

Autonomous Weapons Systems and the Contextual Nature of *Hors de Combat* Status

Steven Umbrello ^{1*}, Nathan Gabriel Wood ²

¹ Department of Philosophy and Educational Sciences, University of Turin, 10124 Turin, Italy; steven.umbrello@unito.it

² Department of Philosophy and Moral Sciences, Ghent University, Blandijnberg 2, 9000 Gent, Belgium; Nathan.Wood@UGent.be

* Correspondence: steven.umbrello@unito.it; Tel.: +393518238010

Abstract: Autonomous weapons systems (AWS), sometimes referred to as “killer robots”, are receiving evermore attention, both in public discourse as well as by scholars and policymakers. Much of this interest is connected with emerging ethical and legal problems linked to increasing autonomy in weapons systems, but there is a general underappreciation for the ways in which existing law might impact on these new technologies. In this paper, we argue that as AWS become more sophisticated and increasingly more capable than flesh-and-blood soldiers, it will increasingly be the case that such soldiers are “in the power” of those AWS which fight against them. This implies that such soldiers ought to be considered *hors de combat*, and not targeted. In arguing for this point, we draw out a broader conclusion regarding *hors de combat* status, namely that it must be viewed contextually, with close reference to the capabilities of combatants on both sides of any discreet engagement. Given this point, and the fact that AWS may come in many shapes and sizes, and can be made for many different missions, we argue that each particular AWS will likely need its own standard for when enemy soldiers are deemed *hors de combat*. We conclude by examining how these nuanced views of *hors de combat* status might impact on meaningful human control of AWS.

Keywords: autonomous weapons; meaningful human control; *hors de combat* status; killer robots; military ethics

1. Introduction

Autonomous weapons systems (AWS) are likely to become a mainstay of modern advanced militaries. These systems come in different forms, shapes, and sizes, and imbued with different levels of autonomy, and thus have different capabilities in the field. There are numerous reasons for the rise of AWS in contemporary militaries. Aside from their use reducing the number of humans on the battlefield and thus reducing possible casualties, such systems can feasibly be capable of waging war more ethically, given their potential speed, efficiency, and precision. Moreover, it has been argued that with the advancement of computer vision and other means of surveillance, the epistemic gap in selecting targets for engagement can be closed even more so than in the case where human operators decide their targets (Arkin 2009; Guetlein 2005).

However, despite these advantages, there remain objections to AWS and their potential deployment. In particular, there is heavy opposition to AWS which are capable of lethal engagement, with organizations like the Campaign to Ban Killer Robots, the International Committee for Robot Arms Control, and the International Committee of the Red Cross (among others) advocating for either partial or full bans on the development and deployment of such systems. The arguments against AWS range from principled objections concerning their deleterious effects on human dignity, to more practical concerns about command and control over systems with opaque machine learning algorithms, or targeting systems which are technically insufficient for meeting the principle of

discrimination (Johnson and Axinn 2013; Purves et al. 2015, Sparrow 2007; 2016; Roff 2013; Guarini and Bello 2012). Proponents of a ban often argue that these points alone require prohibition. However, even granting that there currently exist technical obstacles to properly meet these challenges, this does not preclude such a possibility in the future as technology surpasses those barriers.

In this paper we avoid these principled and practical arguments, and instead explore how existing law might impact on AWS. In particular, we examine the notion of *hors de combat* status, as it is defined in the Geneva Protocol I Additional to the Geneva Conventions (AP I), arguing that advances in AWS will necessitate a more nuanced and varied approach to determining this status for enemy combatants. Our aim is to show that the law demands a contextualized appraisal of whether or not an enemy is in fact *hors de combat*, or “out of combat”, and that, based on this appraisal, more combatants will be deemed *hors de combat* than is usually taken to be the case. These factors imply that AWS will need to be capable of making finer-grained distinctions of *hors de combat* status of enemies based on subtle contextual factors. In making this final point, we also pay heed to the fact that there are many different types of AWS, and that the determination of *hors de combat* status may vary depending on the particular AWS under examination.

The paper is structured as follows. In section 2 we discuss the legal foundations of *hors de combat* status, in particular, AP I, Article 41 and its commentary. We show that *hors de combat* status is likely broader than one might initially suppose, and more importantly, that determining *hors de combat* status requires a contextualized approach, where the relative power of opposing belligerent agents will impact upon the final judgment. Section 3 applies this finding to AWS, showing how different autonomous weapons systems will need to be able to make different *hors de combat* determinations depending on their own abilities and on their precise mission objectives. In section 4, we discuss how these points effect the notion of “meaningful human control” over AWS, and indicate future research which might further the debate.

2. The Safeguards of *Hors de Combat* Status

Contemporary discussions on the ethics and laws of war often place heavy emphasis on the protections afforded to noncombatants. However, in addition to safeguards for noncombatants, historically, the development of many ethical and legal norms was centered on minimizing, or at least mitigating, harm to combatants (Best 1979; Gardam 1993; 1863 St. Petersburg Declaration; ICRC 1977, pp. 480-481). This should not come as a surprise, given that in the 19th century – when the laws of armed combat (LOAC) were being first codified – wars were “conducted by professional armies and the civilian population was not involved to any great extent” (Gardam 1993, p. 397). And it was against this backdrop that the legal status of being *hors de combat*, and its associated protections, was codified.

The currently binding legal articulation of *hors de combat* status can be found in AP I, Art. 41, which maintains that

1. A person who is recognized or who, in the circumstances, should be recognized to be *hors de combat* shall not be made the object of attack.
2. A person is *hors de combat* if:
 - (a) he is in the power of an adverse Party;
 - (b) he clearly expresses an intention to surrender; or
 - (c) he has been rendered unconscious or is otherwise incapacitated by wounds or sickness; and therefore is incapable of defending himself;
 provided that in any of these cases he abstains from any hostile act and does not attempt to escape.

For the purposes of our argument, we will primarily be concerned with 2(a) and 2(c), but before moving onto our discussion, there are two general points worth indicating

with regards to this treaty instrument. First, the protections associated with *hors de combat* status are extended not only when a person is recognized to be *hors de combat*, but also when they *should be recognized* to hold that status. With this, the protections of *hors de combat* status are made more objective than some other protections of the LOAC, as they are in force even when an adversary fails to recognize that his opponent is truly *hors de combat*.¹ Second, the conditions (a)-(c) are disjunctive, meaning that one is *hors de combat* so long as any one of those conditions is met, “provided that in any of these cases he abstains from any hostile act and does not attempt to escape”.

The core question we will now address is who precisely, given the above treaty instrument, is to be deemed *hors de combat* during the course of hostilities? There are some obvious classes of individuals specifically picked out in Art. 41, such as those who are unconscious, those who are so severely sick or wounded as to be utterly incapable of defending themselves, and those who have surrendered. However, what of individuals who are generally combat-effective, yet for some reason cannot defend themselves for contextual reasons, or cannot defend themselves during the duration of a combat engagement (even though they might be threatening later)? Or what of individuals who are unarmed but still dangerous, or those who are armed but so laughably underequipped or undertrained as to be virtually incapable of creating a threat? In what follows, we examine these questions in full, and argue that in most cases, the deciding factor will be contextual, and that by and large, the set of individuals deemed to be *hors de combat* will likely be larger than is usually taken to be the case.

2.1. In or out of combat?

The core intent of Art. 41 is to provide a defense for persons who are no longer a part of combat, yet who have not yet been taken into custody and made prisoners of war.² It is for this reason that it picks out individuals who cannot defend themselves due to unconsciousness, wounds, or sickness – as these persons could be taken into custody at any time – or picks out individuals who have expressed an intent to surrender (but who have not yet been taken into custody). By the same line of reasoning, combatants “in the power of an adverse Party” are taken to be protected due to the fact that they cannot be seen to be in combat any longer (otherwise they would not be in the power of their enemy); a man with effective means to defend himself and thwart capture cannot be seen to be in anyone’s power. Moreover, the AP I official Commentary maintains that “[a]defenseless adversary is *hors de combat* whether or not he has laid down arms” (ICRC 1987, p. 484), showing that the core point of Art. 41 is not whether or not enemies have surrendered (or even intend to surrender), but whether or not they can still be seen to be a part of combat. If they cannot be seen that way, that is, if they are out of combat (*hors de combat*), then they are protected under Art. 41.³

An important implication of this element of Art. 41 is that whether or not an enemy is to be deemed *hors de combat* is likely to depend upon very specific contextual factors in a given case. To see this, let us consider a handful of examples.

First consider a hypothetical encounter set in the First Gulf War, which we will call *Tanker*. Suppose a Coalition tank brigade is rolling through the desert, far from any other friendly soldiers or civilians, and they happen to come across an Iraqi platoon of footmen. The Iraqi soldiers have small arms, but nothing besides that, something which the Coalition tankers can see through their sights. The tankers also do not see any means of communication with which the footmen could call for support or radio in the tanks’ position. The tanks approach the Iraqis, and are presented with a number of choices: they could continue on their way toward their objective and ignore the Iraqis; they could disarm the Iraqis and then continue on their way; they could take the Iraqis prisoner and return to base; or they could engage and kill the Iraqis (something they can do with impunity, given that the Iraqis have no available means for effectively attacking tanks).

In such a case, the Iraqis are armed, and have shown no intent to surrender, but in this precise case, they are also clearly defenseless. Moreover, given that they could fire

their rifles at the tanks all day to no effect, they are, from the tanker's point of view, out of combat. That is, to use the French, they are *hors de combat*. And importantly, they are *hors de combat* from the tanker's perspective not because they are wounded or unarmed (in fact they are both healthy and armed), but because they could not possibly engage the tanks in a meaningful manner. It is the utter and complete irrelevance they hold with regard to the tanks' mission that makes them effectively out of combat, or *hors de combat*.

However, one may object that soldiers are often targeted even when they have no defense, and that this is perfectly correct. More strongly, one might object that it is the goal of military men and women to do their utmost to outmatch their enemies, such that they may strike with impunity. After all, that is the point of calling in air or artillery strikes, to hit an enemy in such a way that they cannot defend themselves, and thus preventing any risk to your own people in the process.⁴ However, the point in the above example is not precisely that the Iraqis are defenseless, but rather that they are out of combat, or *hors de combat*. To put it differently, in order for the words of the AP I Commentary to sensibly capture the customary legal understanding of *hors de combat* status, it should not say that a "defenseless Adversary is *hors de combat*", but rather that a "powerless Adversary is *hors de combat*".⁵ This is because whether or not a soldier can effectively defend himself is beside the point as to whether or not he is deemed "in combat". However, a soldier who is powerless, that is, one who can have no impact on his enemy, truly is out of combat, or *hors de combat*. To better see this, consider a variant of the above example, which we will call *Tanker with Trooper*.

Imagine that everything is as described above, except that the tank brigade is providing escort to a group of Coalition foot soldiers. These additional Coalition troops are militarily on par with the Iraqis encountered, with both groups being capable of inflicting casualties on the other.

In this scenario, the Iraqis are just as outmatched as before. In fact, with a smattering of foot soldiers alongside the tanks, the tank brigade is, if anything, more overpowering to the Iraqis, making the Iraqis, if anything, more defenseless. However, the Iraqis, though clearly outgunned and without hope, can still inflict casualties on this attacking force. As such, the Iraqis are not properly powerless against this enemy (despite being effectively defenseless), and so the Iraqis are clearly not out of combat. That is, they cannot be seen to be *hors de combat*, despite the fact that they are just as, if not more, defenseless than in the previous case.⁶

If one agrees that the Iraqis are not *hors de combat* in the *Tanker with Trooper* case, then it follows that there is more to being *hors de combat* than merely being defenseless. Moreover, it seems that the essential component to being deemed *hors de combat* turns out to be the very intuitive notion of whether or not a soldier is in fact in or out of combat, which itself hinges on whether or not a soldier holds any power to harm his or her enemy. If I can harm my enemy, then I am obviously a part of combat, but if I may be killed and can do nothing to prevent that, and I cannot do anything to harm my enemy in other ways (say, by killing enemy soldiers other than my attacker), then I am no longer a part of combat. (It is also worth pointing out that there is little military reason to kill such an individual, as it costs time and money – in the form of spent ammunition – to kill such a person, but provides only minute military gain. Given this, there is a potential case to be made that killing such men would violate the principle of necessity, as embodied in AP I, Art. 35. (cf. Garraway 2009; Goodman 2013; Meyer 2012; see also Ohlin 2012; Schmitt 2013; Wood 2020).)

At this point, one may object that our position is too broad, in that many soldiers far from the front lines may permissibly be targeted, even though they are not part of combat and may also be defenseless against the attacks on them. For example, one might imagine a military supply train bringing tanks and their crews to the front, but which is still very far from the front and will take some time to get there. Even though it is far away, and even though it won't be dropping off its tanks and soldiers anytime soon, it is perfectly legal to target such a train. As such, basing *hors de combat* status on whether or not a sol-

dier is in combat *at that moment* seems to provide an unfounded (and extreme) widening of the protected status.

To this objection, there is no direct response which can be given. However, it is worth pointing out that *hors de combat* status is not understood in a fully coherent manner. For example, a combatant rendered unconscious during fighting is protected under Art. 41, yet one who is sleeping in his barracks is not so protected. Thus, it is not unthinkable that *hors de combat* status may be extended in one situation and withheld in another which is structurally quite similar. Moreover, even if the military supply train might permissibly be targeted in the above example, it almost certainly should not be targeted in a minor variant of the case, to which we now turn.

Imagine the case as above, but further suppose that the supply train is across a deep river, and as part of the war effort, every strong bridge over that river has been destroyed. In this scenario, the train is trapped and cannot move to the front lines. Every tank aboard the train is still combat effective, but there is no way for them to reach the actual fighting. Given this, the tanks are, for all intents and purposes, out of combat. In this case, we should regard such units as *hors de combat*, not because they are necessarily defenseless (though they may also be defenseless against certain forms of attack), but because they are quite simply out of combat, *hors de combat*. And if this granted, what makes them out of combat in this situation has nothing to do with the tanks or soldiers themselves, but is rather a feature of the context of this scenario. They are out of combat because the bridges are out.⁷

Thus, whether or not an individual (or unit) is deemed to be *hors de combat* will depend upon contextual factors, most importantly, whether or not the individual (or unit) can actually contribute to the fighting. In some cases, combatants will be rendered *hors de combat* because there is no way they can harm their adversary, and in others some cases it will be because they are excluded from the fighting altogether due to the actions of others (or possibly due to environmental hindrances). But in any case, one's status as *hors de combat* will be impacted upon by both that fighter's capabilities, and the capabilities of his enemy. And as a related point, whether or not one is armed need not impact on *hors de combat* status in the least. An unarmed man may still be very capable of killing his enemies, and there may also be armed men who are utterly powerless against their enemies, because their enemies are far better armed and armored. The important point is not what one carries or what one can do, but rather what both you and your enemy carry, and what both you and your enemy can do to one another. The greater the gap between the two, the more likely the weaker party will be likely to be seen as out of combat, simply because the stronger will be impervious to harm from the weaker.

At this point, one may grant all of the above arguments, but still wonder what the precise upshot of this is for commanders and soldiers on the ground. Put differently, assuming the Iraqis are *hors de combat* in the *Tanker* case, what does that mean for the tankers themselves? Must they take the Iraqis prisoner? Must they disarm the Iraqis? Even more strongly, if we have assumed the Iraqis are "in the power" of the tankers, does that mean the tankers are required to provide the protections and supplies which would be demanded if they had taken the Iraqis prisoner?

In answer, we would stress that whether or not the Iraqis (or anyone else) are deemed to be *hors de combat* provides no guidance with regards to these further questions. All that *hors de combat* status demands is that the tankers (or anyone else) refrain from making the Iraqis the object of attack, provided the Iraqis abstain from hostile acts and do not attempt to escape. However, it is up to the tankers themselves whether they choose to simply continue on with their mission, or instead to disarm the Iraqis, or instead to take the Iraqis prisoner. What the tankers may not do is simply engage the Iraqis, as they are to be treated as *hors de combat* in that situation. Importantly, however, they are only *hors de combat* for as long as they act in accordance with the final proviso of Art. 41. Thus, if the tankers demand that the Iraqis throw down their weapons so that the tankers may run over their rifles (effectively disarming those footmen), the Iraqis must comply with this

or accept that they have made themselves liable to attack; they must act the prisoner or the enemy, but cannot enjoy the protections of the former while remaining the latter. In this way, Art. 41 provides a protection for persons who are powerless, but does not give blanket permission for powerless individuals to exploit their weakness. It guards them against initial and unnecessary violence, but simultaneously demands that they either peaceably comply with their victors' requests, or accept that they have forfeited the safeguards of *hors de combat* status.

As a final point, it is worth clarifying that we do not think that the foot soldiers in *Tanker* must necessarily be viewed to be *hors de combat*. Moreover, we accept that there is, and will continue to be, reasonable disagreement about when one party is "in the power of an adverse Party", or when one party is truly defenseless in the face of its enemy's might. Our point is that what renders one defenseless or "in the power" of another cannot be determined with a one-sided assessment of a single parties' capabilities; it must be put in the context of a relational assessment based on the capabilities and limitations of both parties. Thus, being armed or unarmed, being wounded or healthy, or being conscious or not, rarely provides adequate information for determining *hors de combat* status. Rather, that will virtually always demand a view to the capabilities of combatants on both sides of a conflict, and to the differences between them. Moreover, one's status as *hors de combat* (or not) may depend on factors wholly outside one's control, like whether or not bridges have been destroyed, or whether or not the enemy you face has men who are vulnerable to your arms. The core lesson is that persons are deemed out of combat, *hors de combat*, in virtue of a myriad of factors, many of which will be derived from the context within which the assessment is being made.

3. *Hors de Combat* Status and AWS

What exactly does this understanding of *hors de combat* then mean for AWS? There will likely be many implications, but we contend that in light of the widely varied, and potentially dynamic situations in which AWS will be deployed, these systems must be capable of responding to changing and contextualized evaluations of an enemy's status as *hors de combat* (or not). AWS systems should be treated individually given that the contexts of their use – aerial, naval, and ground-based – are substantively different. And even within these broader categories, different types of AWS will come with their own capabilities and limitations, something that will (potentially) change when an enemy is deemed *hors de combat*. For example, a lightly armored autonomous drone may often encounter enemies who are neither defenseless nor powerless, while a heavily armored autonomous assault platform will likely encounter individuals fitting both of those descriptions. This, in turn, requires that such systems possess a level of technical sophistication high enough to allow for calculations which are sensitive to the many contextual factors that will impact upon the relative strength and power of all belligerent groups. Although this is not technically impossible, the viability as well as necessity of this is beyond technical plausibility at present. Nonetheless, what this betrays is that *hors de combat* status is fundamentally tailored by the entities making such evaluations, given both their capabilities and limitations. A foot soldier makes different evaluations than a tank commander, who makes different evaluations than an autonomous turret sentry, all of which would likewise make different evaluations from a Reaper drone.

In order to illustrate this for AWS, we can take the examples above as inspiration. To begin, we will compare two cases to show how mission objectives might alter what AWS would determine as *hors de combat*.

First, imagine a fully autonomous Reaper drone designated to neutralize an insurgent leader, a case we will call *High-Value Target*. The commander of a forward operating base, alongside his tacticians, legal professionals, and other experts, determines that the most efficient plan is to neutralize the target via an aerial strike, and that such a strike is lawful. The commander has a fully autonomous Reaper drone outfitted to undertake the mission. The drone is tasked with taking off, arriving at the target's location, confirming

that the target is present, confirming that the target is not in the vicinity of so many noncombatants as to render the strike disproportionate, releasing its payload, and then flying back to base. However, suppose that while en route to its target, the drone passes over a company of heavily armed enemy combatants who are isolated in the hills. Despite the fact that such a group is heavily armed, their offensive and defensive capacity against a Reaper drone is functionally irrelevant. In such a case, the hostile party is rendered, as in the *Tanker* case, *hors de combat*.

However, how would an AWS fare in a case similar to *Tanker with Trooper*? Let us imagine that the base commander, instead of sending the Reaper alone, decides to deploy a team of Navy SEALs to neutralize the target, and that they are to travel using ground vehicles. In this case, *High-Value Target with SEALs*, an autonomous Reaper drone is deployed to provide close air support for the SEALs, but all other factors are the same as in *High-Value Target*, with the SEAL team encountering the same heavily armed company of enemy troops. In this case, the Reaper should arguably not view the enemy combatants as being *hors de combat*, because those enemies can inflict casualties on the SEALs, and thus are not powerless and are therefore legitimate targets for the Reaper drone. Yet in this case, the Reaper drone plus the SEALs forms an even greater asymmetric advantage over the enemy combatants. However, like the *Tanker with Trooper* case, the factors that determine *hors de combat* status are not simply whether or not one is able to defend oneself, but rather whether or not one has power to affect one's enemy. In this case, despite the advantage held by the SEALs and Reaper, the enemy troops are nonetheless able to inflict casualties, whereas in *High-Value Target* they are powerless against the Reaper, and thus are (arguably) to be deemed *hors de combat*.

Taken together, these points demonstrate that certain classes of people are not to be treated as *hors de combat a priori*. Rather, contextual factors can change the status of the same group of individuals, all other things remaining equal, simply by changing the other actors (machine or human) involved in a given scenario. For AWS, this means that it would be nonsensical, if not technically unfeasible, to create a blanket method for such systems to determine whether enemies should be classed as *hors de combat*. Moreover, this would be the case even for specific types of AWS, because any given combat scenario is marked by dynamism, something which must be reflected in the way AWS operate in order for them to accurately determine whether or not enemies are *hors de combat*. Given the current technical obstacles for such nuanced programming in AWS, we contend that commanders should hold the final say regarding the rules of engagement and adequate standards of due care in such engagements.

To see the value of this in practice, consider the differences between when military forces are taking a city held by enemy forces as compared to when they are occupying said city. (As a real world example, we might envision the differences between taking Mosul from ISIS versus holding Mosul afterward.) During the course of a large-scale operation to clear enemy combatants from an area, and where there are large numbers of noncombatants who might be unintentionally or intentionally harmed by the enemy, it is sensible to treat all enemy combatants as "in combat", regardless of whether they are at a disadvantage or not. This is because, even if they cannot strike directly at their opposing forces, they still hold power to harm noncombatants, and so, unless they fit one of the stricter categories of *hors de combat* (they are seriously wounded or sick, or are attempting to surrender), they should be seen as legitimate targets. However, once the city has been taken, arguably all persons within its environs should be viewed as "in the power" of the forces currently holding the city. As such, many more persons will need to be considered to be *hors de combat*. Moreover, the degree to which persons may be seen to be in the power of attacking forces may change on a daily basis, with districts shifting hands regularly. As such, there may be a need to alter rules of engagement regularly, or to utilize rules which are conditional on certain contextual factors like one's control of an area. All of this points to the need for commanders and combatants on the ground to be able to quickly and effectively alter the way AWS acting in their theater of operations view the

hors de combat status of the enemy. And this, in turn, argues against trying to determine single overarching means of determining this status for all AWS, or even for all AWS in a given environment. Each battle is unique, and commanders should have the ability and means to ensure that their soldiers and hardware, AWS included, are compliant with the laws of war.

To reiterate, what these illustrations aim to demonstrate is that there is no base standard that works for all scenarios. As such, the governing norms in place during deployments of AWS will be highly contextual. In order for AWS to be lawfully deployed, they must be able to pay due heed to the dynamism necessary for determining *hors de combat* status, and it must likewise be assured that humans, in this case deployment commanders, retain the ability to make on-the-fly changes to the targeting principles that are in force in such systems.

In sum, AWS can come in many forms, some more or less fragile or vulnerable. As such, they need to be able to evaluate whether enemies are *hors de combat* in nuanced and contextual ways, paying heed to not just their abilities and their enemies', but also to how changing mission parameters may affect whether or not an enemy is in combat. This, in turn, has implications for our understanding of meaningful human control in the domain of AWS governance and deployment, a point to which we now turn.

4. Regaining Meaningful Human Control (MHC)

Meaningful Human Control (MHC) is a complex, albeit relatively modern concept that emerged from the growing discourse on the ethical and legal issues of AWS. Although there are a number of different frameworks for what constitutes MHC, all frameworks agree that humans must remain in control, or at least have oversight over the decision-making of a system in a non-arbitrary, and thus "meaningful" way. What constitutes this meaningfulness remains debated by scholars, leading to (at least) six different approaches to MHC:

1. Preserving MHC through proper preparation and legitimate context for use, viz. through current NATO targeting procedures (Roorda, 2015);
2. Attaining MHC by having a human agent make "near-time decision[s]" in AWS engagement (Asaro, 2012);
3. Preserving MHC through adequately training commanders in the deployment and function of AWS to ensure proper attribution of responsibility (Saxon, 2016);
4. Attaining MHC through apprising designers/programmers of their moral role in the architecture of AWS (Leveringhaus, 2016);
5. Attaining MHC through design requirements involving necessary conditions to *track* the relevant moral reasons for agent actions and *trace* the relevant lines of responsibility through design histories (Mecacci & de Sio, 2019; Santoni de Sio & van den Hoven, 2018); and
6. Preserving MHC by distributing responsibility for decisions through the entirety of the military-industrial complex (Ekelhof, 2019).

This paper does not aim to propose or endorse any of these approaches to MHC. However, it does merit noting that human agents are never extricated from decision-making in any of the above approaches, nor do they abdicate full decision-making to the system, even if the system can technically be designed with 'full autonomy' (i.e., the ability to select and engage a target without contemporaneous human input). This central tenet is supported by the results of this paper's exploration of *hors de combat*; as we have demonstrated, it would be incorrect and possibly even dangerous to have a single set of targeting principles for AWS designed to determine *hors de combat* status in enemy combatants. This also applies to specific and narrow types of AWS, showing that control must necessarily remain in the hands of commanders for AWS to be lawfully deployed.

This, of course, does not mean that lawfully fielded AWS must be devoid of any programming to that end more generally. On the contrary, it is imperative that AWS possess some general abilities to discern the status of enemy combatants. However, such general patterns will rarely be sufficient alone, and responsible deployment of AWS will have to address this fact. This means that the human overseers must be able to impact on the way AWS view enemy combatants, and the way AWS determine whether enemy combatants are still in combat.

5. Conclusions

Autonomous weapons systems remain a hotly debated topic in both academia, but also in international public spheres. The debate over the ethics and legality of their design and deployment is further complicated by how *actual* military operations are currently carried out, how the letter of the law regarding military operations relates to and differs from the spirit of the law, as well as how different AWS change how they legally relate to potential combatants. In an effort to clear up some of these complications, this paper drew on the legal articulation of *hors de combat* status as found in AP I, Art. 41, showing how this legal principle might impact on the design and use of AWS. What we aimed to show is that *hors de combat* status cannot be wholly reduced to a set of clear categories that can be programmed into AWS, and that some relevant categories are contingent on dynamically changing contexts in the field. These contexts change when one can be seen to be *hors de combat* and thus systems, in order to remain lawful, must be sensitive to this changing status. As such, this paper concludes that military commanders must retain control over AWS' general targeting behavior, in order to be able to respond to the shifting legality of certain targets due to changing contexts. Meeting this imperative would also provide a greater degree of meaningful human control of AWS, an aim which has independent merit and may prove necessary for addressing other legal and ethical concerns related to autonomous weapons systems.

Author Contributions: The authors have contributed equally to the paper. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: We would like to thank Maciek Zajac for providing useful feedback on an earlier draft. All remaining errors are the authors' alone. The views expressed in this paper are not necessarily those of the authors' affiliations.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Arkin, R.C. *Governing Lethal Behavior in Autonomous Robots*, CRC Press: Boca Raton, USA, 2009; pp. 256.
2. Guetlein, M. A. *Lethal autonomous weapons: ethical and doctrinal implications*. Technical report, 2005, United States Naval War College, pp. 34.
3. Johnson, A.S.; Axinn, S. The morality of autonomous robots. *Journal of Military Ethics*, 2013, 12(2), 129-141. <https://doi.org/10.1080/15027570.2013.818399>
4. Purves, D.; Jenkins, R.; Strawser, B. J. Autonomous machines, moral judgment, and acting for the right reasons. *Ethical Theory and Moral Practice*, 2015, 18(4), 851-872. <https://doi.org/10.1007/s10677-015-9563-y>
5. Sparrow, R. Killer robots. *Journal of Applied Philosophy*, 2007, 24(1), 62-67. <https://doi.org/10.1111/j.1468-5930.2007.00346.x>
6. Sparrow, R. Robots and respect: Assessing the case against autonomous weapon systems. *Ethics & International Affairs*, 2016, 30(01), 93-116. <https://doi.org/10.1017/S0892679415000647>
7. Roff, H.M. Killing in war: Responsibility, liability, and lethal autonomous robots. In *Routledge Handbook of Ethics and War: Just War Theory in the 21st Century*, Allhoff, F., Evans, N. G., Henschke, A., Eds.; Routledge: pages 352-364. Routledge: Milton Park, UK, 2013, pp. 352-364.
8. Guarini, M.; Bello, P. Robotic warfare: Some challenges in moving from noncivilian to civilian theaters. In *Robot Ethics: The Ethics and Social Implications of Robotics*, Lin, P., Abney, K., Bekey, G.A., Eds., MIT Press: Cambridge, USA, 2012, pp. 129-144.
9. Best, G. Restraints on war by land before 1945. In *Restraints on War: Studies in the Limitation of Armed Conflict*, Howard, M. E., Ed., Oxford University Press: Oxford, UK, 1979, pp. 17-37.
10. Gardam, J. G. Proportionality and force in international law. *The American Journal of International Law*, 1993, 87(3), pp. 391-413.

11. ICRC. *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949*. MartinusNijhoff Publishers: Geneva, Switzerland, 1987.
12. Restrepo, D. Naked soldiers, naked terrorists, and the justifiability of drone warfare. *Social Theory and Practice*, 2019, 45(1), pp. 103-126. <https://doi.org/10.5840/soctheorpract201911153>
13. Restrepo, D. Excuses, justifications, and the just war tradition: Are there good reasons to kill the naked soldier? *Journal of Global Ethics*, 2017, 13(1), pp. 58-69. <https://doi.org/10.1080/17449626.2017.1302494>
14. Restrepo, D. In defense of mercy. *Journal of Military Ethics*, 2020, 19(1), pp. 40-55. <https://doi.org/10.1080/15027570.2020.1738981>
15. Zając, M. Spare not a naked soldier: a response to Daniel Restrepo. Forthcoming.
16. Garraway, C. To kill or not to kill?: dilemmas on the use of force. *Journal of Conflict & Security Law*, 2009, 14(3), pp. 499-510. <https://doi.org/10.1093/jcsl/krp032>
17. Goodman, R. The power to kill or capture enemy combatants. *European Journal of International Law*, 2013, 24(3), pp. 819-853. <https://doi.org/10.1093/ejil/cht048>
18. Mayer, C. Minimizing harm to combatants: Nonlethal weapons, combatants' rights, and state responsibility. In *Routledge Handbook of Ethics and War: Just War Theory in the 21st Century*, Allhoff, F., Evans, N. G., Henschke, A., Eds.; Routledge: pages 352-364. Routledge: Milton Park, UK, 2013, pp. 301-311.
19. Ohlin, J. D. (2012). The duty to capture. *Minn. L. Rev.*, 2012, 97, pp. 1268-1342.
20. Schmitt, M. N. Wound, capture, or kill: A reply to Ryan Goodman's the power to kill or capture enemy combatants. *European Journal of International Law*, 2013, 24(3), pp. 855-861.
21. Wood, N. G. The problem with killer robots. *Journal of Military Ethics*, 2020, 19(3), pp. 220-240. <https://doi.org/10.1080/15027570.2020.1849966>
22. Roorda, M. NATO's Targeting Process: Ensuring Human Control Over (and Lawful Use of) 'Autonomous' Weapons, In *Autonomous Systems: Issues for Defence Policymakers*, Andrew Williams, A., Scharre, P. Eds., NATO Headquarters Supreme Allied Command Transformation, 2015, Norfolk, USA, pp. 152-168.
23. Asaro, P. On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross*, 2012, 94(886), pp. 687-709. <https://doi.org/10.1017/S1816383112000768>
24. Saxon, D. Autonomous drones and individual criminal responsibility. In *Drones and Responsibility: Legal, Philosophical, and Socio-Technical Perspectives on the Use of Remotely Controlled Weapons*, Di Nucci, E., Santoni de Sio F., Eds., 2016, CRC Press: Boca Raton, USA, pp. 17-46.
25. Leveringhaus, A. (2016). Drones, automated targeting, and moral responsibility. In *Drones and Responsibility: Legal, Philosophical, and Socio-Technical Perspectives on the Use of Remotely Controlled Weapons*, Di Nucci, E., Santoni de Sio F., Eds., 2016, CRC Press: Boca Raton, USA, pp. 169-181.
26. Mecacci, G.; Santoni de Sio, F. Meaningful human control as reason-responsiveness: the case of dual-mode vehicles. *Ethics Inf Technol*, 2020, 22, pp. 103-115. <https://doi.org/10.1007/s10676-019-09519-w>
27. Santoni de Sio, F.; van den Hoven, J. Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Front. Robot. AI*, 2018, 5, p. 15. <https://doi.org/10.3389/frobt.2018.0001>
28. Ekelhof, M. Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation. *Global Policy*, 2019, 10(3), pp. 343-348. <https://doi.org/10.1111/1758-5899.12665>

¹Note that though one is to be protected whenever one should be recognized to be *hors de combat*, it is only a war crime to make "a person the object of attack in the knowledge that he is *hors de combat*" (AP I, Art. 85.3(e)). Thus, it is possible to fail to meet the demands of Art. 41, while still not acting in a criminal manner, as long as that failure is based in ignorance. Once one knows an enemy is *hors de combat*, any targeting of that enemy will constitute a "grave breach" of the treaty, and thus be deemed a war crime.

²ICRC 1987, paragraphs 1601-1603 (pp. 480-482) Note also that the relevant factor is whether or not a person is him- or herself in combat, but not necessarily whether or not a person is still in a military engagement. An individual may still be (partially) involved in a military engagement in virtue of being physically present where hostilities continue to unfold, yet be *hors de combat* due to wounds. As a result, it is possible that some *hors de combat* persons may suffer harm simply due to their proximity to battle, but this will not mean that those who harm them have breached Art. 41, so long as the *hors de combat* persons are not made the object of attack.

³Recent debates on the killing of so-called "naked soldiers" make arguments that are similar to those presented here. However, these debates reach a much stricter conclusion than what we are advancing, and moreover seem to be far less sensitive to the particularities and importance of context in determining the permissibility of targeting (lethal or otherwise). For these reasons, we view these positions with caution, as they seem liable to prove too much. For such

arguments, see the recent position developed in Restrepo 2017; 2019; 2020. For a response to this see the forthcoming Zajac n.d.

⁴For example, such arguments are put forward in Zajac n.d. Zajac's points are made in relation to the so-called "naked soldier" debates (supra, note 3 above), but with very minor adjustments they would be relevant here as well.

⁵Alternatively, one might also simply ignore the words of the Commentary. However, it is our belief that if a possible (and plausible) interpretation of the Commentary allows one to reconcile its central findings with customary international law, then such is preferable, especially in light of Art. 38(1)(d) of the Statute of the International Court of Justice, which maintains that the court shall apply "judicial decisions and the teachings of the most highly qualified publicists of the various nations, as subsidiary means for the determination of rules of law". Given that the International Committee of the Red Cross represents one of the most respected authorities on International Humanitarian Law, we hold that its Commentary cannot be simply ignored in those instances where it appears to run against common opinion or state practice, but should instead be interpreted in the most charitable light which might allow it to be reconciled with such positions. Thanks to Maciek Zajac for pushing me on this point.

⁶This point reflects the facts of many cases in modern warfare. For example, aircraft providing close air support nearly always target individuals who are defenceless against them. However, these craft are providing support to ground troops who are in harm's way, and are under threat from precisely those individuals being targeted by the aircraft providing support. Thus, though the individuals targeted by close support aircraft are defenceless against such craft, those individuals are not powerless, as they can still harm the ground troops being support by air cover. Perhaps more controversially, aircraft or drones carrying out kill strikes in counterinsurgency environments may be viewed as striking enemies who are defenceless and out of combat. However, we would argue that if those targeted are in a position to harm others (friendly or third party individuals), then they may be justifiably deemed to be in combat, or at least close enough to combat to be permissibly made the object of attack. At any rate, whether or not they are *hors de combat* will depend on the precise nature of the situation, and most importantly (for our purposes) on a number of contextual factors.

⁷If the state contemplating such a strike has a legitimate war aim of disarming its enemy, then it may still be permissible to destroy these tanks. However, this would arguably only be permissible if the tanks could be struck without harming the crews of said tanks. The reason for this is because, so long as the tanks have no possibility of joining the fighting, the war may be carried out without doing anything with regards to these units, and they may still be destroyed after the war has ended (such disarmament may even be made an explicit condition of the peace settlement). It is also worth emphasizing that this example is predicated on their being some degree of certainty that the bridges will stay down. Such certainty will be possible in some cases, but less so in others, and whether or not one knows the bridges will stay down will certainly affect the permissibility of targeting decisions against units across the river. Again, thanks to Maciek Zajac for suggesting these points (among others).