Ethics and Information Technology (2023) 25:13
https://doi.org/10.1007/s10676-023-09689-8

ORIGINAL PAPER

AWS compliance with the ethical principle of proportionality: three possible solutions

Maciek Zając[1]

Abstract
The ethical Principle of Proportionality requires combatants not to cause collateral harm excessive in comparison to the anticipated military advantage of an attack. This principle is considered a major (and perhaps insurmountable) obstacle to ethical use of autonomous weapon systems (AWS). This article reviews three possible solutions to the problem of achieving Proportionality compliance in AWS. In doing so, I describe and discuss the three components Proportionality judgments, namely collateral damage estimation, assessment of anticipated military advantage, and judgment of "excessiveness". Some possible approaches to Proportionality compliance are then presented, such as restricting AWS operations to environments lacking civilian presence, using AWS in targeted strikes in which proportionality judgments are pre-made by human commanders, and a 'price tag' approach of pre-assigning acceptable collateral damage values to military hardware in conventional attritional warfare. The article argues that application of these three compliance methods would result in AWS' achieving acceptable Proportionality compliance levels in many combat environments and scenarios, allowing AWS to perform most key tasks in conventional warfare.

Keywords Autonomous weapon systems · Ethics of War · Ethics of Military Technology · Just War Theory · Military Ethics · Principle of Proportionality · Law of Armed Conflict · International humanitarian law

Introduction

Autonomous weapon systems (AWS)[1] are currently at the center of vigorous debates on the ethical and responsible use of AI in military contexts. One of the key issues in these debates is AWS' capacity for compliance with the ethical rules for right conduct in warfare, or *Ius in Bello*. Among the moral principles that form the core of *Ius in Bello*, the Principle of Proportionality[2] that is frequently considered one of the hardest principles for an AWS to comply with (Amoroso, 2020, 76–96; Asaro 2012; Brenneke, 2018; Foy, 2014). Without offering a solution for the compliance-with-Proportionality problem one cannot make a genuinely compelling case for the moral permissibility of using AWS. This article aims to offer some such solutions.

I begin by analyzing the Principle, both in its general formulation and in its application via a three-part Proportionality Test. I argue that the three components of this test may be performed independently of each other by different persons, and that one component – the collateral damage estimation – could be competently performed by dedicated software. I then proceed to discuss three mutually supporting solutions to the problem of AWS compliance with Proportionality, showing how these could assure sufficient compliance in a substantial number of combat environments and scenarios, including some types of attacks in conventional land warfare. I close with defending these solutions from anticipated objections.

[1] US Department of Defense defines AWS as "a weapon system that, once activated, can select and engage targets without further intervention by a human operator". In this article I will adopt this commonly used definition - https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf.

[2] To avoid confusing the Reader I will capitalize the term 'proportionality' whenever I refer to the *Ius in Bello* principle, rather than to the concept itself.

✉ Maciek Zając
maciekzajac1@gmail.com; meciej.zajac@uw.edu.pl

[1] Institute of Philosophy and Sociology, Polish Academy of Sciences, Warsaw, Poland

Published online: 13 February 2023

The Principle of Proportionality

This article is concerned with ethics, not law, and so with the ethical rather than legal Principle of Proportionality. However, I will treat the formulation of the Principle being used in the Laws of Armed Conflict (LOAC) as the basic formulation. I do so even though the requirements of law and ethics may be quite different, basing this choice on several considerations. First of all, the legal formulation does indeed capture the substance of the ethical Principle. Secondly, critics of AWS frequently use the legal formulation (Amoroso, 2020, 77–78; Brenneke 2018, 74; Foy 2014, 55), rather than consistently using some alternative one. And thirdly, the legal formulation is representative of actual military practice, and not only of what most military personnel actually do, but of what they ethically aspire to do. Consequently, it is the Principle as formulated in the law that is the most frequent object of scholarly debate.

This is not to say that the current legal standard is identical with the ethical one. The legal standard is, indeed, quite permissive, and following a more restrictive standard may very well be ethically required. For the purposes of my argument, however, I will remain agnostic as to whether the ethical Principle of Proportionality is more restrictive than the current legal standard[3]. Indeed, one of the virtues of my argument is the fact that if it is correct, it is also correct in case a more restrictive standard should be followed. Let us now examine both the current legal standard and a proposition of a more restrictive one.

Rule 14 of the ICRC's compilation of Customary International Humanitarian Law (IHL) states that.

> [l]aunching an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated, is prohibited.

The wording of Rule 14 is for all practical purposes identical with the four different formulations of the Principle found in Additional Protocol I: "in Article 51(5)(b), as part of the prohibition on indiscriminate attack; (2) in Article 57(2)(a)(iii), as part of the precautionary considerations when launching an attack; (3) in Article 57(2)(b), for when an attack is in progress and may need to be cancelled or suspended; and (4) in Article 85(3)(b) on acts which, when done willfully, are regarded as grave breaches of AP I" (Homayounnejad, 2019, 234). Article 8(2)(b)(iv) of the Rome Statute for the International Criminal Court also criminalizes breaches of proportionality, calling attacks involving incidental damage that "would be clearly excessive in relation to the concrete and direct overall military advantage anticipated" "serious violations" of the LOAC[4]. The Principle of Proportionality is thus firmly embedded in both statutory and customary law, and in the ethical theory of *Ius in Bello* (McMahan, 2009, 18–32; Orend 2006/2013, 125-6). The Principle may even be considered quintessential, the *Ius in Bello* in microcosm, as it "is aimed at establishing an equitable balance between humanitarian requirements and the sad necessities of war" (Sandoz, Swinarski & Zimmerman 1987, 685, paragraph 2219).

As mentioned, the Proportionality standard in its legal form may be considered to be overly permissive. As stated by the ICRC Commentary, "the provision allows for a fairly broad margin of judgment" (1987, 684, paragraph 2210). This streams from the fact that "there is no matrix, no order, no formula that resolves that [proportionality] dilemma" (Solis, 2016, 297). Reasonable persons may disagree in edge cases whether a certain attack was proportional or not. As the presumption of innocence is a principle of all criminal law, in cases of doubt the law permits an attack to be carried out (Haque, 2017, 200). According to Yoram Dinstein, "the adjective 'excessive' means proportionality is not in doubt" (2016, 155, paragraph 417).

An ethical standard might indeed be much stricter. Adil Haque claims that in cases where an attack's proportionality is unclear, the attack ought to be deemed ethically impermissible:

> Harming civilians is not wrong because it is disproportionate. On the contrary, harming civilians is wrong unless it is proportionate. Attacking forces should presume that collaterally harming civilians is wrong unless they acquire decisive reason to believe that they are justified in doing so. Accordingly, combatants should refrain from attacks that are not clearly proportionate (2017, 201).

I believe that if AWS can meet the current legal standard in the circumstances I describe below, then they would also meet Haque's more restrictive one. Consequently my argument does not depend on the validity of Haque's critique of the current standard, or on the validity of other critiques aimed at establishing the ethical standard as stricter than the legal one. It does, however, depend on another, less controversial set of claims: that a Proportionality judgment can be divided into three distinct components; that collateral damage estimation may be performed independently by different agents than those acting in combat, and without lowering the Proportionality judgment's validity; that collateral

[3] Neither do I make any claims as to whether the current legal standard entails any obligations additional to the ethical ones.

[4] The addition of the adjective 'clearly" makes the Rome Statute's formulation the most permissive one.

damage estimation is machine feasible, that is, that software could perform in this respect on par with human analysts, if not much better; and that two other components of the Proportionality judgment – the assessment of the anticipated military advantage and the judgment of excessiveness – can sometimes be performed several hours or even days ahead of a strike without their validity being diminished.

The three components of the proportionality test

"Proportionality analysis is made of three steps: i) collateral damage estimation; ii) military advantage assessment; iii) determination of excessiveness" (Amoroso, 2020, 78). Before discussing each of these, it is important to note that Proportionality is a requirement attached to a commander's decisions *ex ante*, not *ex post* (ibid.); "the terms 'expected' and 'anticipated' make clear that the commander's judgment of the opposing variables (Estimated Collateral Damage and Military Advantage Anticipated) is made before an attack is launched (…) Should this information be subsequently found to be flawed or incomplete, it is still the ex ante situation that matters for the legal assessment of proportionality." (Homayounnejad, 2019, 235).

Collateral Damage Estimation (CDE). CDE consists of predicting the incidental harm caused by using a particular weapon against a particular target in a particular area at a particular time. An example of CDE would be trying to predict whether a building inhabited by civilians may suffer structural damage or collapse if a 2000 pound bomb is used to destroy a nearby bridge, or whether a missed shot by a sniper could harm a civilian located behind the sniper's intended target.

Some scholars believe CDE can be automated (Homayounnejad, 2019, 237). This belief is well justified: reliable methodologies for estimating collateral damage already exist and are used by commanders and their staffs (though the exact details of these methodologies are often classified – Wright 2012, 831-3). These methodologies are presently incorporated into software and such software is already used to aid human decision makers. What it essentially does is physics, not ethics – predicting the effects of explosive and other area effecting weapons on their area of impact. Consequently, even AWS-skeptical authors do not question AWS' ability to get CDE right in its basic form (Amoroso, 2020, 78–81; Brenneke 2018, 75).

That does not mean, however, that automating this step of the proportionality analysis is completely free of controversy. The first point of contention is the issue of so-called reverberating effects that "generally ensue when an infrastructure providing critical services to the civilian population is destroyed or severely damaged" (ibid., 80). Suppose a strike against a strategic bridge also predictably damages a power plant located next to it, cutting off a city's electricity for a few days. A number of civilians die not in the bombing itself, but because of the power outage (e.g., they cannot call ambulances because their phone batteries have run out, and so die of perfectly treatable ailments). Should these indirectly caused casualties be included in proportionality calculations? My answer is resolutely "yes". Interruption or lasting denial of access to basic commodities like potable water, electricity, transport infrastructure, and healthcare predictably and undeniably result in civilian casualties, and not taking this fact into account constitutes indefensible moral callousness. While some LOAC scholars might disagree that the currently existing law should be interpreted in this way, I believe that ethically, the matter is clear.

That said, taking this more restrictive approach would not create a problem for AWS users. As mentioned, the reverberating effects come from attacks on essential infrastructure whose location hardly ever changes throughout the conflict (and if it does – for example when a hospital is relocated – this should be communicated to the enemy under provisions of API's Article 48). Thus, battle staffs should have no problem with compiling a database of such locations and inserting it into an AWS' memory with proper restrictions on engaging such targets (or other targets within or nearby these locations).

A final, yet crucial observation about CDE – there is no reason why CDE could not be performed by a different agent than the one assessing anticipated military advantage and/or making the judgment of excessiveness. Indeed, the commander, especially a higher level one, sometimes both has to and should rely on the estimates made by expert analysts or specialized software. If one person (or a piece of software) performs reliable CDE, there is no reason why another could not make a valid proportionality judgment based on this estimate[5].

Anticipated Military Advantage (AMA). While CDE may often be performed with rough exactness and render a quantifiable result, AMA can almost never be calculated in this way (Brenneke, 2018, 84; Foy 2014, 60). War is not a game of chess – there is no omniscient analytical engine that can rate combatants' decisions and output their impact on the prospect of achieving a given goal in numerical form[6].

[5] To say otherwise would be to say that people, including lawyers and ethicists writing on this issue, have no capability for making valid judgments of excessiveness in real or hypothetical cases, since presumably few of them are qualified to perform CDE.

[6] While one may envision some "AlphaGoWar" doing exactly that in a more distant future, such a prospect may be considered deeply undesirable for a number of reasons. Nor will such technology be accessible to AWS users within the timeframe discussed in this paper.

Strategy and tactics may have their rules, but warfare is not an exact science. That does not mean, of course, that skilled commanders or even rank-and-file combatants cannot asses specific military developments as good or bad – they routinely do so. Yet while CDE may usually render a more or less broad estimate of how many will be *killed* if an attack is launched, it is comparatively rare for commanders' to be able to estimate how many will be *saved* if the attack goes through. Nor is there usually available any other simple currency in which AMA may be expressed.

What does AMA actually consist of then? AMA needs to be "concrete and direct", as stipulated by Rule 14 and AP I. Yoram Dinstein explains that AMA "must be perceptible, particular and real as opposed to general, vague and speculative" (2016, 161). It "can only consist in ground gained and in annihilating or weakening the enemy armed forces" (Sandoz et al., 1987, 685, paragraph 2218). These limitations exclude any actions aimed at civilian populations or the adversary's economy in general, hallmarks of twentieth century total war, and leave two currencies for measuring AMA – casualties (combatants and materiel) and territorial gains[7].

However, the value of positional gains, or even the value of materiel gains, is highly relative and context-dependent, to the point where some, like Dinstein (2016, 162) and Homayounnejad, who approvingly quotes him (2019, 237), believe that these are to be assessed according to the attacking commander's "subjective state of mind", that is, in relation to the battle plan he makes. On the one hand, that is obviously true – all tactical gains are ultimately being achieved to realize operational and strategic plans, and so should be evaluated in their context. Destroying a single tank may not seem significant, but if that tank is preventing traffic on a road that is to become a main route of a major offensive, the calculation changes; had the commander planning the offensive chosen a different route, it would not. In this sense, AMA is relative to the objectives contingently set by a commander and thus subjective. However, it seems that at least as far as the ethical aspect of the matter is concerned, the standard of a "reasonable military commander", binding for the third step of the proportionality analysis, should also be applied to AMA. A commander or a combatant may be and sometimes is patently wrong is his estimation of the military advantage attached to a certain outcome; a commander may overvalue holding a position that is clearly useless, or may attach too much value to destroying an outdated platform. Such tactical blunders may render their AMA assessments flawed. A commander being irrationality attached to an erroneous military theory or tactical decision should not serve as an excuse, let alone as a justification, for a disproportionate attack.

I have so far talked of AMA estimation as if it was performed mainly at the strategic and operational level – the level at which specialist battle staffs plan and deliberate on large operations and targeted strikes. However, low-level tactical decisions, which also require AMA estimation, are entirely different. The scope and depth of AMA estimation performed by the rank-and-file combatants is quite limited in comparison to the strategic, operational, or even battalion level. It would be completely wrong to imagine infantry sergeants, individual pilots, or tank commanders having perfect knowledge of the tactical, operational, and strategic consequences of their actions prior to undertaking these actions, thinking several steps ahead. "I destroyed this tank despite the presence of human shields around it, as I was aware this act would start a cascade of events that would lead to us winning the battle" is a justification one will almost never hear at this level, as this kind of analytical clarity is rarely achieved by top generals, and uncommon even in historians analyzing engagements *post factum*.

Indeed, it is rare for combatants in the heat of tactical-level combat to understand the link between attacking a given target and the achievement of a higher-level objective. "I destroyed this tank because I was afraid it would kill me"; "I destroyed this tank because if we destroy more of their tanks than they destroy ours, we will eventually win"; "I destroyed this tank worth several million dollars with a missile worth eighty thousand dollars, a good bargain" – these are the types of reasons that drive tactical decisions at a squad or platoon level. The goods secured via such tactical moves are limited to the security of the platoon members and those they are meant to safeguard, the attrition of the enemy force, and territorial and tactical gains. The sum of such goods secured through tactical moves aids higher goals, as when the resultant imbalance of strength, achieved through attrition, leads the enemy to withdraw or surrender. Yet combatants are rarely contemplating how a given tactical action may secure a great strategic good, and even if they are, they can only imperfectly estimate the scope of the goods they are securing. What is mostly being achieved, predictably and immediately, is military advantage quantified in terms of the number of enemy casualties, materiel lost, and ground gained – as per the aforementioned instructions of the ICRC Commentary[8].

[7] Territory is not valuable in itself, but only insofar as it is strategically or economically vital or populated by people who are in need of liberation from unjust occupation. It goes without saying the liberating a city of two million is more of a gain than liberating a much larger but desolate area, assuming the strategic value of both is equal.

[8] It is true that combatants at a tactical level may sometimes obtain reliable knowledge of the operational/strategic importance of their tactical level actions – think of history's famed rearguard actions. But even then, and especially then, the value of specific tactical goals is quantifiable and does not require complex strategic reasoning to

To summarize, AMA is estimated differently at different levels of military hierarchy, as combatants at different levels have varying epistemic access to the overall picture of events. Lower-ranking combatants have to trust that the assessments of their epistemically privileged superiors are simply better than their own; absent instructions from their superiors, they use the much more restricted information available to them to conduct much simpler estimations of their own. Consequently, they cannot and are not required to conduct complex assessments – they either trust the orders sent down the ranks or perform their own AMA estimates, taking into account just their local tactical situation.

Determination of Excessiveness. "Once expected collateral damage and anticipated military advantage are attributed a value, the attacking force has to establish whether the former is 'excessive' in relation to the latter. Here lies the core of the principle of proportionality, as well as the source of the most intractable problems that it poses" (Amoroso, 2020, 84). The biggest of these is posed by "the need to compare *contradictory* and *dissimilar* values with *no common metric*" (Homayounnejad, 2019, 244), which " introduces a strong element of subjectivity, which bedevils consensus in a given case or consistent application across cases" (ibid., 234). The other is engendered by "the *inherently indeterminate* nature of 'excessive'" (ibid., 244) and the relationship between the concepts of 'excessive' and 'disproportionate'. Finally, there exists an obvious yet underemphasized problem of undue partiality, as very substantial discretion in Proportionality decisions is granted to the people personally and professionally affected by these very decisions.

The first of these problems may justifiably be considered philosophically intractable. Even if AMA could be estimated with reasonable exactness and objectivity, translating it into the currency of human lives is usually impossible. And if and when this is possible, the issue is still far from straightforward. Even supposedly simple thought experiments may bring out differing intuitions on this issue. Is it permissible to kill three enemy soldiers while collaterally killing two civilians? The thought experiment comes with the levels of certainty quite rarely attainable in combat – we know the exact number of persons involved, their exact fate if a weapon is fired etc. Let us also assume there is no further tactical gain to the situation beyond attrition of the enemy force. Is three combatants for two civilians acceptable? Well, it still depends. What is the usual ratio in that particular combat environment? In fierce urban fighting, three for two might be acceptable, perhaps even better than usual. In most other circumstances, not so much. Killing these five persons would almost certainly not rise to a violation in a legal sense, to be sure – but this is not a decision one would like to be routinely taken if war came into one's own backyard. In the twentieth century, while wars killed more civilians overall, such a ratio of *battle* deaths would be disappointing, if not unheard off[9].

Gary Solis provides an example of an American captain who, being under fire and having two wounded, shot missiles at a building he knew contained a mother and a child (2016, 297), killing them, probably alongside some Taliban fighters (the captain did attempt to flush the occupants out of the building by other methods first). Assuming his two wounded men he was trying to get to safety would die without medevac, was this an acceptable tradeoff? Thomas Hurka believes that such a tradeoff would be acceptable, stating that: "while a nation may prefer its own civilians' lives to those of enemy civilians, it may not do the same with its soldiers' lives. Instead, it must trade those off against enemy civilians' lives at roughly one to one." But then he continues: "This is not to say that an act that kills 101 civilians as a side effect of saving 100 soldiers is necessarily disproportionate; the comparisons cannot be that precise. But it does imply that any act that kills significantly more civilians than it saves soldiers is morally impermissible" (Hurka, 2005, 64). In other words, if two mothers with two children each were seen inside the house, no missiles could have been fired and the soldiers would have to accept that their two comrades would die, or attempt an effort more heroic than firing a Javelin missile from behind cover to save them.

Situations in which both possible outcomes are as clear and isolated from other circumstances are relatively rare in war. Nor are human lives the only objects of ethical value to be weighed against AMA. Comparing the value of human health, the value of property (including essential infrastructure and housing), or the value of cultural objects, religious sites, or scientific installations all pose their own conundrums. Making such comparisons requires answering complex questions to which philosophers do not currently have established answers.

Nor can we be certain such questions can be answered at all, let alone with exactness. Is maiming five better than killing one? Is rendering twenty families homeless better than killing one person? How about preventing the destruction of the Notre Dame? Is this worth a life? Is this worth risking a life? We would probably consider an exact answer to any of these questions – for example, saying that rendering seven families homeless to save a single life is acceptable but doing this to eight families is not – a marker of insanity. It seems that the best we can do is ruling out some answers as wildly implausible and perhaps even wholly unreasonable. Yet even here there exists a space for disagreement, as answers differing by an order of magnitude or even more

arrive at. Holding this pass is worth the lives of the Frankish army; one knows how many comrades are to be saved.

[9] In the conflicts in which civilians were being directly attacked, much worse ratios could naturally be expected.

can sometimes both be plausible. Saving the Notre Dame cathedral from complete destruction may be worth sacrificing one or ten lives, depending on who you ask, though sacrificing a thousand seems clearly excessive and sacrificing one hundred thousand seems insane. No wonder that when discussing Proportionality, publicists usually resort to paradigm cases that set the boundaries of a large gray area (Amoroso, 2020, 85; Dinstein 2016, 156).

The framers of the LOAC knew that weighing collateral damage and AMA with any kind of exactness is not possible. Consequently, the term 'proportionality' does not figure in the law, supplanted instead by prohibitions on 'excessive' or 'clearly excessive' collateral damage. In the words of Gary Solis, " 'close' [collateral damage – M.Z.] issues do not rise to a violation" (2016, 294). "A significant and unreasonable outweighing of military advantage anticipated by estimated collateral damage is needed before the rule is violated" (Homayounnejad, 2019, 238).

Of course the term 'excessive' is also vague, and introducing it does not solve borderline cases. However, it radically changes the character of the burden placed on combatants by the Proportionality Principle; they may proceed unless what they are planning to do is clearly wrong. The ICRC Commentary may be read as attempting to restrict action in case of doubt: "the disproportion between losses and damages caused and the military advantages anticipated raises a delicate problem; in some situations there will be no room for doubt, while in other situations there may be reason for hesitation. In such situations the interests of the civilian population should prevail, as stated above" (Sandoz et al., 1987, 625-6, paragraph 1979). Yet this passage refers only to doubts as to whether excessive damage is caused. Even on this interpretation, some degree of disproportion is clearly allowed.

It must be noted that the problem of vagueness cannot be surpassed by creating more classification methodologies for CDE and AMA, it is merely transferred onto those methodologies. J.D. Wright proposes classifying both CDE and AMA as either marginal, moderate, or substantial, creating a 3×3 matrix for estimating whether collateral damage would in fact be excessive (2012, 852)[10]. Matching substantial military advantage with marginal collateral damage is clearly proportionate; matching marginal AMA with substantial CDE, clearly excessive. When classifications are equally matched, say, both expected values are moderate, Wright deems an attack proportionate but advises refraining from it in counterinsurgency scenarios. While his guidance is an improvement over instructions provided by the law itself, the problem of vagueness still haunts it all the same. Instead of deliberating over the meaning of 'proportionate' or 'excessive', Wright's readers are invited to deliberate over the meanings of 'marginal', 'moderate' and 'substantial'.

In summary, incommensurability and vagueness cause Proportionality to be clearly more permissive than it may appear on first glance, or on some interpretations[11]. The Principle's restrictive strength is further weakened by the fact that a person undertaking the proportionality analysis, or at least making the final call, is by law a military commander, desirous of success for personal reasons and sharing a special emotional and professional bond with his troops. Unsurprisingly, commanders tend to imagine the military advantage of their preferred moves to be greater than it actually is (Haque, 2017, 197), and value the lives of their troops higher than those of civilians, especially civilians from the enemy nation (Kowalczewska, 2021, 101). Thus, they may be counted on to make use of all the elasticity that the vagueness of Proportionality allows. As Judith Gardam puts it, "the lack of precision operates in the interest of the military rather than that of civilians" (1993, 407). The situation is no different when a civilian politician – a president or a defense minister – is making a proportionality call on a particularly high-collateral-damage strike (Solis, 2016, 297). Biases common among the military brass are different than biases of top-level politicians, but both are equally likely to introduce non- or un-ethical considerations.

Given all of the above, it may be tempting to consider Proportionality as overly permissive, and therefore easy to comply with, even for AWS. However, this is not actually the case[12]. First of all, Haque's very demanding yet simple solution of considering impermissible all attacks that are not clearly proportionate provides a way to avoid making our philosophical limitations a source of ethical permissiveness. Secondly, the fact that humans are bad at making precise proportionality judgments, and at explaining the intuitions that lead to those judgments, does not change another fact, namely that such judgments are clearly beyond the capacity of any artificially intelligent software below the level of human intelligence and general understanding of the world, precisely because they require the agent making them to have detailed understanding of the social and ethical landscape to which they pertain. Homayounnejad calls Proportionality judgments "arguably, the most 'cerebral' and abstract judgment that calls for metacognitive thinking" (2019, 97, also 243). No matter how permissive the standard

[10] Ronald Arkin has proposed an even more elaborate 5×4 matrix (2009, 188, Fig. 12.8) which runs into identical problems.

[11] The Commentary passage quoted in the previous paragraph is immediately followed by Paragraph 1980, which equates the prohibition on excessive damage with a prohibition on extensive damage, something which is clearly not the case (Dinstein, 2016, 156-7; Homayounnejad 2019, 240). This misinterpretation by the Commentary is one of the reasons for which Proportionality may commonly be considered far more restrictive than it really is.

[12] I am thankful to an anonymous Reviewer for pressing me to reflect more deeply on this line of argumentation.

would be, AI agents short of human intelligence could not fit it (Kowalczewska, 2021, 102), because they are not the kind of agents that may reason in the fashion required by the judgments of excessiveness.

We do well to acknowledge and remember this. Yet we also do well to remember that AWS, as we have defined them, do not make decisions – their human programmers and commanders do. The question we should be asking is therefore not "can these machines make Proportionality judgments?"[13], but "can humans make Proportionality judgments for AWS ahead of time in some circumstances?". Answering this latter question requires reflecting on how frequently and in what circumstances and roles AWS would have to make proportionality judgments. As argued below, the need to perform these would hardly be universal.

Applying proportionality to AWS

Proportionality Judgments As Frequently Unnecessary. Proportionality judgments are not necessary in case of every attack. In fact, two conditions have to be jointly fulfilled to necessitate a Proportionality analysis: (1) civilians or civilian objects have to be present close enough to the attack's military objective to make collateral damage a possibility and (2) precautionary measures have to fail to eliminate the possibility of collateral damage. As noticed by Homayounnejad, non-combatants other than civilians are not covered by the protections of Proportionality (2019, 240), at least not as far as the law is concerned (Amoroso, 2020, 79). Ethics can, of course, be more restrictive than law, and there is good ground to believe non-civilian and yet protected persons and objects, such as POW or military hospitals, should be included in Proportionality calculations as well (Dinstein, 2016, 154 – 55). However, soldiers who have become non-combatants through wounds, incapacitation, or individual surrender cannot be included so long as they are in the midst of comrades who continue to fight, as this would entail their side unduly benefiting militarily from the protected status being granted them.

Given that both (1) and (2) have to be fulfilled for the Proportionality analysis to be necessary, it is quite obvious that AWS will not have to engage with Proportionality questions in quite a few combat scenarios (Amoroso, 2020, 91; Homayounnejad, 2019, 245-6; Van den Boogaard, 2015, 262). Lack of civilian presence should make engaging most naval and almost all aerial targets unproblematic. Crucially, it is precisely naval and especially aerial supremacy that today's military planners are most interested in. Civilians would also be absent from a not insignificant portion of land environments, such as the polar regions, jungles and deserts, and from restricted or fortified zones.

Even in circumstances where civilian presence should be expected, AWS' increased potential for taking precautionary measures should on many occasions enable them to eliminate or significantly reduce CDE. There are two reasons for which such an increase in ability is to be expected of AWS. First, as machines, they should simply be more precise than human combatants, lacking certain physical limitations and not being subject to psychological stresses of combat. Proportionality analysis is usually necessary when dealing with weapons of limited precision and/or large area of impact; we rarely have to ask whether it is proportionate for a sniper to take a shot. Yet as AWS would frequently be much closer to the latter than to dumb- or even precision-guided munitions, some models could simply be made technologically incapable of causing excessive collateral damage, with all civilian casualties attributable to such platforms being due to either Distinction errors or hardware malfunctions. Secondly, AWS would be expandable, and so could "shoot second" or sacrifice themselves in order to limit collateral damage in ways which would be morally, psychologically, and politically impossible for and with human combatants. It is thus not unreasonable that an improvement, and perhaps even a qualitative jump in ability to take precautions in combat would follow from mass introduction of AWS, just as it followed from the mass introduction of unmanned platforms. Finally, even in complex environments, human commanders may delineate spatio-temporal zones in which civilians would be absent or in which precautionary measures alone would assure no harm to civilians and civilian objects, and use AWS exclusively in such zones (Crootof, 2015, 1878).

[13] When posed in this way the question implies that AWS would not merely follow strictly a set of pre-programmed instructions, but that they would somehow learn to make the judgments, presumably from human-fed examples. Acting in this way, AWS would merely be imitating moral reasoning rather than actually performing it (Foy, 2014, 61). Ronald Arkin entertained the possibility of having AWS learn to imitate and extrapolate human proportionality judgments (2009, 47 – 8), as did Jeroen Van den Boogaard (2015, 277). Inference from human demonstrations and feedback has since proven to be a very promising approach to other problems in robotics (Christian, 2020, 253 – 73). However, like all machine learning approaches, this one would also depend on the quality and variety of initial examples provided (Amoroso, 2020, 93–96). It is hardly certain that at this point military ethicists and lawyers could provide a good enough database of properly solved cases. More importantly, one is to be wary of introducing machine-learning approaches to elements that may be handled via direct instructions. As Van den Boogaard puts it, "autonomous weapons systems must be prevented from making decisions on an operational or strategic level without the input of the relevant human military commander because at that level, the equation must also include other factors to determine the military advantage sought, including both strategic and political factors" (2015, 277). Machine learning these aspects from a limited database of examples does not seem feasible.

The precautionary measures an AWS may be able to implement – something to be empirically verified for each design and accompanying set of tactics – will ultimately determine how often, or how rarely, AWS will need to engage in Proportionality analyses. Yet even if their capacity for precautionary actions remains merely at a human level, or even falls below it, they will still be capable of fighting compliantly in some selected land environments, in most naval ones, and in almost all aerial actions. That is, even if we solve the Proportionality conundrum by requiring all AWS to refrain from any action that may cause any non-negligible amount of collateral damage[14], they will still retain their potential for revolutionizing air and naval warfare, and some selected aspects of land warfare. This demonstrates that even such very conservative restrictions would not, as Amoroso suggest, make AWS useless and "pointlessly expensive piece of weaponry" (2020, 92). Moreover, with the use of sliding autonomy, no single piece of equipment would have to be an AWS and an AWS only. In fact, one can easily imagine a fighter jet being optionally-manned for complex ground attack missions, remotely controlled for some other missions and switched into partial or full autonomy in civilian-free environments.

Proportionality Judgments in Targeted Strikes. Extremely conservative targeting is one option for AWS users willing to comply with Proportionality, but not the only one. Moral decisions, including Proportionality judgments, are frequently undertaken at levels of military hierarchy that AWS will certainly not occupy. This is only logical, as „Military advantage anticipated and proportionality are more appropriately assessed at the operational or strategic level. At these levels, senior commanders are able to consider the larger operational picture and come to a more comprehensive judgment" (Homayounnejad, 2019, 236). This is certainly the case of the most high value objectives that become the goal of dedicated targeted strikes, and of complex objectives achievable only via complex, coordinated attacks. "It goes without saying that an attack carried out in a concerted manner in numerous places can only be judged in its entirety" (Sandoz et al., 1987, 685, paragraph 2218). As only these high-level commanders understand the true AMA of such objectives, only they are capable of correct Proportionality judgments in the case of such attacks.

This assertion has an important upshot. "Conversely, small-unit commanders and individual combatants often have a limited objective and are not briefed on its value within the overall scheme of attack, nor on its changing value as battlefield dynamics evolve. Accordingly, such personnel should avoid unilaterally halting an attack, except in the case of extreme or unconscionable collateral damage" (Homayounnejad, 2019, 236). If that is true of human pilots or low-level commanders, it would also be true of AWS. Some targets designated by higher-level commanders are so important that their value vis-à-vis collateral damage can only be estimated by those higher-ups, and only extremely violent swings in the amount of CDE can make the collateral damage excessive.

Yet if the Proportionality judgments involving most important targets can be made ahead of time by generals, there is no moral reason why the proportionality judgments regarding other less important targets should not be made by colonels or majors. As long as the AMA of striking a specific target is known and stable within the timeframe for which the attack is authorized, commanders at an appropriate level could determine the amount of collateral damage they consider excessive for that target and send AWS to strike it unless this amount would be unavoidably exceeded by such a strike (Thurner, 2012, 81) [15]. In such a case, the only element of the proportionality analysis performed by the AWS tasked with the mission would be CDE, of which the machines could in principle be capable of at some point, as discussed above.

Consequently, targeted strikes against specifically identified targets, conducted within a reasonable timeframe from the conclusion of proportionality analysis, could also be compliantly performed by AWS, as all the moral and qualitative work inherent in assessing AMA and assigning it an acceptable level of collateral damage would be performed by human commanders and their staffs, with the machine itself only estimating collateral damage and comparing that estimate to a pre-programmed threshold. An attack conducted in this way would not be at all different from the targeted strikes being routinely conducted today. It could in fact benefit from the more accurate computer CDE techniques and the lack of bias in comparing the CDE to the pre-programmed threshold value allowed[16]. The only difference

[14] As civilian objects are defined negatively as all non-military objects, any attack may be expected to cause some amount of collateral damage to the environment, and, in inhabited areas, to private property. While killing or injuring a human being, or rendering a structure uninhabitable, is certainly to be taken into account, felling a single tree in a dense forest or creating a pothole in a country road does not seem to rise to the threshold.

[15] The AMA of attacking a given target may of course change greatly in a relatively short span of time. In that case, the AWS programmed with instructions no longer reflecting reality may effect a disproportionate strike if they cannot be reached with new instructions in time. Yet this situation would be no different from human pilots or submariners becoming unreachable in a communications-challenged environment and acting on obsolete instructions, and no more criminal than such an incident.

[16] It is easy to imagine a human pilot, charged with executing a targeted strike with a pre-set level of acceptable collateral damage, trying to consciously or unconsciously tweak his collateral damage estimation so that the strike can in fact go through. After all, the pilot will probably understand the military benefit of the strike to some extent, although not nearly as well as his commanders, and wish to have it

would consist of a human pilot being obliged to follow his commanders' proportionality conditions, and so used as a tool, being replaced by another, mechanical tool – and this is not a morally relevant difference.

It is important to stress the limited timeframe in which targeted strikes of this sort would have to take place in order for the proportionality assessment, or rather the AMA it is based on, to remain fresh enough. A human pilot will usually execute a targeted strike within hours, perhaps within a day, of receiving his instructions from a commander responsible for pre-making a proportionality judgment. An AWS would have to stick to a similar timeframe, making the military situation highly unlikely to change substantially enough for the proportionality judgment to become outdated. The character of the target being struck will dictate the "expiry date" of the judgment. It is perfectly reasonable to assume that a rail link important to enemy logistics will remain equally important a day later, but a civilian building may cease to be the linchpin of enemy defenses within an hour.

Proportionality Judgments in Attritional Warfare. As just discussed, targeted strikes by AWS would be aimed at specific objectives hand-picked by human commanders. If only strikes of this type were allowed, this would radically limit AWS' usefulness against striking targets of opportunity outside of civilian-scarce environments. However, I will now argue that in some circumstances commanders may be able to pre-make proportionality judgments not only about specific targets, but about certain classes of targets. This would be true of counter-platform strikes in attritional warfare, that is, strikes of opportunity against autonomously detected military hardware the minimal military value of which remains constant throughout a conflict. The paradigmatic situation of this sort would involve an AWS being sent to detect and destroy enemy air-defense platforms, tanks, or artillery pieces within a particular sector and within a particular time-frame.

Compliance with Proportionality on such missions may be approached in two ways. First, extremely conservative targeting rules may be imposed, prohibiting the AWS from carrying out any attack that would result in non-negligible collateral damage. This approach could be coupled with the platform attempting to dial-in, or communicate with a human commander in order to have him asses collateral damage. This would be the best approach whenever communications superiority over the enemy could be had. Moreover, if this level of communications could be reached, the AWS could also be constantly updated on the battle staff's estimate of their targets' AMA (Van den Boogaard, 2015, 275-6). However, if this was the case, why communicate with the platform via orders and updates and not just take in-the-loop remote control of it? Perhaps if only small data packets could be transferred due to bandwidth limitations or enemy interference, remote control would not be possible while limited information exchange would. In such cases the machine could transfer its CDE to its commander, have her make the proportionality judgment and then act on her instructions, just as with a targeted strike.

This, however, may not always be possible due to communications problems or the pace of the fighting, two chief reasons for which AWS would need to be deployed in the first place. The other solution would be to have "human-made decisions that are normally carried out in real-time (…) replaced with (or supplemented by) more general programmatic instructions that are fed into the machine's software in advance. This effectively means that individual decisions on the use of lethal force are substituted by broader policy-like choices, which are applicable to the range of situations matching the pre-programmed parameters" (Homayounnejad, 2019, 111). In short, the AMA of destroying a tank at a given sector of the front would be fixed, and periodically updated, by commanders and their battle staffs, and weighted against possible amounts of collateral damage to render a threshold of excessiveness (Scharre, 2018, 257). This would replicate the process to be engaged in in the case of targeted strikes, yet instead of a specific target a member of a class of targets would be assessed. To put it bluntly, commanders would decide every few days how many civilian lives are worth sacrificing in order to destroy a generic target such as a tank at that specific sector.

I follow William Boothby in believing that "generic military advantage to be anticipated from the attack of an object, say, that the algorithm software is designed to identify, may be known in advance" (2014, 111). That is not to say that the entire military advantage is knowable – it obviously is not – but rather to say that the part of military advantage that is inherent in imposing attrition on the enemy force by destroying a vital piece of military hardware will obtain in almost any circumstances. Destroying an enemy tank may be more or less advantageous, depending on the scenario – but it usually is advantageous at least to a certain extent, and this extent is substantial enough to justify causing a non-zero amount of collateral harm. This is untrue only with regards to an enemy force that is about to surrender, either locally or generally, as in forfeiting the war. Yet if such a surrender is about to happen, then this is either knowable to an AWS' commander, who may cease employing AWS upon learning this, or it is not known to her, in which case she or her AWS cannot be required to act on this knowledge. It is of course possible that the commander will learn of the pending surrender but will not manage to contact the AWS

realized. AWS CDE software would not, obviously, have desires of that or any other sort, eliminating this particular source of bias. This, however, should not be understood as implying the stronger claim that AWS would be free of bias in general.

in time; but this would be no different than any other case of inability to communicate some crucial piece of knowledge down the chain-of-command, which routinely happened in communications-denied frontline environments without creating ethical problems. An AWS that did not learn of the enemy surrender in time would be just like a human combatant that did not learn of it[17].

Another type of case in which generic military advantage may not gained by destroying high-value platforms is in situations involving abandoned or broken-down equipment. If an AWS failed to recognize inoperable platforms as such, which is likely, especially at the current technological stage, it could strike needlessly, perhaps even denying its own side a valuable capture. Such a needless strike would of course fail to justify collateral harm. The AWS' commander would have to take into account its inability to recognize abandoned equipment and consider how likely it was that such would be struck needlessly and possibly with collateral harm. However, humans, especially those having to make split-second decisions in combat, are also likely to make such mistakes, and mistakes these would be. AWS cannot be required to be entirely free of error, any more than any other system can be, and their commanders simply need to take the specific propensity for certain errors into account. And striking an abandoned vehicle would still have a smaller expected value of denying the enemy its recovery, which might be as likely as capture by friendly troops.

Except for the scenarios outlined in the preceding paragraph, a substantial military advantage is always attached to destroying certain enemy platforms. Usually, this generic advantage is coupled with the additional tactical advantage of destroying a piece of hardware in some particular tactical situation. However, AWS would not be able to discern that more specific advantage, and so that value ought to be assumed to be zero. Thus, AWS performing counter-materiel strikes would routinely underestimate the AMA of their strikes, making them *more conservative in targeting* than a manned or remotely controlled equivalent would be. On the model I am proposing, AWS' inabilities would be mitigated by a more restrictive use policy.

The 'price tag' solution is, however, a proposition that has been rejected by some authors writing on the topic. Let us therefore engage with their objections. Marcus Wagner writes that "the value of destroying a radar installation at a particular moment is impossible to determine *a priori*" (2014, 1397). If this means it is impossible to determine the value of this particular piece of equipment outside the context of a particular war or even of a particular engagement, he is certainly correct. However, were Wagner to mean that it is impossible to determine such a value an hour, or several hours, before the attack takes place, this would no longer be correct for many cases. The length of a timeframe within which such estimates would remain sufficiently fresh would vary from engagement to engagement, and commanders would definitely have to adjust it to the circumstances ruling at the time, but surely in each case there is a timeframe prior to attack in which such a determination is possible. If not, attacks with weapons such as artillery or torpedoes, which feature a delay between launch and impact, would not be possible. Therefore I believe that Wagner's reservations are directed at a much cruder process than the one I am proposing.

Amoroso also rejects such a concept, writing that

> (…) the pre-determination of acceptable collateral damage should not be conducive to a trivialization of proportionality assessments, by turning it into a merely quantitative analysis. This means that human operators cannot limit themselves to putting a "price tag", in terms of civilian harm, on each set of targets (e.g., by valuing the destruction of an enemy tank according to the number of permissible civilian deaths (2020, 91)

In support of his stance, Amoroso puts forward three different statements. First, he cites Wagner's reservations, which I have already addressed. Second, he quotes Louise Doswald-Beck, saying that "it is impossible to state that a factory is worth x civilians" (ibid., 92, footnote 235). As with Wagner's, Doswald-Beck's statement can be interpreted in two ways. If it is to mean that it is impossible to arrive at a precise number, say, 241, then it is clearly correct. As discussed above, such precision is unavailable for proportionality judgments. However, if the statement is interpreted to mean that for any given target it is impossible to uncontroversially state that a certain number of collateral deaths would clearly be excessive, while a certain lower number would clearly not be, then believing such a statement would undermine the entire enterprise of assessing proportionality. If this cannot be done, what exactly can be done, and what is routinely being done, by military commanders, lawyers, judges, and ethicists? I believe reasonable persons can agree that in an overwhelming number of possible scenarios, killing one hundred civilians collaterally while destroying a tank is excessive. I also believe they would agree that in most circumstances killing one civilian collaterally to take out a tank is not excessive. If the purpose of Doswald-Beck's statement would be to undermine the very possibility of making such broad and uncontroversial judgments, then it

[17] An even rarer case of an enemy unit that became permanently isolated geographically, for example isolated on an insignificant island, and so effectively rendered *hors do combat*, is analogous to the about-to-surrender scenario.

should be rejected as clearly wrong (though I do not believe this would be the correct interpretation).

Amoroso's third reason for rejecting the "price tag" approach is that it seemingly precludes one from appropriately making a diligent CDE assessment, which would have to include taking into account "the relative weight of the categories of civilians and civilian objects that may incidentally be affected, as well as reverberating effects" (2020, 92). This statement is influenced by Amoroso's belief, voiced earlier in his book, that the lives of some sub-groups of civilians, such as children, are to be considered more valuable than other civilian lives. This belief is, however, to be rejected, as differentiating between the value of various civilians' lives is much more likely to lead one to view certain groups of non-combatants as targetable than to more careful triage of inevitable casualties. As for assigning value to various classes of civilian objects and to a finite set of key infrastructure installations, destruction of which may lead to reverberating effects, this not only can but has to be done by battle staffs prior to attacks. Individual pilots or tank commanders may not be able to tell a water purification station from a random building, and so are fed this data by much more knowledgeable staffs using maps, open-source intelligence, analysis of human geography, etc. Computer-assisted or computer-performed CDE would offer much speedier and reliable access to such data, and so would actually help in anticipating reverberating effects.

To address another persistent objection, there is nothing inhuman in relying on an algorithm drafted beforehand by persons of proven moral knowledge and sensitivity. Such a solution is not unheard of in other areas of ethics, most notably in bioethics. It is used in emergency triage situations (occurring most often in military medicine), in decisions regarding the pursuit of research and state financing of certain medical and pharmaceutical interventions rather than others, but also in charitable work on global health and animal welfare and, ultimately, in any area of human activity concerned with providing for basic human needs in the circumstances of resource scarcity. Few would argue that the pattern of COVID vaccine distribution should be an intuitive call of a local physician, rather than a result of bioethicists, doctors, and health management specialists having carefully crafted a specific, detailed formula for the assignment of priority. Similarly, a given unit's rules of engagement could feature an algorithm for determining an upper bound on collateral damage of various kinds, subject to revisions and corrections with the progress of events. I do not see why robots would have to be inferior to humans in the implementation of such an algorithm, yet I see multiple reasons why they would be much superior in this respect (Van den Boogaard 2015, 269 – 71).

Critics might say that this dehumanizes the process and strips it of much needed sophistication, and I am prepared to grant this point in cases of operational or strategic level proportionality judgments. This is why I believe proportionality analysis for targeted strikes should be performed individually for each target. Yet even on the level when not only strategic but political considerations are involved (ibid., 275 – 77) a measure of consistency is surely desired. Nor are any of the decisions taken at this level truly historically unique, at least not all the time. Killing Bin Laden was much like capturing Saddam Hussein; the decision to liberate Mosul was much like the decision to liberate Raqqa.

But even if these case were all genuinely unique, these are decisions taken at the highest military and political levels, nothing that any AI agents will (hopefully) engage in anytime soon. What compliance with Proportionality entails in attritional warfare is making tactical decisions, decisions about taking pieces, not about winning chess games. And such decisions, tens of thousands of which are made in every war, are essentially and necessarily similar to many other such decisions. How many lives is a tank worth on this area of the front this month? What is the worth of a supply truck? Targeting of this kind is repeatable and may be subject to moral aggregation. Such aggregation is already routinely undertaken by human combatants.

One final objection would consist of pointing out that the mechanisms of compliance I just outlined, and their inability to work in certain combat scenarios, would limit the purported benefit of AWS[18]. The answer would depend on whether one is talking about their military or ethical benefits. As for their military usefulness, it would only be minimally limited by these constraints in air and naval combat, and would still remain quite substantial in conventional land warfare. The fact that ethical constraints limit a system's military usefulness may also be a military or strategic problem, but it is not necessarily an ethical one, at least not in this case. As for the purported ethical benefits of using AWS, I believe these would be realized chiefly, though not exclusively, via their propensity to employ more advanced precautionary measures and by further limiting human exposure to combat. The proposed restrictions aimed at Proportionality compliance would not impact these, while forcing commanders to take a more cautious approach to causing collateral harm. I view the latter change as positive.

Introducing the "price tag" approach to counter-materiel attritional strikes would not introduce crudeness into the analysis performed at this level, it would reduce it while simultaneously acknowledging the extent to which it cannot be reduced. This acknowledgment could then become a springboard towards greater sophistication. After all, AWS

[18] I thank an anonymous Reviewer for bringing this objection to my attention.

would be able to weigh many more factors, and access many more memory resources or databases than humans could in the same time. More and more sophisticated criteria could thus be tried and introduced. "In the case of tactical level combat, changes in military advantage anticipated can even be pre-programmed in accordance with known or anticipated changes on the battlefield, such as how many other targets have been destroyed or neutralised; or whether tanks appear individually or in concentrations" (Homayounnejad, 2019, 247). Returning to an earlier example, a much richer and much more up-to-date picture of the human geography of a particular area would be available to a machine, enabling more sophisticated CDE much more attuned to potential reverberating effects[19].

Conclusion

There is no question that for the foreseeable future, AWS will not reliably perform proportionality analyses in all their complexity, as this requires human-level intelligence and depth of understanding. However, AWS could and would avoid Proportionality conundrums in many cases by employing precautionary measures to which they may be uniquely predisposed, or simply in virtue of being limited to operations in civilian-free environments. However, whenever issues of proportionality would actually arise, they would either have to refrain from launching any attacks or have humans pre-assess the proportionality of various sorts of attacks.

This latter solution has its downsides, but is already being employed in relation to human combatants carrying out targeted strikes, an area where AWS may be used instead of humans. In the context of attritional strikes against enemy combat platforms, having competent lawyers, ethicists, and commanders attach proportionality "price tags" to various target classes within well-defined spatio-temporal constraints would also likely suffice to fulfill the Principle's requirements[20]. AWS, and the commanders issuing their instructions, would have to act more conservatively than a human combatant engaging in real-time proportionality judgments. This would somewhat limit, though by no means negate, their military usefulness, but is not an ethical problem in itself. Indeed, greater caution in causing collateral harm, necessitated by AWS' limitations, should be welcomed by humanitarians wary of more permissive uses of force.

The conclusion that follows is that AWS (and their commanders) could comply with the Principle of Proportionality in quite a number of combat environments. This has obvious implications for their general ability to comply with the ethical principles of *Ius in Bello* as a whole. In fact, it removes the strongest argument of those who claim such compliance is unachievable. This is not to say that this alone should settle our judgment on the possibility of AWS complying with these rules. Indeed, compliance would have to be proven through empirical tests to ensure that these weapons are reliable and may be permissibly used. I also do not claim to have settled the larger debate on the AWS' moral permissibility, which hinges on their ability to clear several other ethical thresholds. However, as this debate can only be resolved one issue at a time, I hope this article, and the discussion it may spark, will ultimately contribute to this more general debate as well.

As for the more general issues of *Ius in Bello* and war ethics, the emergence of AWS brings to the fore the urgency of unanswered questions regarding the nature and application of the Principle of Proportionality. As such, it may be a hidden blessing; "[it] may have the advantage of obliging States to agree on how exactly proportionality must be calculated and also on which parameters influence this calculation" (Sassoli, 2014, 331).

Funding This research was funded in whole by National Science Centre, Poland grant number 2022/44/C/HS1/00051.

References

Amoroso, D. (2020). *Autonomous Weapons Systems and International Law: a study on human-machine interactions in ethically and legally sensitive domains*. Nomos.

Arkin, R. (2009). *Governing lethal behavior in autonomous robots*. CRC Press.

Asaro, P. (2012). On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal

[19] The AWS itself would obviously have no understanding of the reverberating effects, but its programmers would, and so they could code in the fact that any damage to critical infrastructure entails a large number of additional casualties, or, even more restrictively, that any such damage is simply disproportional regardless of other factors.

[20] I do not wish to imply that AWS could not fulfill the requirements of Proportionality in some other cases, nor that they could not – although such compliance would be more difficult in any other case. My claims are restricted to the scenarios I described.

decision-making. *International Review of the Red Cross*, *94*(886), 687–709. https://peterasaro.org/writing/Asaro%20IRRC.pdf.

Boothby, W. H. (2014). *Conflict law: the influence of new weapons technology, human rights and emerging actors*. Springer.

Brenneke, M. (2018). Lethal Autonomous Weapon Systems and their compatibility with International Humanitarian Law: a primer on the debate. In T. D. Gill, R. Geiß, H. Krieger, & C. Paulussen (Eds.), *Yearbook of International Humanitarian Law* (21 vol., pp. 59–98). TMC Asser Press.

Christian, B. (2020). *The Alignment Problem: machine learning and human values*. Atlantic Books.

Crootof, R. (2015). The killer robots are here: legal and policy implications. *Cardozo Legal Review*, *36*, 1837–1915. https://ssrn.com/abstract=2534567.

Dinstein, Y. (2016). The Conduct of Hostilities under the Law of International Armed Conflict, Third Edition. Cambridge University Press.

Foy, J. (2014). Autonomous weapons systems: taking the human out of international humanitarian law. *Dalhousie Journal of Legal Studies*, *23*, 47–70. https://doi.org/10.2139/ssrn.2290995.

Gardam, J. G. (1993). Proportionality and force in international law. *The American Journal of International Law*, *87.3*, 391–413. https://doi.org/10.2307/2203645.

Haque, A. A. (2017). *Law and Morality at War*. Oxford University Press.

Homayounnejad, M. (2019). Lethal Autonomous Weapon Systems Under the Law of Armed Conflict. Doctoral Dissertation, King's College London. https://kclpure.kcl.ac.uk/portal/files/110384075/2019_Homayounnejad_Maziar_0222601_ethesis.pdf.

Hurka, T. (2005). Proportionality in the morality of War. *Philosophy & Public Affairs*, *33*(1), 34–66. http://www.jstor.org/stable/3557942.

Kowalczewska, K. (2021). *Sztuczna inteligencja na wojnie: Perspektywa MPHKZ. Przypadek autonomicznych systemów śmiercionośnej broni*. Wydawnictwo Naukowe Scholar.

McMahan, J. (2009). *Killing in war*. Oxford University Press.

Orend, B. (2013). *The morality of war*. Second Edition. Broadview Press.

Sandoz, Y., Swinarski, C., & Zimmermann, B. (Eds.). (1987). *Commentary on the additional protocols: of 8 June 1977 to the Geneva Conventions of 12 August 1949*. Martinus Nijhoff Publishers.

Sassoli, M. (2014). Autonomous weapons and international humanitarian law: advantages, open technical questions and legal issues to be clarified. *International Law Studies/Naval War College*, *90*, 308–340. https://archive-ouverte.unige.ch/unige:37976.

Scharre, P. (2018). *Army of none: autonomous weapons and the future of war*. W.W. Norton & Company.

Solis, G. D. (2016). *The Law of Armed Conflict: International Humanitarian Law in War Second Edition*. Cambridge University Press.

Thurner, J. S. (2012). No One at Controls: Legal Implications of Fully Autonomous Targeting. Joint Force Quarterly 67 4th Quarter, 77–84. https://ssrn.com/abstract=2296346.

Van Den Boogaard, J. (2015). Proportionality and Autonomous Weapons Systems. *Journal of International Humanitarian Legal Studies*, *6*(2), 247–283. https://doi.org/10.1163/18781527-00602007.

Wagner, M. (2014). The dehumanization of international humanitarian law: legal, ethical, and political implications of autonomous weapon systems. *Vanderbilt Journal of Transnational Law, 47*, 1371–1424. https://ssrn.com/abstract=2541628.

Wright, J. D. (2012). 'Excessive' ambiguity: analysing and refining the proportionality standard. *International Review of the Red Cross*, *94*, 886, 819–854. https://doi.org/10.1017/S1816383113000143.