Jus in bello necessity, the requirement of minimal force, and autonomous weapon systems

Alexander Blanchard,[1] Mariarosaria Taddeo[1,2]

[1] The Alan Turing Institute, London, UK

[2] Oxford Internet Institute, University of Oxford, UK

Abstract

In this article we focus on the *jus in bello* principle of necessity for guiding the use of autonomous weapon systems (AWS). We begin our analysis with an account of the principle of necessity as it entails the requirement of minimal force in Just War Theory, before highlighting the absence of this principle in existing work on the permissible uses of AWS. Overlooking this principle means discounting the obligations that combatants have towards one another in times of war. We argue that the requirement of minimal force is an important requirement for considering ethical uses of force. In particular, we distinguish between lethal and non-lethal purpose of use and introduce the prospect of non-lethal purpose of use of AWS and review a number of challenges which AWS pose with respect to their non-lethal use. The challenges arise where AWS generate unpredictable outcomes impinging upon the situational awareness required of combatants to ensure that their actions meet the requirement of minimal force. We conclude with a call for further research on the ethical implications of non-lethal uses of AWS as a necessary, preliminary step to assess the moral permissibility of AWS.

1

of Defence (MOD) sponsored research and is released for informational purposes only. The contents of this paper should not be interpreted as representing the views of the UK MOD, nor should it be assumed that they reflect any current or future UK MOD policy. The information contained in this paper cannot supersede any statutory or contractual requirements or liabilities and is offered without prejudice or commitment.

1. **Introduction**

The operational advantages offered by autonomy in weapon systems looks set to proliferate its future use. For example, in 2021 the UN Security Council reported the first use of Autonomous Weapon Systems (AWS) in the spring of 2020 as part of the on-going conflict in Libya. The forces of the Government of National Accord "hunted down and remotely engaged" a retreating column of Libyan National Army Forces using AWS and loitering munitions. These systems, as the UN Security Council reports,

> "… were programmed to attack targets without requiring data connectivity between the operator and the munition: in effect, a true 'fire, forget and find' capability" (Choudhury L et al. 2021).

There remain difficulties in defining and thus identifying AWS. But if this was indeed an example of the use of AWS, it illustrates the urgency of the task of addressing ethical and legal questions about their use. As a report by the US Air Force recognises,

> "Authorizing a machine to make lethal combat decisions is contingent upon political military leaders resolving legal and ethical questions" (USAF 2009, 41).

These ethical and legal questions relating to the use of AWS are wide-ranging and profound: AWS raise questions about the locus of responsibility for their deployment, particularly in cases where these systems malfunction (Matthias 2004; Sparrow 2007; Gerdes 2018; Taddeo and Blanchard Forthcoming); about whether the use of algorithmic targeting decisions potentially infringes the right against arbitrary execution (Heyns 2013; 2017); and about whether their use violates the principle of dignity in death.

A central point of debate, the one to which this article speaks, is whether AWS can be used in line with the Just War Theory principles underpinning International Humanitarian Law (IHL). The principles of IHL seek to regulate warfare and limit the harmful effects of armed conflict. If AWS cannot be used in a way which complies with the principles contained in IHL – necessity, proportionality, and discrimination – then there are grounds for an international ban on AWS. Scholarly debate about the compliance of AWS with IHL has been voluminous. Some commentators have argued that AWS will comply with IHL in a way that surpasses human combatants (R. C. Arkin 2010; R. Arkin 2018). This relates especially to the principle of distinction which upholds the

3

protection of civilians in times of conflict. Advocates argue that since AWS lack the emotions that are said to 'cloud' the judgements of human combatants, AWS will have lower rates of target misidentification (Marchant et al. 2011).

At the same time, a number of commentators have argued that, for the foreseeable future, AWS will lack the contextual awareness required to distinguish between civilians and military targets (N. Sharkey 2008; 2010; N. E. Sharkey 2012; ICRC 2021). This work has highlighted key areas where ethical challenges to do with the deployment of AWS may arise and has helped to define the parameters of debate. In light of this, it is puzzling that the principle of necessity has been largely overlooked in discussions of the regulation of AWS under IHL. As we describe below, whilst the principles of distinction and proportionality have been afforded considerable weighting in this debate, necessity is either discounted as an inconsequential contribution to ethical thought on warfare or as inapplicable to AWS (see for instance: Grut 2013). This is a limitation on addressing the problems raised by AWS for two reasons. The first is that necessity will be an element in determining the legality of AWS under IHL. Second, understanding whether the use of AWS complies with IHL will require recourse to Just War Theory, the ethical system which underpins IHL. Under Just War Theory, the principles of proportionality, distinction, and necessity inform one another. A complete assessment of permissible and impermissible acts of war cannot be accomplished without considering the three principles together. In this article we assess the contribution that necessity makes to these ethical assessments of AWS.

Our aim is twofold. Firstly, it is to show that once properly understood necessity provides an important principle for informing ethical use of AWS. For this alone necessity is deserving of greater attention in work on the ethical challenges posed by AWS. Secondly, we aim to set out the limits at which AWS can be used within the parameters dictated by necessity. In doing so, we argue that the use of AWS will present challenges in meeting the requirement of minimal force and the principle of necessity more broadly.

We will clarify the nature of necessity in section 2, particularly in its relation to combatants. Frequently the principle of necessity is confused with the principle of proportionality. Disentangling these two principles is important for clarifying their application to AWS and to identify the requirement of minimal force. In section 3, we explore the principle of necessity with respect to AWS, before considering the use of AWS in light of the requirement of minimal force. We argue that the unpredictability of these systems challenges their use in line with this requirement.

Before beginning the analysis, we must define its conceptual space in two important ways. Firstly, we define an AWS as,

> "an artificial agent which, at the very minimum, is able to change its own internal states to achieve a given goal, or set of goals, within its dynamic operating environment and without the direct intervention of another agent and may also be endowed with some abilities for changing its own transition rules[1] without the intervention of another agent, and which is deployed with the purpose of exerting kinetic force against a physical entity (whether an object or a human being) and to this end is able to identify, select and attack the target without the intervention of another agent is an AWS. Once deployed, AWS can be operated with or without some forms of human control (in, on or out the loop)" (Taddeo and Blanchard Forthcoming, 18)

Under this definition, lethal and non-lethal AWS are each a specific subset of AWS as defined by the purpose of use of AWS. A lethal AWS is used with the goal of exerting lethal force – i.e. resulting in death – against human beings. A non-lethal AWS is used with the purpose, as with non-lethal weapons generally, of incapacitating human beings "without causing death or permanent injury" (Davison 2009, 1). A non-lethal AWS may also be a system used for the purpose destroying for material destruction or disabling equipment.

Second, in discussing the ethics of AWS we are concerned with the ethics of intended use. The debate on implications of AWS for IHL is often at cross-purposes about whether the ethical implications of AWS have to do with its inherent characteristics – e.g. autonomy – or with the manner of its use – e.g. in urbanized settings. As Grut notes,

> "a weapon may present problems for or be contrary to IHL because of the manner in which it is used; or more rarely, the weapon itself may be inherently problematic" (Grut 2013, 9).

While AWS pose ethical problems concerning their entire lifecycle, from procurement and design to deployment, the analysis of these problems is outside the scope of this article. This is because, the underpinning question of the research presented in this article is whether Just War Theory can shed any light on the use of AWS. Just War Theory is not concerned with technical characteristics of the means used as much as it is with the intended purpose of use. Addressing the overall question of the moral permissibility of AWS requires considering different aspects, among which the technical

---

[1] Transition rules allow a Turing Machine, and more in general computational artefacts, to change its internal state when presented with a certain output following a series of if/then or logic doors as defined by its algorithm.

characteristics of these weapons are a central one along with the respect of the principles of Just War Theory. In this article, we tackle a specific aspect – whether non-lethal uses of AWS can respect the principle of *in bello* necessity, and leave the analysis of other aspects to further steps of our work.

2. The Principle of Necessity

The principle of necessity allows for measures which are necessary to accomplish a legitimate military objective. As such the principle entails both a permission and a constraint. The permission states that, whatever means are necessary for attaining a legitimate end, those means are justified.[2] As Lackey has formulated it,

> "the principle [of necessity] does not say that whatever is necessary is permissible, but that everything permissible must be necessary" (Lackey quoted in: Ohlin and May 2016, 77).

The constraint states that, whatever means are unnecessary for attaining the legitimate end, those means are unjustified. If a particular action is considered necessary to accomplish a given military end, then that entails that there are actions which are unnecessary for accomplishing the same end. Actions which are unnecessary for accomplishing a given end must be avoided (Matsumoto 2020). Necessity thereby determines the justified means for pursuing a legitimate end. In this way, the principle of necessity best corresponds to the requirement for 'justified use' of artificial intelligence (AI) proposed for domain-agnostic uses of AI. 'Justified use' cautions against overuse – thereby generating risks – or underuse – thereby generating opportunity costs (Floridi et al. 2020; Taddeo et al. 2021).

The constraint entailed by the principle of necessity introduces the requirement of minimal force, which states that combatants are obliged to use the minimal amount of force to attain a military end. This requirement is crucial for thinking about the use of AWS, both for lethal and non-lethal uses. To understand the requirement of minimal force it is important to clarify how the principle of necessity both differs from, and interacts with, the principles of proportionality and discrimination.

Often the principle of necessity is confused for the principle of proportionality and this is misleading. Consider the following,

> "For autonomous weapon systems to comply with the military necessity rule, the force used will only be deemed to be necessary if it is proportional, targeted on a military objective, and in line with the dictates of humanity" (Chengeta 2016, 131).

---

[2] Here we mean this only with respect to the principle of necessity considered in isolation. Whatever is permitted by the necessity principle must also, to be permissible overall, be permitted by the principles of distinction and proportionality. We consider this point presently.

6

Here necessity is made to supervene upon proportionality. It is important to hold them as separate because they are distinct calculations. First, an act of war can be proportionate because its costs[3] are tolerable relative to its benefits, but it can at the same time be unnecessary because those benefits could have been achieved by less costly means (Hurka 2008, 128). Conversely, an act of war can be necessary in so far as it is the least destructive – perhaps the only – means for attaining a given objective but fail to be proportionate, because the costs were not tolerable relative to the benefits. An AWS may achieve a proportionate balance between benefits and costs when used in a specific context, but that alone does not guarantee it is the *least* costly means. Likewise, an AWS employed as the least costly means does not mean that it brings a proportionate balance of benefits to costs. Being distinct principles, the principle of necessity permits a comparative calculation of the proportionality of conduct in war. As Hurka points out

> "…the proportionality condition considers the relevant benefits and harms of a war or act considered on its own, while the necessity condition compares the result of that calculation with the results of similar calculations for relevant alternatives, allowing a choice only when its balance of benefits to harms is better than that of any alternative" (Hurka 2008, 129).

A proportionality calculation made about one particular act of war, even if yielding a positive benefit, cannot be applied straightforwardly to alternative means because, even if the least destructive, alternative means "will not achieve all the same goods, or not all to the same degree, and sometimes they risk additional harms" (Hurka 2008, 129).

It is also important to note that the necessity principle and the proportionality principle speak to different sets of actors. The principle of necessity considers the harms that are done to combatants and whether such harms are necessary. The principle of proportionality speaks to the unintended but foreseeable harms associated with harms on non-combatants. In that way both principles are modified by the principle of discrimination (Ohlin and May 2016). McMahan has expressed this point by referring to the principle of necessity as 'internal to liability.' Harms inflicted on those who are liable to suffer them, i.e. combatants, have traditionally been assumed to have no role in determining proportionality (McMahan 2009, 19–23). If such harms were to play a role in assessing proportionality, then the resort to war would be ruled out if the expected harm to enemy combatants were to exceed the expected harm to one's own non-combatant population were the war not undertaken. The notion

---

[3] Costs are meant here in a wider sense and not just in a financial sense; e.g. economic, environmental, and human.

that proportionality calculations should include harm intentionally inflicted upon enemy combatants is one which no Just War theorist has been willing to embrace (McMahan 2009, 19).

It is the modification of necessity by the principle of distinction which defines the requirement of minimal force. The requirement of minimal force states that if there are ways to eliminate the threat posed by an enemy combatant, to achieve a given military objective, without killing that combatant, then it would be impermissible to kill that combatant. At the very least, restricting ourselves here to *in bello* considerations of individual acts of war, the act of war must be of a "duration, scale, and intensity" which is the minimum necessary to meet the given threat (United Nations General Assembly 2004, 58). There is no agreed measure for the application of this requirement as it relies on contextual considerations. However, the requirement itself is often found in different formulations across the literature on the ethics of war:

> "In the conduct of war, it is morally obligatory to use the minimum force necessary to realize a military objective" (Lango 2010, 482).

> "…just war theory condemns intentional harm that is not necessary for the accomplishment of the mission, proposing that military forces should not employ disproportionate force against enemy combatants. Instead, combatants do retain some right not to be harmed when this harm does not do anything to further mission accomplishment" (Mayer 2015, 302).

> "…the immediate object is not to kill or even to injure any particular person, but to incapacitate or restrain him" (Childress cited in: Lango 2010, 484).

Given the appropriate context this could require the use of non-lethal means. Where non-lethal means can be used in meeting a given threat, there is a moral obligation to use them.

The principle of necessity is one of the lesser-understood principles of Just War Theory. This is despite its centrality to Just War Theory. Its comparative lack of explicit formulation speaks to the fact that it stands as a broader orientation towards war rather than being a precise doctrine. It represents a compromise between considerations of humanity and the legitimate need of a state to fight effectively in war (Schmitt 2009). This has also meant necessity is confused for expediency. This is under the lasting influence of the idea of *kriegsraison*, that a course of action is permitted because it compelled "the submission of the enemy with the least possible expenditure of time, life, and money"

(Walzer 1977, 144; Downey 1953, 253). The association of necessity with expediency has also continued under a popular interpretation of the concept of 'supreme emergency' found in the work of Walzer (1977, 251–68). According to this popular interpretation the notion of a supreme emergency provides a general license for subordinating ethics to military objectives.

It is important to stress here a limitation in Just War Theory with respect to the requirement of minimal force. The requirement of minimal force determines that it is morally obligatory to use non-lethal means to incapacitate an enemy combatant where lethal means would be unnecessary (in light of a legitimate military objective). Just War Theory therefore determines the point at which non-lethal means are required. However, it does not contain a general set of prescriptions for the use of non-lethal means in war. This is because Just War theory gives paradigmatic status to killing as "both the form of violence whose justification the theory needs to account for and as the type of threat that might justify a resort to war" thereby overlooking 'lesser' forms of violence (Finlay in: O'Driscoll et al. 2020, 19). Intuitively the status that Just War Theory gives to the paradigmatic status of killing represents an economy, for

> "…if we can decide what values or threats justify killing, then it seems likely we'll implicitly have covered all other – presumably lesser – values and threats. That way we don't have to elaborate explicitly a theory of punching, for instance, or wounding (or of smashing shop windows or burning cars). A theory that justifies killing already, implicitly, justifies these lesser harms" (Finlay in: O'Driscoll et al. 2020, 19).

The plausibility of this implicit assumption stems from the fact that warfare is marked as a distinct activity because of the centrality of killing. However, the conceptual apparatus of Just War Theory, built around killing, has left it inapplicable to lesser forms of harm. Indeed, in order to make more appropriate judgements about non-lethal violence we would,

> "…need to look much more closely at the deeper assumptions underlying just war theory and to theorise more fully the evil of violent means as such (as opposed to killing and war using lethal force) on the one hand, and the values that might be commensurable with it, on the other" (Finlay in: O'Driscoll et al. 2020, 19).

There is not scope in this article to elaborate these deeper underlying assumptions. Whilst Just War Theory provides a criterion – the necessity principle – for when force short of lethal force should be used, its failure to provide a theory for justifying non-lethal force presents an obstacle in using it to unpack the potential ethical implications of using AWS in such a way.

9

We note this point because it parallels the treatment of AWS in international debate on these systems whereby they have become known as 'Killer Robots'. This designation, first employed by The Campaign to Ban Killer Robots, has helped to highlight an important issue for the public, but its success in having done so, and the way that such talk has bled into academic philosophical discussion of AWS, has obscured potential problems with non-lethal uses of AWS (see for instance: Sparrow 2007; Krishnan 2009; Human Rights Watch 2012; Goose and Wareham 2016; Taylor 2020). The reason why this is problematic is because non-lethal uses of force might entail ethical problems which are not captured by discussions on *lethal* autonomous weapon systems (LAWS). This paradigmatic focus on killing tends to centre on the permissibility, or not, of taking life. It overlooks problems associated with non-lethal force to do with domination, coercion, or autonomy. These are problems which demand sustained attention. While it is not in the scope of this article to provide a comprehensive account of the ethics of non-lethal force when associated with autonomous systems; it is nevertheless important to raise a potential problem with respect to the use of AWS in a way that complies with the requirement of minimal force. This in turn raises questions about the potential for AWS to be used in a way which complies with the principle of necessity.

3. The Principle of Necessity and AWS

As noted at the outset, the permissibility of using AWS will in part depend on whether AWS can be used in a way which conforms with the principles of IHL. A good deal of scholarship has been devoted to the question of the capacity of AWS to meet the principles of distinction and proportionality, either in dedicated studies or broader reviews of the ethics of AWS and international law (see: Marchant et al. 2011; Grut 2013; Foy 2014; Roff 2015; van den Boogaard 2015; Beard 2018; Davison 2018; Winter 2018; 2020). But the principle of necessity goes either unconsidered or considered in only a perfunctory way. See for example (Grut 2013), whose analysis focuses on exploring the challenges of AWS to IHL, but only with respect to the principles of distinction and proportionality. Grut notes that the principle of distinction and the principle of proportionality are the "[t]wo bedrock principles of IHL." The principle of necessity goes unmentioned. Neither does Wagner's much-cited article (Wagner 2014) on the implications of AWS for IHL deal with the principle of necessity. The principle of necessity is mentioned, to be sure, but only in its counterpoise to the principle of humanity. As Wagner writes,

> "The tension between military necessity and humanity is one of the main characteristics of
> IHL. There is considerable disagreement as to where the balance should be struck on the

continuum of military necessity and humanity. There is also disagreement as to what degree extant circumstances – including advances in military technology, the acceptability of civilian casualties in the court of public opinion, and potentially more fundamental changes in the role of state sovereignty – should influence military decisions. These discussions are evidence of a development away from a military-centric approach towards one that increasingly takes humanitarian considerations into account" (Wagner 2014, 1387).

The growth of the importance of humanitarian considerations in warfare relative to necessity is an important consideration for the ethical use of AWS. However, Wagner does not detail how the line between necessity and humanitarian considerations should be drawn in the use of AWS.

Other commentators have sought to reject the principle of necessity as unhelpful in considering the ethical use of AWS. For instance, Foy writes that

"while the use of AWS engages the principles of military necessity and unnecessary suffering, these principles are engaged in a different way than distinction and proportionality" (Foy 2014, 54).

The reason that this is so, according to Foy, is that the principle of necessity is relatively unaffected by the employment of autonomy in weapon systems.

"The difficulties of adherence to the principles of distinction and proportionality naturally arise when humans are removed from the decisions-making loop. The principles of military necessity and unnecessary suffering are less affected by the removal of a human from the loop and more situation-specific" (Foy 2014, 54).

Schmitt and Thurner (2012) have also expressed scepticism about the applicability of the principle of necessity to the use of AWS. They write that attempts to prohibit the use of AWS on the basis of necessity mischaracterize military necessity. Their objection recalls, again, the status of necessity as a counterpoise to considerations of humanity rather than a rule in itself. As they write, necessity is not "a distinct rule of the law of armed conflict." Rather, it is "a foundational principle that undergirds the entire body of law" (Schmitt and Thurnher 2012, 258). Moreover, even granted that necessity were treated like an explicit rule, it alone would not make AWS unlawful.

"They would not be unlawful *per se* because there are clearly situations in which they are valuable militarily. In other words, autonomous weapons systems are not superfluous, if for no other reason than the fact that unlike manned systems they can attack the enemy without placing an operator at risk."

11

Schmitt and Thurner then go on to claim that acceptable uses of AWS can be exhaustively captured by the principles of proportionality and distinction, quite without the need to use necessity:

"As to prohibitions based on use, the requirement that military objectives yield some military advantage would make any separate condition for military necessity redundant. With regard to situations raising proportionality issues, any strike lacking military advantage but causing harm to civilians of civilian objects would violate the rule…the law of armed conflict already prohibits attacks on those who have surrendered or are otherwise hors de combat. Taking these observations together, the result is that military necessity has little or no independent valence when assessing the legality of autonomous weapon systems or their use" (Schmitt and Thurnher 2012, 258–59).

Indeed, given the distribution of risk and liability in war, distinction and proportionality will remain of central importance in such reflections. However, the assumption, as evidenced in Schmitt and Thurner, that proportionality and distinction alone are important in determining the ethical use of AWS is misconceived. The question here concerns the obligations which enemy combatants owe one another, a key question in assessing just behaviour in war waging. In discounting necessity there is an assumption that an enemy combatant, when neither surrendering nor *hors de combat*, is susceptible to any amount of force. That is, their liability to attack renders them fair game. As we saw above, however, because of necessity, liability to attack does not license an attack of any "duration, scale, and intensity." Rather, there is a requirement of minimal force. A combatant must use the minimal force required to incapacitate an enemy combatant in the pursuit of a military objective. At times this will require the use of non-lethal force. As Mayer states,

"just war theory, in its current state, deems it morally impermissible to employ lethal weapons against a combatant when non-lethal weapons can accomplish the mission just as effectively" (Mayer 2015, 302).

The requirement of minimal force is no less applicable to AWS than it is to other means of warfare. Therefore, discounting necessity means discounting the question of when it is appropriate to use AWS at the threshold below lethality. This will be the focus on our next section.

3.1.   Necessity and the Problem of Predictability

Here we want to raise a challenge that AWS poses to the minimal requirement of force, specifically when the requirement of minimal force calls for non-lethal uses of AWS. The challenge relates to the self-recognition that combatants have of their ability to attain a particular objective, and how the

12

recognition of this ability is important for determining the appropriate use of force. Setting out this challenge requires an account of both (a) intention as it pertains to the requirement of minimal force and (b) predictability as it pertains to AWS.

There can be *practical* difficulties in holding a distinction between lethal and non-lethal weapons. The lethal and non-lethal *uses* of a weapon can frequently be indistinguishable in their effects. Weapons used in a way meant to be non-lethal can frequently be lethal in effect; and weapons used in a way meant to be lethal may also produce non-lethal effects. As Enemark explains, "the difficulty essentially comes down to a potential disconnect between the *intention* behind the use of a weapon and the consequences thereof" (Enemark 2008, 201 - emphasis original). A combatant may have used a weapon only to incapacitate an enemy combatant, but the injuries sustained may have ultimately been lethal. As a 1972 report on non-lethal weapons by the US National Science Foundation explains:

> "All weapons…create some primary or secondary risk of death or permanent injury. The probable seriousness of their effects (their lethality) depends on a number of factors, not all of which are determined by their design. Weapons not intended to kill or create permanent injury, if used with an degree of regularity, would undoubtedly cause some deaths because of the physiological differences among those against whom they employed, physical malfunctioning, improper utilization, and other circumstances" (National Science Foundation cited in: Davison 2009, 1)

Similarly, the weapon used in a way that was meant to be lethal can, in the same way, have non-lethal effects. The indeterminacy of effects is exacerbated in war given the complexity of the combat environments and the inherent dangers the activity poses. Thus, assessing lethal or non-lethal use of a weapon requires an assessment of the intention of use of that weapon. As Ramsey has written,

> "the objective of combat is the incapacitation of a combatant from doing what he is doing because he is this particular combatant in this particular war; it is not the killing of a man because he is a man or because he is this particular man…the difference is to be found in the *intention*" (Paul Ramsey 2002, 397 - emphasis added).

Crucially, the intention required in necessity calculations is an intention supported by contextually bound judgements on the application of the necessity principle in the specific context. As Ohlin and May write,

> "Military necessity, as a form of practical necessity, is the employment of objective legal standards by a person who must judge at the moment from a significantly subjective

13

standpoint. Military necessity is a mixture of these objective and subjective factors, and cannot easily be reduced to merely one or the other perspective" (Ohlin and May 2016, 86).

Here the point is that military necessity is in part governed by the principle of success as pertaining to a given set of circumstances: if an action does not increase the likelihood of attaining a military objective, then that action cannot be deemed necessary to fulfil the objective. Success here is taken to be composed of two factors: luck and ability. Luck depends on whether circumstances are conducive to successfully attain the goal. Ability is the capacities of the human agent to attain a goal taking into consideration the circumstances of that goal. Ability therefore requires a recognition of one's ability to attain that goal, along with an ability to factor that recognition into calculations about success. Such assessments "are to be made from the first-person standpoint of the commander in a given situation" (Ohlin and May 2016, 85).

Here the difficulties arise in the use of AWS for non-lethal purposes. The understanding of success required by necessity is unattainable where the use of AWS leads to unpredictability of outcomes. Let us briefly focus on the limits of predictability of AI systems before continuing our analysis. The question of the predictability of AWS relates to broader discussions about the predictability of AI systems. As noted in (Taddeo and Blanchard Forthcoming), the predictability of AI is assessed by the consistency between past, present, and future outcomes. The predictability of an AWS relates both to the technical aspects of the system and to the context of use of that system. Regarding technical aspects, different models of AI have varying levels of predictability, and not all models are unpredictable. For instance, some AI models are comprised of decision trees rule-based decision-making which, in principle, leaves decisions-making process, and thus the outcomes of that process, known to either or both the designer and the operator of the system. Other models, like reinforcement learning using neural networks may result in a decision-making process which is opaque to both the designers and operators of the system, and whose outcomes change depending on the interactions of the system with the environment. It is this latter type of learning which raises problems for predictability. An AI-enabled system which can transform itself as it changes from each specific deployment to the next and can, thereby, acquire behaviours and generate outcomes which may not have been anticipated, or intended, prior to the point of use. As the UNIDIR reports,

"If a machine learning system can adjust itself in real time while executing a mission - a technique that is gaining favour as a means of continuously improving the system's performance and further enabling autonomous operations in complex environments - that

14

system's specific outputs may be harder to predict as it may acquire new unanticipated behaviours that have not been tested" (Holland Michel 2020, 7).

The extent to which these sorts of AI models are employed by militaries remains to be seen. Likely it will depend on whether the operational advantages offered by continuous learning mitigate for the loss of predictability of outcomes, or the level of generality at which (un)predictability is present (Payne 2021). However, even if the technical aspects of the system lead to (relatively) predictable outcomes when taken in isolation, AI systems nevertheless pose questions with respect to operational unpredictability when interacting with the specific context of deployment. Operational unpredictability relates to *all* uses of AI systems and therefore all deployments of AWS. As Holland Michel writes,

> "all autonomous systems exhibit a degree of *inherent operational unpredictability*, even if they do not fail or the outcomes of their individual action can be reasonably anticipated. This is because by design, such systems will navigate situations that the operators cannot anticipate" (Holland Michel 2020, 5- emphasis added)

This is a point which has also been discussed by the ICRC,

> "autonomous weapon systems are unpredictable in a broad sense, because they are triggered by their environment at a time and place unknown to the user who activates them. Moreover, developments in the complexity of software control systems – especially those employing AI and machine learning – may add unpredictability in a narrow sense of the process by which the system functions" (International Committee of the Red Cross 2019, 11)

Here predictability is at the confluence of a number of variables including the technical features of the system, the complexity of the conflict environment, adversarial and evasive behaviour of opponents regarding the AWS, and the understanding of the system possessed by the operator of the system. These variables converge to limit the foreseeability of the outcomes of the system even where the technical features of the system alone render the system (comparatively) predicable. Holland Michel illustrates operational unpredictability through the example of the interaction of an autonomous drone with its environment, mapping the interior of a network of tunnels:

> "Even if the drone exhibits a high degree of technical predictability and exceptional reliability, those deploying the drone cannot possibly anticipate exactly what it will encounter inside the tunnels, and therefore they will not know in advance what exact actions the drone will take." (Holland Michel 2020, 5)

15

The variables which characterise operational unpredictability in their interaction with the system will be exorbitantly large, generating a significant degree of uncertainty, and making comprehensive assessments about the effects of the AI system intractable.

This has two implications for using AWS in a way which adheres to the requirement of minimal force. Firstly, it raises questions about intent and its relation to the lethal or non-lethal uses of AWS. As noted above, the lethal and non-lethal uses of a weapon can frequently be indistinguishable in their effects. Assessing the intent in the use of the system thereby becomes important for distinguishing between lethal and non-lethal applications of force. However, the operational unpredictability of the autonomous system uncouples the relation between the intent of the operator and the outcome of the use of AWS. Though an operator can be said to have used AWS in a way which was intended to be non-lethal, the AWS may nevertheless act in a way which is lethal. This challenges the capacity of operators to employ AWS in a way which adheres to the requirement of minimal force.

Second, it has implications for the ability of the operator to determine the chances of success of employing AWS in specific contexts. Success, as noted, is dependent on ability. We saw that ability required (a) recognition of one's ability to attain the legitimate military objective and (b) an ability to factor that recognition into calculations about success. When the effects of the AWS are unpredictable this precludes the possibility of having a full understanding of their ability and behaviour.[4]

Until these 'unknowns' are addressed, it will be problematic to use AWS in a way which meets the requirement of minimal force, particularly in circumstances where that requirement obliges combatants to use non-lethal force. The challenge posed by AWS pertains to the way the requirement of minimal force is applied in concrete settings – i.e. the situational judgement it requires of the operator – rather than to the principle itself. There will be instances where the requirement of minimal force necessitates a non-lethal use of AWS employed in combat or other adversarial situations. As such, it will be important to give considerable attention to an ethics of non-lethal uses of AWS. This is why, it is important that further research is developed to understand to what extent are changes in the behaviour of an AI systems are predictable or not, and how this impact the changing actions/ decisions of human command throughout a defence operation.

---

[4] This is not an issue presented by automatic systems. These systems act in a predictable way, producing outcomes that, whilst influence by contextual factors, can nevertheless be anticipated. This helps the operator obtain a fixed understanding of their ability when using that system.

4. Conclusion

In this article we focused on the *in bello* principle of necessity and its implications for the use of AWS. We drew attention to the requirement of minimal force entailed by that principle, arguing that its neglect in work on AWS has obscured important ethical questions. There may be instance in which a combatant has chosen to deploy an AWS and is obliged to deploy it in a non-lethal way. However, the characteristics of AWS pose a number of challenges to the use of AWS in non-lethal ways, raising problems for their use in line with the minimal force necessary to achieve legitimate military objectives. We focused on one particular challenge here: the unpredictability of the system precluding a full recognition of the ability possessed by a combatant when using AWS. However, the ethical challenges posed by non-lethal uses of AWS are wide-ranging and profound and have not been given adequate attention.

References

Arkin, Ronald. 2018. 'Lethal Autonomous Systems and the Plight of the Non-Combatant'. In *The Political Economy of Robots*, 317–326. Springer.

Arkin, Ronald C. 2010. 'The Case for Ethical Autonomy in Unmanned Systems'. *Journal of Military Ethics* 9 (4): 332–341.

Beard, Jack M. 2018. 'The Principle of Proportionality in an Era of High Technology'. *Forthcoming, Oxford University Press, Lieber Institute for Law and Land Warfare Book Series-Complex Battlespaces: The Law of Armed Conflict and the Dynamics of Modern Warfare, Edited by Christopher M. Ford and Winston S. Williams,(2018)*.

Boogaard, Jeroen van den. 2015. 'Proportionality and Autonomous Weapons Systems'. *Journal of International Humanitarian Legal Studies* 6 (2): 247–283.

Chengeta, Thompson. 2016. 'Measuring Autonomous Weapon Systems against International Humanitarian Law Rules'. *Journal of Law & Cyber Warfare* 5 (1): 66–146.

Choudhury L, Aoun A, Badawy D., de Alburquerque L. A., Marjane J., and Wilkinson A. 2021. 'Letter the Panel of Experts on Libya Established Pursuant to Resolution 1973 (2011) Addressed to the President of the Security Council'. s/2021/229. United Nations Security Council.

Chris Mayer. 2015. 'Minimizing Harm to Combatants: Nonlethal Weapons, Combatants' Rights, and State Responsibility'. In *Routledge Handbook of Ethics and War: Just War Theory in the Twenty-First Century*. London: Routledge.

Davison, Neil. 2009. *'Non-Lethal' Weapons*. Palgrave Macmillan.

———. 2018. 'A Legal Perspective: Autonomous Weapon Systems under International Humanitarian Law'.

Downey, William Gerald. 1953. 'The Law of War and Military Necessity'. *The American Journal of International Law* 47 (2): 251–62. https://doi.org/10.2307/2194822.

Enemark, Christian. 2008. "Non-Lethal'Weapons and the Occupation of Iraq: Technology, Ethics and Law'. *Cambridge Review of International Affairs* 21 (2): 199–215.

Floridi, Luciano, Josh Cowls, Thomas C. King, and Mariarosaria Taddeo. 2020. 'How to Design AI for Social Good: Seven Essential Factors'. *Science and Engineering Ethics* 26 (3): 1771–96. https://doi.org/10.1007/s11948-020-00213-5.

Foy, James. 2014. 'Autonomous Weapons Systems: Taking the Human out of International Humanitarian Law'. *Dalhousie Journal of Legal Studies* 23: 47–70.

Gerdes, Anne. 2018. 'Lethal Autonomous Weapon Systems and Responsibility Gaps'. *Philosophy Study* 8 (5): 231–239.

Goose, Stephen D., and Mary Wareham. 2016. 'The Growing International Movement Against Killer Robots'. *Harvard International Review* 37 (4): 28–33.

Grut, Chantal. 2013. 'The Challenge of Autonomous Lethal Robotics to International Humanitarian Law'. *Journal of Conflict and Security Law* 18 (1): 5–23.

Heyns, Christof. 2013. 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions'. United Nations General Assembly. https://doi.org/10.1163/2210-7975_HRD-9970-2016149.

———. 2017. 'Autonomous Weapons in Armed Conflict and the Right to a Dignified Life: An African Perspective'. *South African Journal on Human Rights* 33 (1): 46–71.

Holland Michel, Arthur. 2020. 'The Black Box, Unlocked: Predictability and Understandability in Military AI'. United Nations Institute for Disarmament Research. https://doi.org/10.37559/SecTec/20/AI1.

Human Rights Watch. 2012. 'Ban "Killer Robots" Before It's Too Late'. Human Rights Watch. 19 November 2012. https://www.hrw.org/news/2012/11/19/ban-killer-robots-its-too-late.

ICRC. 2021. 'ICRC Position on Autonomous Weapon Systems & Background Paper'. Geneva: International Committee of the Red Cross.

International Committee of the Red Cross. 2019. 'Artificial Intelligence and Machine Learning in Armed Conflict: A Human-Centred Approach | International Committee of the Red Cross'. https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach.

Jeff McMahan. 2009. *Killing in War*. Oxford: Oxford University Press.

Krishnan, Armin. 2009. *Killer Robots: Legality and Ethicality of Autonomous Weapons*. Burlington: Ashgate.

Lango, John W. 2010. 'Nonlethal Weapons, Noncombatant Immunity, and Combatant Nonimmunity: A Study of Just War Theory'. *Philosophia* 38 (3): 475–97. https://doi.org/10.1007/s11406-009-9231-3.

Marchant, Gary E., Braden Allenby, Ronald Arkin, and Edward T. Barrett. 2011. 'International Governance of Autonomous Military Robots'. *Columbia Science and Technology Law Review* 12: 272–316.

Matsumoto, Masakazu. 2020. 'Amoral Realism or Just War Morality? Disentangling Different Conceptions of Necessity'. *European Journal of International Relations* 26 (4): 1084–1105. https://doi.org/10.1177/1354066120910233.

Matthias, Andreas. 2004. 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata'. *Ethics and Information Technology* 6 (3): 175–83. https://doi.org/10.1007/s10676-004-3422-1.

O'Driscoll, Cian, Chris Brown, Kimberly Hutchings, Christopher J. Finlay, Jessica Whyte, and Thomas Gregory. 2020. 'How and Why to Do Just War Theory'. *Contemporary Political Theory*, November. https://doi.org/10.1057/s41296-020-00453-x.

Ohlin, Jens David, and Larry May. 2016. *Necessity and Discrimination in Just War Theory. Necessity in International Law*. Oxford University Press. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780190622930.001.0001/acprof-9780190622930-chapter-4.

Paul Ramsey. 2002. *The Just War: Force and Political Responsibility*. Rowman & Littlefield.

Payne, Kenneth. 2021. *I, Warbot: The Dawn of Artificially Intelligent Conflict*. London: Hurst & Company.

Roff, Heather M. 2015. 'Lethal Autonomous Weapons and Jus Ad Bellum Proportionality'. *Case Western Reserve Journal of International Law*, 17.

Schmitt, Michael N. 2009. 'Military Necessity and Humanity in International Humanitarian Law: Preserving the Delicate Balance'. *Virginia Journal of International Law* 50 (4): 795–840.

Schmitt, Michael N., and Jeffrey S. Thurnher. 2012. 'Out of the Loop: Autonomous Weapon Systems and the Law of Armed Conflict'. *Harvard National Security Journal* 4 (2): 231–81.

Sharkey, Noel. 2008. 'Cassandra or False Prophet of Doom: AI Robots and War'. *IEEE Intelligent Systems* 23 (4): 14–17.

———. 2010. 'Saying "No!" To Lethal Autonomous Targeting'. *Journal of Military Ethics* 9 (4): 369–83. https://doi.org/10.1080/15027570.2010.537903.

Sharkey, Noel E. 2012. 'The Evitability of Autonomous Robot Warfare'. *International Review of the Red Cross* 94 (886): 787–799.

Sparrow, Robert. 2007. 'Killer Robots'. *Journal of Applied Philosophy* 24 (1): 62–77. https://doi.org/10.1111/j.1468-5930.2007.00346.x.

Taddeo, Mariarosaria, and Alexander Blanchard. Forthcoming. 'A Comparative Analysis of the Definitions of Autonomous Weapons Systems'.

———. Forthcoming. 'Ascribing Moral Responsibility for the Actions of Autonomous Weapons Systems (Forthcoming)'.

Taddeo, Mariarosaria, David McNeish, Alexander Blanchard, and Elizabeth Edgar. 2021. 'Ethical Principles for Artificial Intelligence in National Defence'. *Philosophy & Technology*, October. https://doi.org/10.1007/s13347-021-00482-3.

Taylor, Isaac. 2020. 'Who Is Responsible for Killer Robots? Autonomous Weapons, Group Agency, and the Military-Industrial Complex'. *Journal of Applied Philosophy* n/a (n/a). https://doi.org/10.1111/japp.12469.

Thomas Hurka. 2008. 'Proportionality and Necessity'. In *War: Essays in Political Philosophy*, edited by Larry May and Emily Crookston. Cambridge: Cambridge University Press.

United Nations General Assembly. 2004. 'Agenda Item 55: Follow-up to the Outcome of the Millennium Summit.' In *A/59/565*. New York: UN. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N04/602/31/PDF/N0460231.pdf?OpenElement.

USAF. 2009. 'United States Air Force: Unmanned Aircraft Systems Flight Plan 2009-2047'. Washington DC: United States Air Force.

Wagner, Markus. 2014. 'The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems'. *Vand. J. Transnat'l L.* 47: 1371.

Walzer, Michael. 1977. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*. New York: Basic Books.

Winter, Elliot. 2018. 'Autonomous Weapons in Humanitarian Law: Understanding the Technology, Its Compliance with the Principle of Proportionality and the Role of Utilitarianism'. *Groningen Journal of International Law* 6 (1): 183–202.

———. 2020. 'The Compatibility of Autonomous Weapons Wiith the Principles of Distrinction in the Law of Armed Conflict'. *International & Comparative Law Quarterly* 69 (4): 845–876.