

Autonomous Weapons Systems and the Obligation to Exercise Discretion

Presentation at the 2016 Meeting of Experts on Lethal Autonomous Weapons Systems,
Convention on Certain Conventional Weapons (CCW), Geneva, 14 April 2016

Eliav Lieblich¹

Introduction

This presentation argues that a key problem posed by AWS² is that they constitute a use of administrative powers against individuals without the exercise of proper discretion. AWS are based on pre-programmed algorithms, and therefore – as long as they are incapable of *human*-like metacognition – when they are deployed administrative discretion is bound. Operating on the basis of bound discretion is *per se* arbitrary and contradicts basic notions of administrative law, notions that, as argued here, complement modern standards of international humanitarian and human rights law. This realization explains better some of the concerns relating to AWS, which are usually expressed in circular arguments and counter-arguments between consequentialist and deontological approaches.

That machines should not be making “decisions” to use lethal force during armed conflict is a common intuition. However, the current discussion as to just why this is so is unsatisfying. The ongoing discourse on AWS is essentially an open argument between consequentialists (instrumentalists) and deontologists. Consequentialists claim that if AWS could deliver good results, in terms of the interests protected by international humanitarian law (IHL), there is no reason to ban them. On the contrary, we should encourage the development and use of such weapons. Of course, proponents of this approach are optimistic about the ability of future technology to make such results possible. They also point out the deficiencies in human nature, such as fear, prejudice, propensity for mistake, and sadism that can be alleviated by autonomous systems.

Those who object to AWS on instrumental grounds, conversely, argue that in the foreseeable future AWS will not be able to satisfy modern IHL’s complex standards – such as distinction and proportionality – and therefore will generate more harm than good.³ They also point out that

¹ Assistant Professor, Radzyner Law School, Interdisciplinary Center (IDC), Herzliya. This presentation is based on Eliav Lieblich & Eyal Benvenisti, *The Obligation to Exercise Discretion in Warfare: Why Autonomous Weapons Systems are Unlawful*, in AUTONOMOUS WEAPONS SYSTEMS: LAW, ETHICS, POLICY 245 (Nehal Bhuta et al. eds., forthcoming Cambridge University Press), and Eliav Lieblich & Eyal Benvenisti, *Autonomous Weapons Systems and the Problem of Bound Discretion* 38 TEL AVIV U. L. REV. (Forthcoming, 2016).

² I use here the term “Autonomous Weapons” (AWS) rather than “Lethal Autonomous Weapons” (LAWS) since questions relating to autonomous use of force arise whether or not such force is necessarily lethal.

³ Instrumentalist objections discuss additional problems, such as the difficulty of assigning *ex post* responsibility and lowering the “price” of warfare which can result in diminishing the restraint on the use of force. In this

AWS will also eliminate the good traits of humanity, such as compassion and chivalry, from the battlefield. While these concerns seem convincing, they fail to lay down a principled objection to AWS since they can always be countered, at least analytically, by resort to optimistic hypotheticals regarding future technologies,⁴ as well as to negative examples of human nature on the battlefield, which are unfortunately abound. Thus, a substantive discussion of AWS must transcend speculative claims regarding their ability to deliver end results, whether these are based on future technologies⁵ or on mutually offsetting arguments from human nature.⁶

Deontologists claim that even if AWS could deliver good immediate outcomes, their use should still be prohibited, whether on ethical or legal grounds. The deontological objections focus on the nature of the computerized “decision-maker” and the human dignity of potential victims.⁷ However, deontologists, too, are placed in an awkward position when confronted with extreme hypotheticals. For instance, they have to admit that even if AWS would be better than humans in mitigating civilian harm in warfare, greater loss of life is preferable to lesser loss of life, only because a machine is involved in the process.⁸ Furthermore, deontological approaches to AWS are lacking in that they argue from notions of dignity, justice and due process,⁹ but they do not tell us how and why these are relevant, as such, in situations of warfare.

The discussion, thus, is caught in a loop of utilitarian arguments and deontological retorts, both not entirely satisfying. This presentation offers a middle-way approach to the question, based on an *administrative* perception of warfare. In particular, it serves to bridge the theoretical gap between warfare and administrative concepts of justice and due process, usually understood to be applicable during peace time.

War as Governance: Modern Warfare as an Exercise of Administrative Power

In order to properly discuss whether notions of justice and due process are relevant to the issue of AWS we have to first address the nature of modern warfare. The basic argument presented

presentation however I will focus only on the primary issue of protection of individuals *in bello*. For a useful summary of additional objections, see generally Autonomous Weapons Report.

⁴ See, e.g., K. Anderson and M. Waxman, ‘Law and ethics for robot soldiers’, *Policy Review*, 176 (2012), available at www.hoover.org/publications/policy-review/article/135336; compare P. Asaro, ‘On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making’, *International Review of the Red Cross*, 94 (2012) 687, 699.

⁵ *Ibid.*, 699.

⁶ See, e.g., N. Chomsky and M. Foucault, *Human Nature: Justice vs. Power – The Chomsky-Foucault Debate*, new edn (New Press, 2006).

⁷ See, e.g., Mission Statement of the International Committee for Robot Arms Control (2009), available at <http://icrac.net/statements/>.

⁸ See L.A. and M. Moore, ‘Deontological ethics’, in Zalta, *Stanford Encyclopedia of Philosophy*.

⁹ Asaro, 700–1. This is because according to Asaro, the essence of due process is ‘the right to question the rules and appropriateness of their application in a given circumstance, and to make an appeal to informed human rationality and understanding.’ *Ibid.*, 700.

here is that under contemporary international law war must be understood as a form of governance¹⁰ – a fact that spawns administrative-legal obligations.

It must be conceded that traditionally, justice and due process were foreign to the idea of war. Classic sources of the laws of war, such as the 1863 Lieber Code, treat the citizens on the other side as part and parcel of the “enemy” in the sense that they were expected, as such, to suffer the hardships of war.¹¹ War was thus seen as a violent struggle between collective entities, in which there was no room for individualization.¹² Such collectivization contradicts the idea of individual agency and responsibility and is of course hardly an example of justice or due process as the terms are regularly understood. Under such an assumption, therefore, it was possible to dismiss any special obligations of justice and due process between a state and the enemy’s civilians. If we adopt this view, then AWS should not be assessed in light of such standards, as deontologists suggest.

However, this perception of war was perhaps sustainable when most conflicts were between equally-capable states. Nowadays, most wars are asymmetric conflicts between states and non-state actors. In many cases, armed force is used by advanced militaries against such actors in failing states or territories. This would be the likely scenario in which AWS would be deployed. For various reasons detailed elsewhere – namely, the absent of accountable sovereigns – such conflicts result in a significant gap in the protection of civilians.¹³ This gap must result in some “diagonal” responsibilities of protection between the potentially affected individuals and the attacker, who is in a position to decide their fate and is residually capable to protect them.¹⁴

This suggests that we must view modern warfare as a form of exercise of state power vis-à-vis individuals rather than a horizontal engagement between equal sovereigns. We can thus understand state action during armed conflict as the exercise of administrative, executive action. Once it is perceived this way, warfare should be subjected to widely accepted notions of administrative law that governs executive decision making.¹⁵ Importantly, such obligations can be triggered even before full territorial effective control by the attacker, both due to the

¹⁰ See E. Benvenisti and A. Cohen, ‘War is governance: explaining the logic of the laws of war from a principal-agent perspective’, *Michigan Law Review*, 112 (2013), 1363–1415; E. Lieblich, ‘Show us the films: transparency, national security and disclosure of information collected by advanced weapon systems under international law’, *Israel Law Review*, 45 (2012), 459–91, 483.

¹¹ General Orders No. 100: Instructions for the Government of the Armies of the United States in the Field (Lieber Code), Art. 21.

¹² See L. Oppenheim, *International Law*, para. 58.

¹³ See E. Lieblich with O. Alterman, *Transnational Asymmetric Armed Conflict under International Humanitarian Law: Key Contemporary Challenges* (2005) 18–19.

¹⁴ See generally E. Benvenisti, ‘Rethinking the divide between jus ad bellum and jus in bello in warfare against nonstate actors’, *Yale Journal of International Law*, 34 (2009), 541–8.

¹⁵ Importantly, with the expansion in the understanding of the notion of “control” in international law, such obligations might be triggered.

broadening understanding of “control” in contemporary international law¹⁶ and the emerging general duty of sovereigns to take other-regarding considerations in their decision-making processes.¹⁷

Therefore, the fact that certain armed conflict cross national borders does not, in itself, negate administrative-like responsibilities between a state and civilians on the other side. Does this in itself mean that states must treat “enemy” civilians as their own? A reasonable answer could be found in the principle of equal moral worth, which requires that a state cannot treat “enemy” civilians beneath minimum acceptable standards, in a manner that it would never treat its own.¹⁸ In the specific context of AWS, we may ask whether states would be willing to use them in situations where their own citizenry could be affected by “decisions” made by such systems. It seems that some states already answered this question in the negative.¹⁹ Indeed, if states limit computerized decisions with regard to their own citizens, it is questionable whether they could subject others to such decisions. The effect of an administrative perception of warfare over the question of AWS is thus clear: AWS would be subject to additional, residual constraints, even if such weapons would perform reasonably, in terms of immediate results, under IHL.

Indeed, while IHL does not directly refer to principles of administrative law, some traits of administrative-legal thinking can be found even in positive law. For instance, the duty to take “constant care” in the conduct of military operations²⁰ is reminiscent of the administrative law notion of the obligation to exercise discretion when making decisions, which is central to our argument as well. As we argue, the administrative approach should inform our understanding of the “constant care” standard.

Similar ideas can be derived also from international human rights law, which requires that the limitation or (in some cases) deprivation of rights be subject to due process in the course of limiting those rights. The administrative law perception can thus inform also our understanding what constitutes ‘arbitrary’ deprivation of life during armed conflict.

AWS and the Binding of Administrative Discretion

¹⁶ ECtHR, *Al Skeini v. UK*, Appl. no. 55721/07, Judgment of 7 July 2011, paras 131–40; Human Rights Committee, General Comment 31, Nature of the General Legal Obligation on States Parties to the Covenant, UN Doc. CCPR/C/21/Rev.1/Add.13 (2004), para. 10.

¹⁷ E. Benvenisti, ‘Sovereigns as trustees of humanity: on the accountability of states to foreign stakeholders’, *American Journal of International Law*, 107 (2013), 295–333.

¹⁸ See D. Luban, Risk Taking and Force Protection, Georgetown Public Law and Legal Theory Research Paper no. 11-72 (2011), 12, 46.

¹⁹ For instance, Article 15 of Council Directive (EC) 95/46 enshrines the right of every person ‘not to be subject to a decision which produces legal effects concerning him ... which is based solely on automated processing of data’. Council Directive (EU) 95/46 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281, 1995 Art. 15(1).

²⁰ Additional Protocol I, Art. 57(1).

As aforementioned, the administrative law perception allows us to understand the duty to take “constant care” as a requirement to exercise continuous discretion during hostilities. This obligation requires the administrative authority the duty to consider each decision, within the confines of its legal authority, in light of the specific goals of the authorizing norm, as well as the rights and interests of those affected in the specific circumstances. Of course, this duty implies a prohibition of fettering one’s discretion in advance, since binding one’s discretion negates the ability to consider each case individually and make appropriate adjustments when needed.²¹

The justifications for the obligation not to fetter one’s discretion are twofold. The first stem from notions of human dignity, which require the executive to give due respect to the individual by considering her specific case.²² The second justification relates to decision-making quality, and assumes that in the long run, good administrative decisions cannot be made, in a complex world, without making constant adjustments.²³ This is due to epistemological limitations, which limit human ability to prejudge or foresee complicated situations.²⁴

Deployment of AWS contravenes this duty. This is because during hostilities, the duty to exercise discretion requires the active, ongoing intention not to inflict harm on civilians.²⁵ It requires the commander (and subordinates) to exercise discretion both when planning the attack and, importantly, during the attack, up to the last moment before pulling the trigger.²⁶ In our context, in an out-of-the-loop scenario, the duty to exercise discretion “in the last moment” would have to be performed by the AWS. As aforementioned, AWS cannot engage in the metacognition required for “true” discretion. Their use thus reflects the stringent binding of executive discretion in advance – through the pre-programmed algorithms that govern their behavior.

Binding discretion during warfare through AWS seriously contravenes both rationales of the duty to exercise discretion. First, it runs counter to the obligation to give due respect to the individual, since at hand are life or death decisions in which the potentially harmed individual is not considered at all but is ‘factored’ into the pre-determined processes. Second, war is an extremely complex environment, by nature requiring constant adjustments. In such an environment, any operation based on rigid, pre-determined decision-making patterns is unlikely to yield good, all-things-considered, long-term results.

²¹ See, e.g., *British Oxygen v. Minister of Technology*, [1971] AC 610, HL (UK). See generally J. Jowell et al., *De Smith’s Judicial Review* (Sweet and Maxwell, 2014).

²² Benvenisti, ‘Sovereigns as trustees’, 314.

²³ And this remains true even if acting without exercising discretion would bring good results in this or that specific case. The duty to exercise discretion is concerned with long run decision-making quality.

²⁴ See, e.g., T.J. Barth and E.F. Arnold, ‘Artificial intelligence and administrative discretion’, 338, 348–9; see also H.L.A. Hart, ‘Discretion’, *Harvard Law Review*, 127 (2013), 652–65, 661–4.

²⁵ See M. Walzer, ‘Coda: Can the Good Guys Win’, *European Journal of International Law*, 24 (2013), 433–44, 437.

²⁶ Additional Protocol I, Art. 57(2)(a)(i), 57(2)(b).

Some proponents of AWS claim that human discretion is indeed exercised, by stressing that it is embedded into the system through the human discretion of the programmers. Discretion, then, is exercised but on a different *temporal* level.²⁷ However, this argument is unconvincing since it precisely points out the problem we highlight: that AWS do not (and cannot) exercise discretion in real-time, as an administrative legal perception requires.

Even if the discretion is exercised by the deploying commander,²⁸ this will not change our conclusion. This is because it is unlikely that the deploying commander would be able to predict how the AWS operates, a problem aggravated by the well documented “computer bias” which causes humans to rely heavily on computer decision-making when available.²⁹ If, conversely, commanders *are* able to foresee the exact manner in which the AWS operates, this would to that the weapon is not sophisticated enough to satisfy the complex standards of IHL to begin with. Now, if we suggest that the system can be significantly adjusted in real time – up to the last minute – then possibly, the system is not autonomous at all. A further related argument is that an AWS can be programmed to freeze in complex situations it cannot resolve. However, ascertaining that a certain situation is complex is in itself a substantive decision that requires discretion.

Indeed, one can raise the question whether this analysis would hold in cases where “friendly AWS” are deployed. For instance, let’s assume that an AWS is charged with rescue operations in hazardous areas. Can such system be weaponized for “self” protection? Of course, it seems unreasonable that machines be allowed to kill just in order to preserve themselves, since this, in essence, is tantamount to recognizing a right to kill in the defense of property. In principle, such a “right” might make sense only as an extension of the right to life of the persons the AWS sets out to rescue. If the AWS is bound to rescue person X, and person Y attempts to destroy the AWS, person Y is in essence killing person X. The AWS could then be justified in killing person Y in order to save X, as an instance of defense of others. However, this might all be true in schematic theoretical cases, but in real life, significant discretion must be exercised in order to ascertain that this is indeed the situation. In essence, such cases merge with the (very) controversial question of whether AWS can be deployed in law enforcement operations and to perform the complex standards of use of force applicable in such contexts.

In sum, the duty to take constant care and thereby exercise discretion, results in the conclusion that, AWS – as long as they do not possess the ability to exercise true discretion – cannot be allowed to make final targeting or killing decisions.

²⁷ See, e.g., M.N. Schmitt and J.S. Thurnher, “‘Out of the loop’: autonomous weapons systems and the law of armed conflict”, *Harvard National Security Journal*, 4 (2013), 232–81, 266.

²⁸ *Ibid*, 267.

²⁹ Barth and Arnold, 348. It should be clarified, that AWS can be both based on rigid programming and still be unpredictable.

What if AWS are Deployed in Circumstances where Only Combatants or Direct Participators in Hostilities are Targeted?

Until now, the discussion focused on the relations between the state and the adversary's civilians. However, does this analysis hold even in cases where AWS are deployed in "traditional," symmetric battlefield scenarios, in which it is clear that no civilians are present? Similarly, what if they are deployed in uncluttered environments such as in deserts, the open sea or in space? Indeed, some proponents of AWS claim that for a weapon to be banned, it needs to be shown that its deployment would be unlawful in *all* circumstances.³⁰ If, the argument goes, combatants can be targeted at all times merely by virtue of their status, then AWS might be lawful in uncluttered environments in which only combatants are present. If, however, there are some administrative-like obligations between the attacker and enemy troops, then the same problem of bound discretion applies also in such cases.

At first sight, this question seems strange: enemy soldiers are fighting the state, and therefore it is unreasonable to argue that there are any administrative-trusteeship relations between the state and troops fighting against it. However, this intuition is not sufficient, since the mere fact that an individual, say, a known terrorist, threatens state security, does in and of itself negate the state's administrative obligations towards him or her.

Therefore, in order to answer the question, we need to say something about the nature of the moral justification to target enemy soldiers during war. Chiefly, the question is whether the morality of targeting is status based or threat base. On the one hand, we might say that soldiers are *per se* targetable by virtue of their legal status. Indeed, this is the way positive law was traditionally interpreted.³¹ If this is true, there is no need to exercise substantive discretion when deciding to target them. However, legal status notwithstanding, the common moral justification for targetability of combatants is still constructed around a notion of threat: meaning, combatants can be attacked since they are presumed to be threatening.³² If we admit that threat plays any part in the moral justification of targeting – and even if we concede that a strong presumption of threat exists when combatants are involved – this must mean that some kernel of discretion must remain throughout. For instance, the fact that combatants rendered *hor de combat* cannot be targeted, requires exercising discretion in determining whether, in a specific case, a person is indeed *hor de combat*.

In the same vein, in recent years there are significant challenges – albeit still, perhaps, *de lege ferenda* – to the traditional view that combatants are *per se* targetable. At first, some authorities pointed to a requirement that civilians, even if directly participating in hostilities, must be arrested whenever possible – and killed only if impossible.³³ The kill/capture debate

³⁰ Schmitt and Thurnher, 266.

³¹ See e.g., G. Blum, 'The Dispensable Lives of Soldiers', 2 *Journal of Legal Analysis* 115, 123–26 (2010)

³² M. Walzer, *Just and Unjust Wars* 145 (4th ed. 2006); J. McMahan, *Killing in War* 32–37 (2009).

³³ Case HCJ 769/02, *The Public Committee against Torture in Israel v. The Government of Israel (Public Committee v. Israel)*, 62(1) PD 507, para. 40 [2006] (Isr.); N. Melzer and International Committee on the Red Cross (ICRC),

then migrated to the question whether the rights of *soldiers* also spawn a “duty to capture.” In this context, some scholars press for the recognition of such an obligation, whether through a narrow interpretation of the concept of military necessity or through an expansive understanding of the notion of *hors de combat*.³⁴ Whatever the basis for possible obligations to prefer the capture of enemy combatants, their underlying assumption correlates with diagonal, administrative-like obligations between a state and enemy combatants *qua* individuals. To the extent that these perceptions gain traction, they will of course strengthen the need for substantive discretion even in “pure,” symmetrical battlefields, where civilian casualties are unlikely.

It must be emphasized furthermore that the claim that in such “pure” situations AWS would be legal notwithstanding their inability to exercise discretion, assumes, in fact, that such sterile environments do exist nowadays (or are at least common enough to make a difference). However, as aforementioned, most modern conflict tend to be asymmetric and occur in complex environment. Moreover, in asymmetric conflict, significant discretion is required in order to ascertain whether a person is a combatant, or otherwise directly participating in hostilities to begin with.³⁵

In sum, it seems that administrative legal obligations – chiefly the duty to exercise discretion – might apply even *vis-à-vis* enemy troops. If this is indeed the case, the problem of bound discretion posed by AWS in other contexts, applies to enemy troops also.

The Challenge of “Dumb” Time-Suspended Weapons

One key challenge must be addressed. It is arguable that “dumb” kinetic weapons such as bullets, ballistic rockets or artillery rounds also constitute cases of bound discretion. This is because once fired, there is no turning back and thus discretion is terminated upon the weapon’s discharge. However, this *absurdum* challenge is unconvincing since the time gap between the exercise of human discretion and the weapon’s impact is negligible in such cases. Change of circumstances between release and impact, of the type that requires the reengagement of discretion, is highly unlikely. AWS, on the other hand, would probably be designed to act independently for prolonged periods of time, in which circumstances might change significantly, and thus require renewed human discretion. Moreover, such dumb weapons do not presume to exercise discretion to begin with – rather, they merely follow simple physical rules and thus human discretion upon launch is in general sufficient.

Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law (2009) part IX. But see W. Hays Parks, ‘Part IX of the ICRC ‘Direct Participation in Hostilities’ study: no mandate, no expertise, and legally incorrect’, *NYU Journal of International Law and Politics*, 42 (2010), 769–830, 783–5.

³⁴ R. Goodman, ‘The power to kill or capture enemy combatants’, *European Journal of International Law*, 24 (2013), 819–53; but see M.N. Schmitt, ‘Wound, capture or kill: a reply to Ryan Goodman’s “The Power to Kill or Capture Enemy Combatants”’, *European Journal of International Law*, 24 (2013), 855–61

³⁵ See, e.g., the elaborate debate in ICRC Interpretive Guidance on the Notion of Direct Participation in Hostilities.

A related argument is that some “dumb” weapons such as landmines indeed exhibit a time gap between deployment and impact, just as AWS. However, this claim does not vindicate AWS as much as it exposes the problems with landmines and similar weapons: they are perhaps indiscriminate because they involve binding of discretion. Moreover, the “dumbness” of landmines will be taken into consideration by the reasonable commander and is likely to restrict their use. Conversely, the perceived sophistication of AWS, and their chimera of discretion, will achieve just the opposite: the commander will regard this “discretion” as sufficient from the perspective of the law. As opposed to the case of landmines and similar “dumb” weapons, the commander will be more inclined to absolve herself from exercising discretion by relying on the weapon’s discretion instead.

Conclusion

Much of the current debate on AWS refers to the need for “meaningful human control” or “appropriate levels of human judgment.” Such terms are not self-explaining and require significant interpretation. As this presentation argues, human control or judgment must be understood in the context of modern warfare, which is closer to executive-administrative action than to classic wars. Since AWS do not exercise substantive discretion – because they cannot engage in metacognition and are based on predetermined algorithms – their use amounts to executive action based on bound discretion. Meaningful human control or appropriate judgment must thus be understood to imply that human beings – as the only agents of “true” discretion – should make final targeting decisions when human lives are affected.