The Future of Killing: Ethical and Legal Implications of Fully Autonomous Weapon Systems

Martin Lark[§]

ABSTRACT

Warfare is moving towards full weapon autonomy. Already, there are weapons in service that replace a human at the point of engagement. The remote pilot must adhere to the law and consider the moral and ethical implications of using lethal force. Future fully autonomous weapons will be able to search for, identify and engage targets without human intervention, raising the question of who is responsible for the moral and ethical considerations of using such weapons. In the chaos of war, people are fallible, but they can apply judgement and discretion and identify subtle signals. For example, humans can identify when an enemy wants to surrender, are burying their dead, or are assisting non-combatants. An autonomous weapon may not be so discerning and may not be capable of being programmed to apply discretion, compassion, or mercy, nor can it adapt commanders' intent or apply initiative. Before fully autonomous weapons use lethal force, it is argued that there needs to be assurances that the ethical implications are understood and that control mechanisms are in place to ensure that oversight of the system is able to prevent incidents that could amount to breaches of the laws of armed conflict.

Keywords: Autonomous weapons systems, unmanned combat aerial vehicle, UCAV, ethics, lethal force, accountability

INTRODUCTION

This paper considers the legal and ethical accountability for the actions of fully autonomous weapon systems and will discuss the challenges that leaders need to consider in deploying such systems. Warfare, whether on a global or more localised scale, is inevitable and the need for those engaging in war to have superiority over their adversaries is fundamental. For centuries, this concept has equated to the need to deploy superior numbers of troops, ships, and more recently, aircraft to overwhelm the opponent. This rudimentary power balance allowed dominance in the battle space, ultimately leading to defeat of the enemy. Combatants may die in huge numbers, but the rules of war in recent times,

[§] Corresponding author: mglark01@gmail.com

generally taken to be the Geneva Conventions of 1949 (subsequently updated by the additional protocols of 1977), are in place to provide a legal framework to protect "non-combatants." For example, the following groups are wherever possible to be spared and treated with dignity: civilians; non military infrastructure, such as hospitals, schools, and significant architecture; and even enemy combatants, for instance, those who are deemed *hors de combat*[1] or are identified as child soldiers; (International Committee of the Red Cross ICRC, 1977). Historically, a human (usually the commander in the field) has made the decision to target, or not target, a particular person, place or thing, however, in more recent conflicts, this decision process has moved away from the point of engagement, for example, when semi-autonomous weapon systems are utilised (Boothby, 2012).

The removal of humans from the battlefield by the deployment of robots is a military objective that is gaining momentum. Historically, man has attempted to wage war from an ever-increasing distance; for example, the slingshot, longbow, rifle, artillery piece, and now the remotely piloted aircraft, are continually evolving technological means of delivering lethal force against an adversary, while reducing the physical and mental risk to one's own combatants.

Ethical concerns on the use of remote weapons are not new; indeed in his history of war, weapons and aggression, O'Connell (1989: 48) identified that in Homer's *Iliad* (considered to have been written between 760BC and 710 BC) there is comment that the use of a bow to fight at distance was not in keeping with the spirit of heroic and confrontational warfare. Despite Homer's misgivings, attempts to preserve one's own force while exposing the adversary to increased risk is fundamental to strategic planning, and the historical desire to fight at ever increasing distance to gain tactical advantage raises the notion that the absence, or remoteness of a weapon operative per se does not amount to either a legally, or ethically important change to what has gone before (Boothby, 2012).

CURRENT BATTLESPACE

There are already semi-autonomous weapons systems in service that replace the need for human presence at the point of conflict (Krishnan, 2009). While systems that move on land or sea are under rudimentary construction, it is in the air that technology is most advanced and semi-autonomous unmanned combat air vehicles (UCAV) are currently being used in conflict zones such as Afghanistan and Iraq (Kellenberger, 2011; Sharkey, 2008). By way of example, heavily armed systems,

such as the *Predator* and *Reaper*, each capable of carrying up to fourteen Hellfire missiles can autonomously navigate within a programmed area and can search for and identify targets. However, it is ultimately a remote human "pilot" sitting thousands of miles away, who assesses the validity of the target and whose decision it is to use deadly force. In practical terms, although the UCAV is physically pilotless; the end result, being the release of weapons systems against a target, is undertaken within the same legal bounds as would be employed in a conventional piloted aircraft. The remote pilot must give regard to the law and consideration of the ethical implications of the decision to use lethal force (Altman, 2013).

BUILDING DISCRIMINATION INTO AUTONOMOUS SYSTEMS

The future of warfare is not likely to follow such conventional, human-controlled decision making processes (Boothby, 2012). Technology is advancing at such a rate that the future battle space is likely to be centrally managed from great distances, and the decision to search for, locate, and use lethal force against enemy targets could be conducted entirely autonomously by robots (Arkin, 2010). For the purpose of this paper, the term *autonomy* is defined as the absence of human control in the weapon's algorithm-based target acquisition, identification and engagement process (Boothby, 2012) and discounts the necessary involvement of humans in the logistical effect, as in the case of, say, the launch, re-fuelling, or recovery of the weapon system.

It is clear that even in such a case where there is an absence of human decision-making, the requirement to show appropriate target discrimination (e.g. between hostile and non-hostile actors), proportionality, and precaution in attack remains necessary (Boothby, 2012; ICRC, 1977). It is not concealable that a robot weapon system is created to perform autonomously, would not relieve a command structure involved in ordering the operation of their responsibilities under international treaties and law to ensure that "everything feasible" was done to identify legitimate military targets and to reduce the impact on the civilian population. This is likely to apply to those involved in programing the weapon. (Boothby, 2012; ICRC, 1977, Art.57-2(a)(1))

In a command and control environment, the commander's intent is provided to subordinates, who develop strategies, plans, and tactical actions to achieve an effect that aligns with the commander's requirements (Adlam & Villiers, 2003; Commonwealth of Australia, 2009; Smith & Flanagan, 2000). At all times in this

process, the commander retains ultimate responsibility for the actions taken to achieve the effect relying on the expertise and moral and ethical judgment of the subordinate to understand and act within the law, and to apply compassion and appropriate discretion dependent on the circumstance of the event they are dealing with, irrespective of a commander's orders to the contrary (Adlam & Villiers, 2003).

In a combat situation, a soldier is expected to apply discretion in a humane way, for example, when an enemy combatant surrenders, their life is spared and they are afforded protection (ICRC, 1977). In the battlefield of the future, when fully autonomous weapons are deployed, will they be capable of making the distinction between enemies actively participating in combat and one who is injured, or *Hors de Combat* who should be spared? This dilemma is one which the commanders of the future must understand, for it is they who may be called to account for an autonomous weapon systems action (Arkin, 2010; Boothby, 2010; Sharkey, 2010). If such distinction is not possible or routinely accurate, then should they be deployed at all?

ELIMINATION OF HUMAN ERROR

Humans are fallible. In the so-called "fog of war," it can be difficult for a soldier to decide whether a target is legitimate or not. There is also a myriad of accounts of soldiers acting inappropriately; for instance, by exacting revenge on surrendered enemy combatants, civilians, or property they have captured. This can include crimes such as summary execution, rape, or wanton destruction of goods and chattels.

There is an expectation that the overall military commander, and even a nation's leader, can be held accountable for the actions of a soldier under their command in the field (Howard, Quigley, & Robinson, 2011: Von Knieriem, 1959). In this context, there is a strong case for the development of autonomous weapons that eliminate human frailty, both in a mental, emotional and physical sense. If systems could be developed that would at least match, or exceed, a soldier's observance of extant laws of war then, it could be argued, a reduction in war crimes incidents would follow. It is anticipated that there would be an additional benefit to deployment of autonomous weapons, such as a reduction in own force casualties and a reduced logistic burden (Arkin, 2010).

There are a number of areas where autonomous weapon systems are expected to out-perform humans (Arkin, 2010). These include:

1. An ability to act fearlessly: Autonomous unmanned combat air vehicles will not need to consider self-preservation when tracking targets. They will be able to view a target much closer than a human would and will be programmed to "self-sacrifice" before they release the weapon if required, thereby eliminating a "shoot first, ask questions later" mindset.
2. Lack of emotion: An autonomous weapon's decision-making will not be clouded by fear, hysteria, or emotion that currently exists in battle. These reactions are symptomatic of behaviours that tend to increase the incidence of criminal actions (Walzer, 2006).
3. Battlefield observation and information assessment processes would be developed to operate more quickly than human response capability. Through employment of enhanced optical sensors, ground and wall penetrating radar, acoustics and data analysis, the autonomous weapon will be closer to the enemy and better equipped to make an informed decision on the use of lethal force, than is currently possible for humans to emulate (Arkin, 2010).
4. Avoidance of the psychological condition of "scenario fulfilment" where in stressful situations, humans will distort or ignore new information that challenges their pre-existing belief patterns (Sagan, 1991, as cited in Arkin, 2010, p. 333). This condition was found to be a causal factor in the destruction of an Iranian commercial aircraft by US forces in 1988 when it was mistaken to be an attacking military jet fighter (Arkin, 2010). Autonomous weapons would not be susceptible to such cognitive bias.

Adams (2002, as cited in Arkin, 2010) argued that autonomous military systems, including weapons currently on the verge of production, will eventually be too numerous, too swift, and will generate an arena too complex for humans to direct, which raises serious concerns about the ability of humans to remain in control in a future conflict zone. Conversely, the commander must acknowledge the positive impact that such technological advances may have on his own force, or the way in which a nation's leaders could use the deployment of autonomous weapons to engender positive effects amongst its citizens. An illustration of this is the media interest over the death of an Australian soldier in a conflict zone. Removal of soldiers from direct conflict would exponentially reduce the risk of own force casualties.

There are important health and logistics benefits as well. To keep an army in the field requires a significant logistic effort. To maintain a human combatant

in top condition they need to be fed, clothed, and equipped, they also need shelter and regular rest. The commander in the field must take account of all aspects of their subordinates' existence and they must also provide strategic guidance on tasks and objectives. As an army closes with the enemy, the risks increase. Soldiers may be killed or injured by enemy action, or by environmental factors, such as terrain or weather. They may suffer immediate psychological issues that will cloud their decision-making and objectivity, or these symptoms may manifest sometime later, resulting in the permanent loss of that individual to the commander for future conflicts.

In addition, the financial burden to the respective state in maintaining a physical presence in a war zone is substantial. It is estimated that the financial cost of the current relatively small Australian Defence Force deployment to Iraq, (circa 600 troops, eight strike aircraft and logistic support), costs approximately A$400 million per annum (Toohey, 2014). The future battle space, employing autonomous weapon systems in place of troops, will counteract some of these significant issues. However, the transition to fully autonomous weapons has its problems, which can cause some serious ethical concerns.

ETHICAL CHALLENGES

The remote pilots of today's unmanned combat air vehicles are in no danger of being killed. They sit in relative comfort, thousands of miles from the battle space, and generally, work standard shifts (Sharkey, 2010). Nonetheless, it is this very distance that can create an ethical dilemma for command. It is clear that fear of death diminishes the further a combatant is from the enemy; indeed, many conflicts have been lost by men running in panic from an opposing force (Holmes, 2003).

Studies conducted on UCAV pilots in the US have identified a phenomenon termed *moral disengagement* in which remote pilots disengage emotionally from their actions in using lethal force (Royakkers & Van Est, 2010; Singer, 2009). The issue appears to be that the weapons control system is configured to resemble a computer gaming console, which has the effect of engendering a so-called "Play Station" mentality (Royakkers & Van Est, 2010). Interviews with US UCAV pilots have showed that, in one instance, a remote pilot described his view of fighting from his "cubicle" by commenting, "…it's like a video game. It can get a little bloodthirsty, but it's fucking cool" (Singer, 2009: 308–309). Or another who mused:

> The truth is, it wasn't all I thought it was cracked up to be. I mean, I thought killing somebody would be this life changing experience. And then I did it, and I was like 'All right, whatever.' (…) Killing people is like squashing an ant. I mean, you kill somebody and it's like 'all right, let's go get some pizza' (Singer, 2009: 391–392).

This almost impassive response raises the argument that UCAV pilots may be so detached from the battle area that they cease to act in an ethical manner. This may cause them to deploy lethal force more readily than they would do if physically present on the battlefield. So, will the future transition to fully autonomous weapon systems remove these issues? Or, will the ethical focus merely shift from the *presence* of a remotely situated human pilot to the *lack* of a human in the decision-making process to use lethal force?

During an attack, the application of discrimination and proportionality in attack, to satisfy the laws of war, can be difficult. People make mistakes and can act immorally, especially when under severe stress (Sharkey, 2010; Sparrow, 2007). It is conceivable that autonomous weapons could also act immorally, but unlike humans, they cannot be held responsible (Sharkey, 2010). A robot does not "think," therefore it cannot be disciplined. So, the dilemma lies in identifying where in the process responsibility for violations of law reside. Sparrow (2007) argued that if it is not possible to identify someone who can be held accountable—be that the operator, the commander in the field, the strategic commander in control of the operation, the manufacturer, the weapon-programmer, or indeed the Minister of Defence—then the fully autonomous weapon should not be deployed at all.

The case for caution is further made when we examine the confusion that exists in a war zone. In cases where non-uniformed combatants are engaged—as in a counterinsurgency conflict—the ability to identify who to kill is predicated on situational awareness, having a perception of current intention and probable actions of those observed in the combat zone.

There are also instances when the use of lethal force against enemy combatants is not lawful, as in the case of when combatants are burying their deceased, or when they have made clear their intent to surrender (ICRC, 1977). Humans can pick up on such "signals" even when they are subtle, while an autonomous weapon may not be so discerning. Davis (2007) argued that an autonomous weapon system may not be capable of being programmed to apply

discretion, compassion, or mercy, nor can it adapt commanders' intent or apply initiative.

Kellenberger (2011: 5–6), the former president of the International Committee of the Red Cross (ICRC), expressed numerous concerns relating to the potential deployment of autonomous weapons, for example, by warning:

> A truly autonomous system would have artificial intelligence that would have to be capable of implementing IHL (international humanitarian law) … Their development represents a monumental programming challenge that may well prove impossible. It would also raise a range of fundamental legal, ethical, and societal issues, which need to be considered before such systems are developed or deployed.

Kellenberger's (2011) warnings may be pertinent; however, there is no indication of a reversal in the rate of technological research and development in the move to the deployment of a fully automated weapon system (Asaro, 2012; Boothby, 2012). Asaro (2012) added a further ethical issue to the deployment of such weapons, by commenting that the removal of humans from the battlefield may reduce the threshold for states to initiate conflicts, especially when there is an asymmetric imbalance in forces or technology. He included that autonomous weapons able to employ lethal force without human oversight may ultimately do so, without specific authority from executive or political leaders, culminating in the unintentional start or intensification of conflicts.

On the one hand, the future where autonomous weapon systems, whether they are acting alone or as part of a networked system, elect to initiate conflicts with belligerents without human sanction is not accepted as inevitable. On the other hand, a number of military observers see beneficial effects of deploying these systems, especially from an ethical viewpoint (Boothby, 2012; Royaker & Van Est, 2010; Sharkey, 2010). It is argued that the use of such autonomous systems, especially in a conventional war scenario — broadly defined as a *state-against-state* conflict, using uniformed forces and weapons in open confrontation (Gorka & Kilcullen, 2011) — could be programmed to seek, identify and engage opponents, with knowledge of enemy strength, disposition and intended tactical movement, with little risk of breaching the laws of war.

Moreover, through the aggressive presence of such systems, the enemy could be denied ground without the need to commit one's own forces. The impact of such an event on civilians caught in the contested zone, may directly reduce

non-combatant loss of life and result in less non-military property destruction, due to the absence of dynamic actions. Autonomous systems may also reduce the incidence of criminal activity against non-combatants or civilian property and infrastructure (Boothby, 2012). It could be envisaged that autonomous surveillance systems may be programmed to observe the actions of both friendly and opposing forces, enabling commanders to ensure that kinetic actions are conducted within the laws of war. If breaches are identified, the swift reporting process, including the evidence of high resolution imagery, would provide a solid base for the strategic commander to take dynamic action to cease (own forces), eliminate (enemy forces) or prevent further occurrence (Boothby, 2012). This surveillance and reporting function would provide effective real-time ethical oversight to the battlefield in a way that has not been previously possible.

CONCLUSION

Despite the warnings and misgivings raised by some, the continued development of fully autonomous weapons systems seems inevitable (Boothby, 2012; Royaker & Van Est, 2010; Sharkey, 2010; Singer, 2009). So, the challenge for politicians and the defence leadership is to understand the strengths and limitations of these systems, and to prioritise the response to future developments.

This could be done by considering the legal and ethical implications of deploying these weapon systems and by examining the need for amendments to the conventions and protocols as they relate to the protection non-combatants. This course of action should not be viewed as without precedent, as the so-called Ottawa Treaty (Casey-Maslen, 1997) set in place international laws to prohibit the use, stockpiling, production and transfer of anti-personnel mines and set a timetable and procedures for their destruction. The analogy between anti-personnel landmines and autonomous weapon systems of the future should not be underestimated, as once placed on the ground mines become indiscriminate in their destruction and require no human interaction to engage with lethal force. It is reasonable to ask whether anti-personnel mines would have been developed and deployed had the ethical issues their use caused been better understood at that time?

The transition to full autonomy has provided insight into the ethical issues that the use of fully automated weapons will generate. This not only relates to the destruction of targets where deployment of lethal force is of dubious legality, but also in relation to the incidence of moral disengagement seen in some UCAV

pilots. Encapsulated within the ethical implications of using fully autonomous weapons is a psychological duty of care for commanders to consider in the current UCAV pilot cohort. Further debate needs to be held at an international governmental level to determine protocols that will guide the development, deployment and use of fully autonomous weapons. This needs to establish just who in the process is likely to be accountable for breaches of law that may occur in the conduct of an autonomous weapon strike.

There are good reasons why the development and deployment of fully autonomous weapons should continue, however, these benefits could be lost by the actions of a system that although superior to humans in speed of reaction, technology and physical endurance, lack the human instincts of compassion, proportionality and situational awareness, and as a result are unlikely to be able to differentiate between an opponent who is willing and able to fight and one who is surrendering or injured. Understanding the chain of accountability for the unlawful actions of robots must be an integral component of the weapons introduction to the battle space.

NOTE

1. French*: "…*outside of combat. A civilian or a soldier who has relinquished or been extricated from combat status. A person is hors de combat if: he is in the power of an adverse Party; he clearly expresses an intention to surrender; or he has been rendered unconscious or is otherwise incapacitated by wounds or sickness, and therefore is incapable of defending himself" (ICRC, 1977, Protocol 41*).*

ABOUT THE AUTHOR

Martin Lark, MPolLeadMangt, is a Major in the Australian Defence Force. He is Provost Marshal at the Headquarters Joint Operations Command, where he acts as principal Service Police advisor to the Commander of Joint Operations.

REFERENCES

Adlam, R and Villiers, P (2003). *Police Leadership in the 21st Century: Philosophy, Doctrine and Developments.* Winchester, UK: Waterside Press.

Altmann, J (2013). "Arms Control for Armed Uninhabited Vehicles: An Ethical Issue," in *Ethics and Information Technology*, Vol. 15, No. 2, 137–152.

Arkin, RC (2010). "The Case for Ethical Autonomy in Unmanned Systems, in *Journal of Military Ethics*, Vol. 9, No. 4, 332–341.

Asaro, P (2012). "On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making," in *International Review of the Red Cross*, Vol. 94, No. 886, 687–709.

Boothby, W (2012). Some Legal Challenges Posed by Remote Attack," in *International Review of the Red Cross*, Vol. 94, No. 886, 579–595.

Casey-Maslen, S (1997). *Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction*. Retrieved from http://legal.un.org/avl/ha/cpusptam/cpusptam.html

Commonwealth of Australia (2009). *Department of Defence Executive Series: Command and Control, ADDP00.1*. Canberra, Australia: Author.

Davis, D (2007). *Who Decides: Man or Machine?* Retrieved from http://armedforcesjournal.com/who-decides-man-or-machine/

Gorka, SL and Kilcullen, D (2011). *An Actor-Centric Theory of War: Understanding the Difference Between COIN and Counterinsurgency*. Washington, DC: National Defense University.

Holmes, R (2003). *Acts of War: The Behaviour of Men in Battle.* London, UK: Cassell.

Howard, J; Quigley, J and Robinson, KJ (editors). (2011). *Genocide in Cambodia: Documents from the Trial of Pol Pot and Ieng Sary*. Philadelphia, PA: University of Pennsylvania Press.

Kellenberger, J (2011). *Keynote Address*, presented at the *International Humanitarian Law and New Weapon Technologies*, 34th Round Table on Current Issues of International Humanitarian Law, San Remo, Italy, September 8-10. Retrieved from http://www.iihl.org/iihl/Documents/JKBSan%20Remo%20Speech.pdf

Krishnan, A (2009). *Killer Robots: Legality and Ethicality of Autonomous Weapons*. Farnham, UK: Ashgate.

International Committee of the Red Cross (1977). *Protocol Additional to the Geneva Conventions of 12 August 1949 and Relating to the Protection of*

Victims of International Armed Conflicts (Protocol I). Geneva, Switzerland: Author.

O'Connell, RL (1989). *Of Arms and Men: A History of War, Weapons and Aggression.* Oxford, UK: Oxford University Press.

Royakkers, L and Van Est, R (2010). "The Cubicle Warrior: The Marionette of Digitalized Warfare. *Ethics and Information Technology*, Vol. 12, No. 3, 289–296.

Sharkey, N (2010). "Saying 'No!' to Lethal Autonomous Targeting," in *Journal of Military Ethics*, Vol. 9, No. 4, 369–383.

Singer, PW (2009). *Wired for War: The Robotics Revolution and Conflict in the Twenty-First Century*. New York: Penguin.

Smith, N and Flanagan, C (2000). *The Effective Detective: Identifying the Skills of an Effective SIO*. London: Policing and Reducing Crime Unit.

Sparrow, R (2007). "Killer Robots," in *Journal of Applied Philosophy*, Vol. 24, No. 1, 62–77.

Toohey, B (2014). "Iraq Deployment Easily Affordable," in *Australian Financial Review*, October 11. Retrieved from http://www.afr.com/news/policy/budget/iraq-deployment-easily-affordable-20141013-11coda

Von Knieriem, A (1959). *The Nuremberg Trials*. Chicago: H Regnery Co.

Walzer, M (2006). *Just and Unjust Wars: A Moral Argument with Historical Illustrations*. New York: Basic Books.

- o O o -