Democracy and regulating autonomous weapons: biting the bullet whilst missing the point?

John Williams

This is the pre-peer-reviewed version of the following article: John Williams (2015), 'Democracy and Regulating Autonomous Weapons: Biting the Bullet Whilst Missing the Point?' Global Policy *?? (?): ??-?? which has been published in final form at www.????*

Abstract

This paper argues that current public policy debate around the regulation of emerging autonomous weapons systems is vital, yet also in danger of neglecting crucial challenges. Current analysis focuses around efforts to define autonomy and to incorporate 'autonomous' systems within established regulatory systems, particularly international law and arms control treaties and conventions. This leads to an emphasis on two key decision moments as the focus of regulation: the initiation of hostilities and target engagement, reflecting the just war tradition that provides the intellectual backdrop for much of this debate. The paper suggests this underestimates the significance of the potential consequences of such weapons systems, arguing that this consensus disguises the extent to which autonomy can only be meaningfully engaged within the specific context of the circumstances when such systems may be deployed, and that the speed of decision-making by such systems will outstrip regulatory endeavours focused on the two decision moments. The paper thus argues that only wide-ranging debate, especially within democracies leading the development of such systems, about the relationship of autonomous systems to the nature and purpose of military violence and underpinning democratic values and principles, can adequately address the full extent and significance the challenge presented by the emergence of contextually autonomous weapons.

Author

John Williams is Professor of International Relations in Durham University's School of Government and International Affairs. He is also a member of the Durham Global Security Institute. His research focuses on normative elements of international relations theory, especially the English school, and on contemporary aspects of the just war tradition.

Democracy and regulating autonomous weapons: biting the bullet whilst missing the point?

The emergence of weapons systems capable of a growing range of functions independent of human control is a source of significant debate, especially within democratic societies. At present, democracies leading the development of such technology, principally the US and UK, affirm their policy is not to deploy systems without human involvement in, or monitoring of, command and control decisions, especially around target engagement (HRW, 2012, pp. 7-8). However, public policy debate about regulation of 'autonomous' weapons is significant and the source of substantial public concern (e.g. ICRAC, 2009, 2014; HRW, 2012). Some fear present assurances about continued human oversight will be incrementally eroded to the point where oversight is, effectively, impossible or meaningless (e.g. Heyns, 2013, p. 6; HRW, 2012, p. 20). Whether governments of other states interested in acquiring such systems will exercise the same rigorous control as the US and UK is also potentially troubling. Addressing the challenges of Lethal Autonomous Weapons Systems (LAWS) is a bullet that must be bitten and a responsibility the leading democracies must accept. Development of increasingly technologically advanced weaponry, up to and including artificially intelligent systems, could dramatically alter conduct of military operations, changing strategic, operational and tactical thinking and the way in which democratic states engage with the role of military force in politics.

As already apparent with extensive use of unmanned aerial vehicles (UAVs – or drones in common parlance), which have extended their role from intelligence, reconnaissance and surveillance into strikes against a range of targets, public policy concerns are real and significant. As unmanned ground vehicles are increasingly weaponized and the capabilities of next generation UMS become clear, there is a real possibility of the use of weapons systems that undertake an increasing range of functions, including target identification and engagement, with limited human oversight or possibly without any human involvement (e.g. DOD, 2011). Existing systems with this capability are restricted to specific roles and environments – anti-ballistic missile systems such as Patriot and Iron Dome; short-range ship protection systems such as Phalanx; counter rocket, artillery and mortar (C-RAM) systems; and SGR-1 'guard robots' deployed by South Korea in the DMZ between itself and North Korea[1] – yet those restricted, defensive roles are unlikely to persist.

Regulating new military technology is always challenging, but the argument here is that autonomous UMS pose particular challenges as yet insufficiently addressed in public policy debate, especially as their capabilities increase and their capacity for autonomous decision-making is extended. Present regulatory debate focuses on the potential impact of autonomous weapon systems at two key decision points: the initiation of hostilities, where there are concerns that autonomous systems may take people to war inadvertently; and the moment of engaging human targets, where there are concerns about maintaining discrimination and proportionality. These two decision points reflect deep-rooted thinking and practice in regulating military violence, yet, I argue, autonomous weapon systems potentially render those key decision points inappropriate as the prime focus for regulatory

policy and practice. Democratic public policy debates needs to be extended to engage seriously the extent to which autonomous weapons systems have a role to play in military violence consistent with the values that characterise democracies and how regulatory policy and practice must develop to retain the realistic possibility of exercising effective control over the deployment of autonomous systems.

I elaborate this argument in four main steps. Firstly, I consider regulatory consequences of the problem of defining 'autonomy', showing how variable definitions reflect assumptions about the two key decision points of initiating hostilities and engaging targets. Secondly, I look at why initiation of hostilities and target engagement are the two central regulatory moments, arguing this reflects the structure of the dominant ethical discourse surrounding war: the just war tradition. Thirdly, I consider perspectives derived from science and engineering which begin to broaden debate. Finally, I look at how the complex interaction of these different perspectives is shaping debate and how consensus around initiation of hostilities and target engagement may miss the point about the public policy significance of emergent LAWS.

Autonomy

Do we already live with autonomous weapons systems? For some, the answer is, 'Yes – and we have done for quite some time.' I have identified existing systems (Patriot, Phalanx, C-RAM and their ilk) that seem to fulfil the US Department of Defense (DOD, 2012, p. 13) criterion of 'autonomy': 'once activated, [they] can select and engage targets without further intervention by a human operator.' In contrast, the UK Ministry of Defence (MOD, 2011, para 206), sets the bar for autonomy very high: 'autonomous systems will, in effect, be self-aware and their response to inputs indistinguishable from, even superior to, that of a manned [system]. As such, they must be capable of achieving the same level of situational awareness as a human.' These systems could be given a set of mission parameters and be expected to decide the best tactics to deploy to fulfil that mission, set appropriate levels of force and targeting for those tactics, and assess when mission parameters had been fulfilled and action should cease (MOD, 2011, paras 205-206). Anything short of human-level situation awareness is 'automation'.

'Detect, identify, engage and destroy an incoming ballistic missile within this defined spatial volume' is a very different mission from 'cooperate with friendly human forces to identify and suppress armed insurgents sheltering amongst a civilian population within this city block, exercising utmost care to avoid civilian casualties and capturing insurgents alive whenever possible.' A system capable of the former would count as autonomous for the DOD, but not the MOD, which would require something close to fulfilment of the latter. But are these efforts at absolute definition useful and appropriate? Appeals for a ban, or at least a moratorium, on the development of autonomous UMS are not uncommon, with the UN Special Rapporteur on Extrajudicial, Summary or Arbitrary Execution; Human Rights Watch (HRW); and the International Committee on Robot Arms Control (ICRAC), amongst others, calling for such steps (Heyns, 2013; HRW, 2012; ICRAC, 2009, 2014).

The focus of these calls is systems which meet the DOD definition: 'weapons that could select and engage targets without human intervention' (HRW, 2012, p. 1); 'systems that,

once activated, can select and engage targets without further intervention by a human operator.' (Heyns, 2013, pp. 7-8). The lack of human oversight ('out of the loop') systems, or systems where human oversight ('on the loop') is minimal or ineffective (HRW, 2012, p. 2) is central to these concerns, especially because of the increased operational speed of such systems (HRW, 2012, pp. 9-12). This highlights how target engagement as a critical decision point is embedded in definitional debates about autonomy. 'Meaningful human control' at the moment of target selection is a common point at issue in this debate: with all three of those terms contestable for some (e.g. Sharkey, 2014).

However, discussion excludes systems like Phalanx, Iron Dome and C-RAM, which are seen as 'automatic' (e.g. HRW, 2012, pp. 12-13). The basis for distinguishing between 'automatic' and 'autonomous' seemingly rests on two factors. Firstly, the complexity of the environment within which the systems operate: shooting down incoming ordnance is contrasted with more complex missions (HRW, 2012, pp. 12-20). Heyns (2013, p. 8) suggests a contrast between 'structured and predictable' and 'open' environments, although without clarity on where the line between those two is drawn. Secondly, the nature of the target matters: ordnance, not people. '[T]hey present less danger to civilians because they are stationary and defensive weapons that are designed to destroy munitions, not launch offensive attacks.'(HRW, 2012, p. 12). More extensive discussions of this distinction, and others, such as 'human supervised autonomous weapons' and 'sense and react to military objects' (SARMO) systems (e.g. Sharkey, 2014), highlight how context and circumstance are engaged in debates about the definition of autonomy. That inevitably impinges on the account of acceptable levels of human control: the computers may have more license when targeting incoming munitions on the high seas, less if targeting munitions in an urban environment or if they are targeting people.

This suggests Anderson and Waxman (2013) are correct: autonomy is a matter of degree and context. It will be approached incrementally, with systems that presently perform certain functions autonomously under certain conditions acquiring additional capabilities. In some environments those capabilities will be sufficient that human monitoring of operations is all that is necessary. In others, direct human control over will be needed because of the complexity of the mission or environment or both. Differentiating between the automation of operational capabilities and the acquisition of autonomy over decision-making is not clear cut without reference to context, including the type of target that is engaged and the moment of engagement. Further complicating this definitional issue is the nature and extent of human control necessary for it to be 'meaningful'. That not only relates to the context and circumstances of deployment and engagement, but also to philosophical and psychological issues (e.g. Sharkey, 2014).

Concerns over the potential of autonomous systems to make initiation of hostilities either more likely or even automatic also play an important role in debates over the significance of autonomy. Critics claim they will lower the threshold of war (usefully summarised in Altmann, 2013, p. 140). Furthermore, a new arms race may ensue as increasing numbers of states seek to deploy more and better autonomous systems (e.g. Altmann, 2013), reinforcing increased propensity for war. Sparrow (2009a, pp. 26-27) offers a vision of large numbers of armed UMS on near-permanent stand-by just beyond the borders of potentially

hostile states, shadowed just inside those borders by other LAWS waiting to pounce, drastically raising the risk of war by mistake or non-human decision. Micro- or nano-scale systems, with individually harmless autonomous components 'swarming' at the point of attack to create devastating weapons, is a further instance of this sort of future potential (for wider discussion see Dunn, 2013). The 'dehumanizing' effect of autonomous weapons and the temptations of greatly reduced casualties amongst the armed forces of states deploying such systems reinforce concerns that autonomous war will be seen as easy, quick and clean (e.g. HRW, 2012, pp. 39-41). There are reasons to be sceptical about these claims. Propensity for war is vastly more complex than technological innovation (e.g. Vasquez, 1993). Launching war does not come down to the basis of 'because we can and because technology means we will win'. Regulating, even banning, LAWS is unlikely to have a decisive effect on the war-proneness of world politics.

Recognition of the significance of context and circumstance are crucial to framing effective regulation. 'Contextual autonomy' highlights that this is not an abstract assessment. This approach parallels Thomas Hellström's (2013, p. 101) concept of 'autonomous power': '… the amount and level of actions, interactions and decisions an agent is capable of performing on its own.' Hellström's account fits this sense of autonomy as a spectrum, with the level of significance of autonomy conditioned by the context within which any system operates. Hellström goes on (2013, pp. 102-5) to plot existing systems and common fictional examples using degree of autonomous power and level of lethality to establish the point where some degree of moral responsibility may be assigned to LAWS.

The adequacy and appropriateness of regulatory systems to address the consequences of incrementally increasing levels of automation that phase into autonomy are important questions. 'Contextual autonomy' is, therefore, the appropriate concept, as it recognises that formulation of the problem is not static, and the meaning of autonomy and levels of autonomy weapons systems may possess alter in different contexts. 'Contextual autonomy' requires extension of Hellström's (2013, p. 101) definition: ' the amount and level of actions, interactions and decisions an agent is capable of performing on its own', with: 'within a definable operational context such that the outcome of operations are materially affected by the system's independent actions, interactions and decisions without routine directive intervention by another, human, agent.' That is complex, clumsy and contestable, but such is the nature of definition within problems of this sort.

This brief discussion of debates about 'autonomy' as the crucial characteristic of these emerging weapons systems suggests two conclusions. First, regulatory systems need to be sufficiently flexible to accommodate autonomy's contextual and circumstantial framing. A fixed category of 'LAWS' subject to special rules or restrictions, or even an outright ban, is unlikely to be attainable, because autonomy is not a characteristic that can be judged in a way that makes a distinct category persuasive. Autonomy is contextual, unlike, for instance, characteristics that distinguish chemical, biological or blinding laser weapons that deploy specific and definable means to create lethal effect. The challenges of establishing arms control for autonomous systems are therefore distinct from and greater than those in these areas. Second, current debate reflects the two classic decision points of ethical and regulatory debate about weapons systems: the moment of the initiation of hostilities and

the moment of target engagement. These two problems are connected: seeing autonomy as threatening easier, or even automatic, initiation of war; and weakened or removed human control over who is targeted and how; threatens the basis of established ethical enquiry and regulatory policy and practice. The fear is of autonomous weapons that remove human control over war and central to this is fear that their speed of action will far outstrip human ability to control, or even to monitor and regulate, future weapons.

What this leaves undiscussed is why initiation of hostilities and target engagement are the crucial regulatory moments and whether or not there may be alternatives that better enable effective engagement with the challenges and threats contextually autonomous systems present. I argue this reflects deep-rooted ethical debates about the use of military force, embedded in the just war tradition, and how these are reflected in regulatory policy and practice. The next section summarises how the interaction between military technology and just war thinking underpins concerns about LAWS, before moving on to look at how alternatives may be constructed.

Ethics, public policy and autonomous UMS

At the heart of the matter for some critics of LAWS is a fundamental moral claim that human beings must not be subject to lethal force by non-human weapons systems (e.g. ICRAC, 2009; Altmann et al, 2013, p. 73). Only people may exercise such authority on the basis of a natural, not an artificial, intelligence. No matter how fast, comprehensive and accurate the data-processing capability of computers may be there is something basically morally wrong about handing such authority to a machine. This 'de-humanization' not only risks making war more likely, it also challenges the central moral questions war raises about when and why acts of systematic and organised large-scale violence are justifiable, who may be attacked as part of that activity, and how much force might be used. This is the territory of the just war tradition.

The just war tradition has, from a position of substantial neglect, become the dominant form of ethical discourse about war (Walzer, 2006; Rengger, 2013). It is disparate, hence my preference for 'tradition' over 'theory', but there are unifying features of the various types of theorising reflected in the debate about LAWS. Most important is the structure of ethical judgement into two (or, increasingly frequent in contemporary debates, three) linked but distinctive categories. *Jus ad bellum* provides a focus for ethical assessment of the initiation of war; *jus in bello* addresses the conduct of military operations, especially the engagement of targets; and, the relative newcomer, *jus post bellum* is concerned with the circumstances that enable a just ending to a war. These three points of ethical focus overlap and interconnect, for instance via concerns that war fought unjustly may make a just peace more difficult (Orend, 2000), but remain distinguishable. Initiating war is assessed against a distinctive set of criteria in comparison with its conduct; with the former focused around questions of 'just cause', 'legitimate authority', 'right intention', 'proportionality' (of war as a response), 'last resort', and 'reasonable prospects of success'; conduct is assessed against 'discrimination' and 'proportionality' (of violence used to achieve a specific objective).

There are two principal approaches to applying this just war framework. The edges between them are blurred, but they are worth elaborating as they begin to highlight how and why

ethical debate about autonomous UMS is so complex, with seemingly contradictory claims made about their ethical implications. This in turn explains the depth of the challenge facing public policy. In addition, it reveals why initiation of hostilities and target engagement are dominant as the key moments of judgement, and why this may be problematic.

First is an approach, increasingly prominent in the last twenty years, that brings the just war tradition into accord with the liberal analytical philosophy that underpins most contemporary human rights discourse. This is a wide-ranging movement addressing *jus ad bellum*, *jus in bello* and *jus post bellum* (for leading general examples of this move see e.g. Orend, 2000; Rodin, 2002; McMahan, 2005). Here, crudely speaking, ethical principles can be identified through the rigorous application of philosophical logic, working from a set of carefully stated and grounded first principles – such as the idea of human beings as rights-holders, amongst which the right to life is especially important – in order to better understand complex ethical choices. The use of carefully constructed 'thought experiments' enable ethicists to control the variables under consideration to better understand how very important ethical principles ought to be resolved when they come into conflict with one another, such that basic ethical insights can be applied to the hard-to-control world of military operations.

The right to life is clearly central, with substantial debate as to how that right, at the individual level, may relate to a state's right to self-defence as a just cause for initiation of war (e.g. Rodin, 2002). Further, killing in combat poses profound ethical challenges, as justifiably killing an individual is typically seen as exceptionally demanding, with many right-based theorists rejecting membership of a class of 'combatant' as sufficient, requiring a higher standard of specific individual liability to lethal force as a result of actions or, possibly, intentions (e.g. May, 2005). Questions of *when* we can go to war and *when* we can kill *which* people are dominant, helping to explain why the initiation of hostilities and target engagement are critical moments of judgement: they are moments when the most fundamental rights are most at stake. Technological mediation of both decisions is a long-standing reality, but technological determination of either or both is what is at stake as automaticity phases into autonomy and humans move from being 'in' to 'on' the loop, and, perhaps, to being 'out' of it entirely. Can this be rendered consistent with the protection of rights? Belief that this is not possible is central to the position of many critics and advocates of a ban, such as Human Rights Watch (HRW, 2012).

Second, is the idea of just war as a tradition of thinking that seeks to provide a set of questions that we must ask about the resort to, conduct and settlement of war. This sees ethics as a situated activity, deploying techniques of case-based reasoning (often known as casuistry, reflecting the origins of the just war tradition in medieval and early-modern Catholic theology) and hermeneutics in order to reach balanced, reasoned judgements about the complex specifics of real-world cases. The phenomenology of war is important here, too, reflecting on the experiences and perceptions of those engaged in military operations. An ethos of 'practical wisdom', as opposed to the philosophical precision of the analytical approach, characterises this account of just war (e.g. Johnson, 2007; Rengger, 2013). The same criteria for conflict initiation and target engagement are applied, but understood differently, as contextual assessment is important.

Timing, though, remains crucial: *jus ad bellum* criteria such as 'last resort' make it necessary to consider whether, in the circumstances, there are other options still available that might offer effective redress; right intention asks whether actors are committed to the long haul to bring about a more just situation in the aftermath of conflict. The 'legitimate authority' condition frames debate about who holds the right to make critical decisions, and on what basis. Within conflict, the *jus in bello* criteria of discrimination asks about the legitimacy of a target in specific circumstances, with the liability of people to lethal force being shaped by the military situation – for instance concern that killing uniformed soldiers of the Iraqi army as they retreated from Kuwait towards Basra in 1991 amounted to (unjustifiable) 'slaughter' (Burke, 2007, p. 205). The ethical justifiability of *when* hostilities are initiated and their conduct, understood in the context of the time, cannot be determined *a priori*, even if philosophical tools can help us in our reasoning. Such an approach, I argue, is more useful to better engaging with the wider implications of LAWS, because it broadens the ethical agenda away from initiating hostilities and target engagement to ask about the context of the conflict itself and the values and standards of those who fight them, which provide important behavioural cues and restrictions on permissible behaviour.

These differences matter, and not just to philosophers, because they have the potential to establish different types of public policy debates that push towards rather different regulatory structures. Acceptance of analytical philosophy as a basis for regulation is likely to reinforce incorporating LAWS into a human rights-based legal regime. Rights-based regulatory structures have many significant benefits, because they help to preserve the central focus on human beings as the ultimate referent object of ethical discourse, addressing a key concern of critics about the de-humanising effect of their use (e.g. HRW, 2012). Analytical philosophy is more suited to the promulgation of rules, including rules about processes to be followed when, inevitably, rules conflict with one another.

Any effort to summarise international legal debate over LAWS in this paper would inevitably be futile, but the centrality of human rights to the analytical approach to just war theorizing is reflected in much legal debate. Schmitt (2013), for example, sees few problems with LAWS, arguing current International Humanitarian Law (IHL) can accommodate such systems with little, if any, substantial modification. The US DOD (2012, pp. 3, 7, 11, 12) repeatedly identifies compliance with IHL as a design requirement of any future systems, alongside 'appropriate human involvement'.

Casuistic approaches, by contrast, are more likely to support regulatory structures emphasising the need for contextual judgement about the appropriateness of using LAWS when set against a broader set of political and ethical objectives, including the nature of the conflict in which their deployment is considered. Rights may well be a primary element of these objectives, but thinking of this sort is more open to balancing rights-based claims against objectives that are harder to express in terms of rights because they appeal to ethical concepts or reference points that are not reducible to individual rights. Instances may include communal well-being and the value attached to shared beliefs, culture or values. UMS may offend against such values, for instance conceptions of military service that emphasise values such as honour, courage and self-sacrifice.[2] They may assist wider communal values such as the obligation to privilege the protection of a shared way of life

and sense of national identity and belonging in the face of existential threat. Importantly, assessing and understanding LAWS is best served by on-going critical reflection on real-world instances. *A priori* judgement is always contingent.

Current armed UAVs effectively illustrate debates that deploy these distinctive approaches and relate to technologies that are a precursor to LAWS. For example, adopting an analytical stance, Strawser (2010) sees a moral duty to deploy UAVs where this reduces the risk to combatants, as long as there is no loss in capability to observe principles of discrimination and proportionality. The technology is a means to an end, and, if the technology produces better outcomes it should be deployed. Schmitt (2013) extends this to autonomous systems, as does Arkin (2009), a prominent advocate of the detached and impartial decision-making capability of autonomous UMS.

Coeckelbergh (2013, pp. 94-6), appeals to more phenomenological ethical reasoning to highlight how current UAVs may be reversing the dehumanising tendencies of technology and distance in war. The extent of surveillance, the quality of the sensory arrays and the immediacy of the experience reported by drone pilots may be restoring the individuality of those targeted. This is something UAV operators also stress – that their experience of long-term surveillance of some of their targets restores an intimacy to combat unavailable from the cockpit of an aircraft or the inside of an armoured vehicle (e.g. Wittes, 2014; Otto and Webber, 2013). This contradicts claims about detached and impersonal 'video game warfare' (HRW, 2012, p. 40).

Those 'benefits' would be compromised were systems to become increasingly autonomous such that the 'human in the loop' provides only final authorisation for a strike carried out following autonomous surveillance, assessment and target selection by LAWS. 'On the loop' systems would further degrade that quality of engagement. Empathy, compassion and inherent reluctance to harm other humans would be lost, but so too would the possibility of vengeful or racist decision-making. Context matters and the values and principles that can, or should, underpin public policy and military behaviour are not always consistent with one another.

The complexity of the debate over the role that machines may play in life and death decisions is partially revealed. At what point does the human being lose control over the decision to go to war or the process of taking a life? What degree of technological mediation is possible before that loss occurs? Does it necessarily matter, if the machines can make 'better' decisions in terms of conserving human lives and respecting human rights? Sharkey (2014) reflects some of these challenges with promulgation of a five-stage spectrum of human control, suggesting that only stages one and two manifest 'meaningful' control. As with debates over autonomy conducted in the abstract, however, this leaves unaddressed the issues of context and circumstance that more casuistic approaches to ethics consider. Is the loss of human control, perhaps temporarily, a price worth paying if it offers decisive military advantage in a conflict where defeat would result in a triumph for a genocidal dictatorship? That construction consciously echoes Walzer's (2006: 251-269) controversial notion of 'supreme emergency', whereby the abandonment of discrimination and proportionality in war may be, *in extremis*, permissible, or at least excusable. For this version

of just war thinking, there are ethical dilemmas in war, and ethical dilemmas are as acute in the field of war as they are in any other field where life and death decisions occur, such as medicine. Judgements about legitimately deploying violence inevitably involve compromising other values focused around peace and the preservation of human life.

A resolution of the ethicality of autonomous UMS on the level of the fundamental moral permissibility of ceding some level of control over life and death to technology is unavailable, especially when ceding such decisions may enable the lives of some to be saved, by removing them from the dangers of combat, for instance, or permit the deployment of systems that could retrieve casualties under intense fire. UAVs already achieve the former, stimulating debate over whether the manifest force protection benefits they offer come at unacceptable costs of reduced discrimination between combatants and civilians – themselves complex, contested and contextual categories (e.g. Strawser, 2010; Plaw, 2013). Those debates also connect to the wider context of the justifiability of war in general and particular wars. For example, McMahan (2005) argues that without the just cause of self-defence, all deaths in war are ethically impermissible, radically challenging the standard just war consensus, and legal position, that soldiers who kill other soldiers in combat are not normally liable for those deaths.

Complexity is no reason to abandon the philosophical debate. Persistent lack of resolution reinforces the need for continuing enquiry to better understand the different arguments made and to relate these to technological, political, social and military developments. Public policy should be honest about the complexity of the issues contextually autonomous UMS present.

The policy implications of this complexity are significant. Firstly, it reinforces the sense that binary positions, whether in favour of or against LAWS, are problematic and offer a poor basis for policy-making. Secondly, it suggests the need to look beyond the initiation of hostilities and target engagement as the focus of debate. These two moments are deeply embedded in the just war tradition but they are also the moments at which the challenges posed by LAWS are at their most intensive, both in terms of intellectual complexity and, consequently, in terms of the unlikelihood of decision-makers (whether human or not) making the most consistently justifiable choices.

UMS and engineering.

Regulating UMS through the way in which they are designed overlaps, but extends, legal and ethical debates. For example, the legality of UMS is often considered in terms of compliance, or otherwise, with rules on ‘means of warfare’. The MOD (2011, para. 502) notes: ‘Most of the legal issues surrounding the use of existing and planned systems are well understood and are simply a variation of those associated with manned systems.’ Article 36 of the 1977 First Additional Protocol to the Geneva Conventions requires all new means of warfare to undergo technical assessment to ensure systems can fulfil the requirement for discrimination (HRW, 2012, pp. 21-26). As Anderson and Waxman (2013, p. 19) note, it is far easier to design, build and deploy autonomous systems that do *not* comply with that requirement than to produce ones that do. Prodigious technical challenges in designing fully autonomous systems capable of distinguishing between human beings fulfilling the roles of

combatant and non-combatant, especially in complex operational environments, mean legal prohibitions on indiscriminate weapons may be effective regulatory bulwarks against autonomous systems for many years (e.g. MOD, 2011, para 508), as long as those developing such systems accept the law. Anderson and Waxman (2013, pp. 19-20, 23-26) and Garcia (2014) suggest the US, as the technological leader in this field, scrupulously observes these requirements and works with allies to embed as deeply as possible legal and behavioural principles and norms, to establish obstacles to states or non-state groups that may see advantages in indiscriminate LAWS.

Weapons that cause unnecessary suffering or superfluous injury are illegal under Article 35 (2) of the First Additional Protocol to the Geneva Conventions (e.g. Anderson and Waxman, 2013, p. 10). Deployment of existing autonomous systems is eased in this respect because they target weapons, rather than humans, in the form of missiles or other ordnance. Advocates of future LAWS highlight their role in disarming enemy combatants, targeting weapons rather than people: 'Our dream machine would confront an enemy combatant on the battlefield; physically remove the rifle from his hands; saw the rifle in half with a diamond tipped saw; hand the two-halves back to him; and then tell him to, "Have a nice day!"' (Canning, 2009, pp. 13-14) That level of discrimination is, indeed, a 'dream': plausible battlespace reality will not approach such a vision. It is, however, illustrative of the focus on target engagement as a critical issue: by avoiding engaging humans, and instead engaging weapons, legal and ethical obstacles fall away.

How far engineers can or should restrict activity to preclude developing unethical or illegal weapons is not exhausted by the law. For some (e.g. Arkin, 2009; Canning 2009) the challenge is largely around understanding IHL demands and incorporating these into the programming of LAWS such that targeting decisions will be legal. For others (e.g. Kovac, 2013), the problems are profound, with almost all weapons-related research being necessarily ethically problematic, if not prohibited, because it may result in the death of innocent civilians. Sparrow (2009b) argues engineers have a special responsibility to develop systems that are safe for human beings to work alongside. That requirement is more far-reaching than may appear at first glance, helping demonstrate how focusing on initiation of hostilities and target engagement may be challenged.

Sparrow (2009b, pp. 171-176) suggests issues that extend beyond the technical challenges of designing and building LAWS that are at least as safe to work with as existing manned systems and that can interact with manned systems in a way that is no more dangerous, and ideally less dangerous, than present systems. In some instances, such as bomb disposal, that may point strongly in favour of developing LAWS. Keeping humans out of harm's way links these technical issues to a common ethical principle of harm reduction, summarised by the Latin tag *primum non nocere*[3] associated with the Hippocratic Oath, but prominent in many statements of ethical conduct in scientific research (e.g. Royal Society, 2005).

Harm, however, is a very complex concept (for discussion with relevance to international politics see e.g. Linklater, 2011) and one that extends in time and space. Sparrow (2009b, pp. 174-175) picks up on some of these issues, for instance the psychological damage to UAV operators because of their radical distancing from the battlefield with resultant limits to

their ability to intervene in distressing events, including the deaths of both civilians and comrades (also Gregory, 2011, pp. 197-199; Otto and Bryant, 2013). Other potential harms may arise from mistakes that can only be addressed by deploying humans into previously unanticipated situations; or over-confidence in the ability of LAWS to achieve objectives that, in reality, require human deployment. Finally, more complex and costly systems mean human beings may be placed in harm's way to recover broken or damaged equipment (Sparrow, 2009b, pp. 172-174). Consequently, current statements of the harm principle in this context (e.g. Hellström, 2013, p. 106) lack sufficient specificity in terms of the 'harm' that should be prevented, how harm ought to be understood and how the inevitable dilemmas that arise in terms of balancing different harms in complex political circumstances can be resolved, or their debilitating effects mitigated.

The ethics of potentially harmful technology and the moral responsibility of scientists is a long-standing debate that produces conflicting answers, even if it were possible to agree on the basic set of ethical tools that ought to be deployed, given the different time-frames in play and the potential degrees of separation between scientific and engineering research and development, and the harmful uses for eventual products (for an interesting discussion of the general ethical responsibility of scientists and engineers see Koepsell, 2010). This highlights that incremental development of autonomy and the differentiated way in which LAWS are likely to comply with ethical and legal strictures within different environments poses serious engineering challenges. An autonomous weapon system deployed on the warship of a sovereign state targeting incoming anti-ship missiles on the high seas subject to continuous human monitoring is, clearly, a different proposition from the 'killer robot' (e.g. Garcia, 2014), targeting insurgents within a densely populated urban environment in pursuit of a pre-defined mission of broad parameters and without oversight. The engineering that enables the former to be a present reality will play a part in the process that could, perhaps, eventually see the latter deployed. Yet the path from one to the other is so lengthy, complex, expensive and unpredictable that knowing where and when a 'red line' may be crossed in the science and engineering of such systems requires the wisdom of Solomon.

Law and ethics cannot establish clear principles which engineers and scientists can apply to their research in order to know whether it is legal and ethical, now or in the future, without their active engagement in reflection and debate (Koepsell 2010). The responsibility of judgement cannot be evaded by passing it to others, seeing legal compliance as a software engineering challenge. Professional codes of conduct and the ethical standards embodied in representative professional bodies, such as the UK Royal Society (2005), acknowledge and institutionalise this two-way process. Situated judgements against broad principles, and recognition of the reality of dilemmas, hard choices and wicked problems characterise this arena, and cannot be wished away.

This problem is not solely confined to expressly military research. The MOD (2011: para. 625) notes how engineering issues are not confined to explicitly military technology: '[N]ew developments in military systems are … likely to come from specialised development of commercial systems, rather than *vice versa*. It is to the commercial sector that we must look for the delivery of future disruptive technology.' This casts the net of engineering ethics very

wide indeed, and the response to these concerns is similarly wide-ranging. Ineke Malsch (2013), for example, aims to use just war categories and concepts to structure governance of scientific research.

This brief summary scratches the surface of the applicability of broad ethical debates to the science and engineering associated with development and deployment of LAWS. It helps explain the attractiveness of following the focus of just war theory and IHL on initiating hostilities and target engagement. The potential for LAWS to lower the threshold of war, or even to initiate hostilities automatically; and their potential to transform the nature of combat in ways that further reduce the value of human life is real and serious.

The accelerated speed that is characteristic of contemporary technological innovation reinforces the sense of concern that surrounds these two crucial decision points. Systems like Phalanx and Patriot 'think' faster than human beings – indeed they have to if they are to fulfil their role. Slowing the systems down so that humans can keep up defeats the object and flies in the face of centuries of military practice where increased speed confers strategic and tactical advantage. Stopping the machines from taking us to war and stopping them from killing people and destroying property at the moment when those decisions are taken will become impossible. The broader questions that come out of science and engineering ethics contribute alternative foci for policy debate, by asking questions prior to those that arise at the initiation of hostilities and the engagement of targets. Which missions are appropriate for LAWS? How far ought autonomy to extend within those missions? How should democratic accountability be maintained over the initiation of war and its conduct, if those processes LAWS? Should democracies accept strategic and tactical limitations in their use of LAWS in order to preserve democratic values of accountability and engagement? Are present international arrangements, like the UN Conventional Weapons Convention (CWC), appropriate for pursuing international agreement on effective regulation?

Regulating UMS: challenges and opportunities

Effective regulation is in the interests of all involved in the development, deployment and governance of these systems. The importance of this work increases as system capabilities increase and proliferation accelerates. Nobody, it seems, disputes the need for effective action in this arena – it is a bullet that must be bitten. The dominant regulatory debates reflect just war thinking, aspects of national and international law around weapons systems and armed conflicts, and the professional environment within science and engineering. All three interact and are, in some important ways, mutually reinforcing, contributing to the present consensus that conflict initiation and target engagement are the key foci.

However, consensus becomes problematic if important issues lie beyond its purview or are unjustifiably downplayed. Mutual interaction and reinforcement to establish a strong consensus risks creating blind-spots. My concern in relation to UMS, especially as the pace and extent of autonomy increase, is that such blind-spots are being created through over-emphasising initiating hostilities and target engagement. Regulatory structures exist that focus on different temporal points: Article 36 points to one such instance, with its focus on the potential of deployed weapon systems to comply with the requirements of discrimination and proportionality, as does the Article 35 (2) requirement not to develop

weapons that cause unnecessary suffering and superfluous injury. All of the load-bearing terms involved – discrimination, proportionality, unnecessary and superfluous injury – are dynamic, contextually specific and demand debate. Yet they point towards the potential to reconsider the temporal perspective of regulatory debate. This is apparent in the advocacy by some (e.g. ICRAC, 2009; Wallach and Allen, 2013; Altmann, 2013) of arms control as a means of banning or restricting the types and capabilities of LAWS that will be developed, controlling their numbers and establishing monitoring and verification protocols. Whether that be including them within existing treaties, such as those on chemical and biological weapons which include delivery vehicles; reviving moribund mechanisms, such as the Conventional Forces in Europe treaty with its numerical and geographical caps on deployments; or developing new treaty arrangements, whether functional, geographical or a mixture of the two; advocates suggest initiating this activity now could build a long-term regulatory framework.

The model for such systems – Cold War arms control – has appeal, given its ability, over years of patient and often frustrating or seemingly fruitless negotiation, to build a complex system of limits to weapons systems, numbers and deployments that contributed to containing tension and preserving stability in the face of ideological and geopolitical hostility. However, the analogy is only partial (e.g. Altmann, 2013) and the challenge of effective verification in particular is enormous, given the dual-use nature of not just the technology but the systems themselves. A nuclear-armed intercontinental ballistic missile has only one purpose (deterrence). Autonomous UMS may well have many potential uses, including several that would be regarded as beneficial (casualty retrieval from combat zones, for instance).

The democratic political culture and the professional standards and values of scientists, engineers, lawyers and service personnel therefore emerge as additional regulatory foci. Rather than concentrating, as in the present consensus, on the moments of initiating hostilities and engaging human targets, there is an opportunity to extend regulatory approaches into these arenas, and to connect them up in presently underdeveloped ways.

The technological potential of LAWS must be judged in relation to debates about the kinds of conflicts it is ethically and legally appropriate to conduct and the types of systems and tactics democratic states are prepared to countenance. Arms control is an important subsequent manifestation of those debates, but cannot be divorced from them. The values and standards democratic societies espouse and seek to defend and promote manifest in how they configure their military capabilities and deploy those capabilities in specific operations. They also manifest in the values and standards their armed forces inculcate amongst their members. For example, this has taken place in democratic states' increasing expectations that armed forces are capable contributors to humanitarian missions, having a role in fulfilling an international 'responsibility to protect'. They are also about professional cultures and the internalisation of behavioural norms, including rejecting technologically deterministic tropes that any and all advantages must be adopted. These are not static debates, set against ethical and legal criteria, or treaty provision, fixed in their present constellations. As the casuistic approach to just war emphasises, they are applied,

contextual and practical judgements informed by principle, but aware of the need for flexibility in the face of the unpredictable and subject to continuous critical assessment of past experience to inform present practice and future planning.

Policy conclusions

Autonomous UMS raise technical and legal issues affecting current standards and practice around initiating hostilities and engaging targets. Policy-makers must respond to the NGO-led challenge in these areas, accepting and pursuing calls for compliance with the highest existing legal standards in development and deployment. That also requires developing workable definitions of 'automaticity' and 'autonomy' that retain sufficient flexibility to permit appropriate technological innovations enabling safer, more discriminate and more proportionate weapons systems.

Arms control, for example via the UNCWC, is a further valuable course to pursue. Overcoming its limitations – encompassing non-state actors and developing effective and timely verification mechanisms in particular – is a key challenge. It took more than twenty years from the first use of nuclear weapons to the first major arms control treaties, a time-frame that, on many current estimates, could see the deployment of fully autonomous UMS.

Political leadership is necessary to shift the established consensus. This demands sustained and serious debate over if and how technologies of automation and autonomy contribute to the protection and promotion of core democratic values, for example: human rights, dignity and equality; accountability; shared identity; popular sovereignty; national self-determination; and citizen participation. Those values do not always coincide and military violence challenges them all in important, even fundamental, ways. The incremental incorporation of contextually autonomous systems into the arsenals of democracies is underway, but there is, as yet, insufficient consideration of how they may affect the ethical principles involved in the democratic endorsement of the permissibility, on occasion, of systematic and organised violence by the state. That agenda ranges wider than conflict initiation and target engagement, yet it is one that has not, yet, been entered into on a sufficient scale. The just war tradition and the ethics of science and engineering in society provide vital resources for addressing questions within arenas of democratic culture and professional standards and values about when and why wars may be fought, to what ends and deploying which means. That may, in time, produce powerful cultural prohibitions against LAWS, likely to prove at least as effective as legal prohibition (e.g. Tannenwald, 2007) as the incompatibility of LAWS with democratic values are tested, debated and internalised in professional and public cultures. Or it may set contextually flexible and case-specific boundaries around LAWS that inform technical and legal analysis, grounding such work and giving shape and direction to domestic and international regulatory practice.

References

Altmann, Jürgen (2013), 'Arms Control for Armed Uninhabited Vehicles: an ethical issue', *Ethics and Information Technology* 15: 137-152. DOI 10.1007/s10676-013-9314-5

Anderson, Kenneth and Matthew C. Waxman (2013), 'Law and Ethics for Autonomous Weapon Systems: why a ban won't work and how the laws of war can help', American University Washington College of Law Research Paper 2013-11. Available at http://ssrn.com/abstract=2250126 (accessed 11 February 2014).

Arkin, Ronald C. (2009), 'Ethical Robots in Warfare', *IEEE Technology and Society* Spring, 2009: 30-33. DOI 10.1109/MTS.2009.931858

Burke, Anthony (2007), *Beyond Security, Ethics and Violence: war against the other,* London: Routledge.

Canning, John S. (2009), 'You've Just Been Disarmed. Have a Nice Day', *IEEE Technology and Society* Spring, 2009: 12-15. DOI 10.1109/MTS.2009.931864

Coeckelbergh, Mark (2013), 'Drones, Information Technology, and Distance: mapping the moral epistemology of remote fighting', *Ethics and Information Technology* 15 (2): 87-98. DOI 10.1007/s10676-013-9313-6

DOD (2011), 'Unmanned Systems Integrated Roadmap: FY2013-2038' Reference no. 14-S-0553. Available at http://www.defense.gov/pubs/DOD-USRM-2013.pdf (accessed 8 August 2014)

DOD (2012), 'Autonomy in Weapons Systems: Directive No. 3000.09' available at http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf (accessed 28 February 2014).

Dunn, David Hastings (2013), 'Drones: disembodied aerial warfare and the unarticulated threat', *International Affairs* 89 (5): 1237-1246. DOI: 10.1111/1468-2346.12069

Garcia, Denise (2014), 'The Case Against Killer Robots: why the United States should ban them', *Foreign Affairs* 10 May. Available at http://www.foreignaffairs.com/articles/141407/denise-garcia/the-case-against-killer-robots (accessed 8 August 2014).

Gregory, Derek (2011), 'From a View to a Kill: drones and late modern war', *Theory, Culture & Society* 28 (7-8): 188-215. DOI: 10.1177/0263276411423027

Hellström, Thomas (2013), 'On the Moral Responsibility of Military Robots', *Ethics and Information Technology* 15: 99-107. DOI 10.1007/s10676-012-9301-2

Heyns, Christof (2013), Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, United Nations Human Rights Council available at: http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf (accessed 9 July 2013).

HRW (Human Rights Watch) (2012), Losing Humanity: the case against killer robots, available at: http://www.hrw.org/sites/default/files/reports/arms1112ForUpload_0_0.pdf (accessed 20 September 2013).

ICRAC (International Committee on Robot Arms Control) (2009), 'Mission Statement' available at http://icrac.net/statements/ (accessed 27 February 2014).

ICRAC (International Committee on Robot Arms Control) (2014), 'New Mission Statement' available at http://icrac.net/2014/07/icracs-new-mission-statement/ (accessed 12 August 2014).

Johnson, James Turner (2007), 'Just War Thinking in Recent American Religious Debate Over Military Force,' in Charles Reed and David Ryall (eds), *The Price of Peace: just war in the twenty-first century,* Cambridge: Cambridge University Press.

Kovac, Jeffrey (2013), 'Science, Ethics and War: a pacifist's perspective', *Science and Engineering Ethics* 19: 449-460. DOI 10.1007/s11948-012-9355-x

Koepsell, David (2010), 'On Genies and Bottles: scientists' moral responsibility and dangerous technology R&D', *Science and Engineering Ethics,* 16: 119-133. DOI 10.1007/s11948-009-9158-x

Linklater, Andrew (2011), *The Problem of Harm in World Politics: theoretical investigations,* Cambridge: Cambridge University Press.

Londoño, Ernesto (2013), 'Pentagon Cancels Divisive Distinguished Warfare Medal for Cyber Ops, Drone Strikes' *Washington Post* 15 April. Available at http://www.washingtonpost.com/world/national-security/pentagon-cancels-divisive-distinguished-warfare-medal-for-cyber-ops-drone-strikes/2013/04/15/62335492-a612-11e2-8302-3c7e0ea97057_story.html (accessed 13 June 2014)

Malsch, Ineke (2013), 'The Just War Theory and the Ethical Governance of Research', *Science and Engineering Ethics,* 19: 461-486. DOI 10.1007/s11948-012-9357-8

May, Larry (2005), 'Killing Naked Soldiers: distinguishing between combatants and non-combatants', Ethics & International Affairs 19 (3): 39-53. DOI: 10.1111/j.1747-7093.2005.tb00553.x

McMahan, Jeff (2005), 'Just Cause for War', Ethics & International Affairs 19 (3): 1-21. DOI: 10.1111/j.1747-7093.2005.tb00551.x

MOD (2011), *Joint Doctrine Note 2/11: The UK Approach to Unmanned Aircraft Systems,* Shrivenham: The Developments, Concepts and Doctrine Centre. Available at https://www.gov.uk/government/publications/jdn-2-11-the-uk-approach-to-unmanned-aircraft-systems (accessed 18 February 2014)

Orend, Brian (2000), '*Jus Post Bellum,' Journal of Social Philosophy* 31 (1): 117-137. DOI: 10.1111/0047-2786.00034

Otto, Jean L. and Bryant J. Webber (2013), 'Mental Health Diagnoses and Counselling Among Pilots of Remotely Piloted Aircraft in the United States Aircraft', *Medical Surveillance Monthly Report* 20 (3): 3-8.

Plaw, Avery (2013), 'Counting the Dead: the proportionality of predation in Pakistan', in Bradley Jay Strawser (ed), *Killing By Remote Control: the ethics of an unmanned military,* Oxford: Oxford University Press.

Rengger, Nicholas J. (2013), 'The Wager Lost by Winning? On the Triumph of the Just War Tradition', in Anthony F. Lang, Jr., Cian O'Driscoll and John Williams (eds), *Just War: authority, tradition and practice,* Washington, DC: Georgetown University Press.

Rodin, David (2002), *War and Self-Defense,* Oxford: Oxford University Press.

Royal Society (2005), *The Role of Codes of Conduct in Preventing the Misuse of Scientific Research,* RS Policy Document 03/05, London: The Royal Society. Available at: http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/publications/2005/9645.pdf (accessed 5 May 2014).

Schmitt, Michael N. (2013), 'Autonomous Weapon Systems and International Humanitarian Law: a reply to the critics', Harvard National Security Journal available at http://harvardnsj.org/2013/02/autonomous-weapon-systems-and-international-humanitarian-law-a-reply-to-the-critics/ (accessed 20 September 2013).

Sharkey, Noel (2014), 'Towards a Principle for the Human Supervisory Control of Robot Weapons', *Politica and Società* 2: 305-324. DOI: 10.4476/77105

Sparrow, Robert (2009a), 'Predators of Ploughshares? Arms Control of Robotic Weapons', *IEEE Technology and Society* Spring 2009. DOI: 10.1109/MTS.2009.931862

Sparrow, Robert (2009b), 'Building a Better WarBot: Ethical Issues in the Design of Unmanned Systems for Military Applications,' *Science and Engineering Ethics* 15: 169-187. DOI 10.1007/s11948-008-9107-0

Strawser, Bradley Jay (2010), 'Moral Predators: the duty to employ uninhabited aerial vehicles', Journal of Military Ethics 9 (4): 342-368. DOI: 10.1080/15027570.2010.536403

Tannenwald, Nina (2007), *The Nuclear Taboo: the United States and the non-use of nuclear weapons since 1945,* Cambridge: Cambridge University Press.

Vasquez, John A. (1993), *The War Puzzle,* Cambridge: Cambridge University Press.

Wallach, Wendell and Colin Allen (2013), 'Framing Robot Arms Control', *Ethics and Information Technology* 15: 125-135. DOI 10.1007/s10676-012-9303-0

Walzer, Michael (2006), *Just and Unjust Wars: a moral argument with historical illustrations (4th edition),* New York, NY: Basic Books.

Wittes, Benjamin (2014), ‘Lt. Col. Matthew Atkins on the “The Personal Nature of War in High Definition”’, Available at http://www.lawfareblog.com/2014/01/matthew-atkins-on-the-personal-nature-of-war-in-high-definition/#.UvLDO2JdV8H (accessed 27 February 2014).

1 SGR-1 robots are not, presently, set to operate fully automatically, engaging targets they have detected without human authorisation. Such capability does exist, however.

2 Illustrative here is controversy over the proposed Distinguished Warfare Medal, to be awarded to drone pilots and cyber operators. This met such sustained criticism, including that it denigrated traditional military virtues of self-sacrifice and courage in the face of the enemy, that it had to be withdrawn (Londoño, 2013).

3 ‘First’ (or, sometimes, ‘above all’) ‘do no harm’.