Journal of
APPLIED PHILOSOPHY

Journal of Applied Philosophy
doi: 10.1111/japp.12200

What's So Bad About Killer Robots?

ALEX LEVERINGHAUS

ABSTRACT *Robotic warfare has now become a real prospect. One issue that has generated heated debate concerns the development of 'Killer Robots'. These are weapons that, once programmed, are capable of finding and engaging a target without supervision by a human operator. From a conceptual perspective, the debate on Killer Robots has been rather confused, not least because it is unclear how central elements of these weapons can be defined. Offering a precise take on the relevant conceptual issues, the article contends that Killer Robots are best seen as executors of targeting decisions made by their human programmers. However, from a normative perspective, the execution of targeting decisions by Killer Robots should worry us. The article argues that what is morally bad about Killer Robots is that they replace human agency in warfare with artificial agency, a development which should be resisted. Finally, the article contends that the issue of agency points to a wider problem in just war theory, namely the role of moral rights in our normative reasoning on armed conflict.*

Introduction[1]

'War', writes Michael Walzer, 'is a relation between political entities and their human instruments', to wit, soldiers.[2] However, it is doubtful that, in the armed conflicts of the future, political entities – most notably states – will exclusively rely on the *human* instruments of war. This is because some states are developing military robots with varying degrees of technological sophistication. It is therefore likely that, in future wars, weaponised robots, often referred to as Killer Robots[3] – a term I shall also use here – are going to complement and even replace 'human instruments' of war.[4] Some commentators welcome this prospect for two reasons. Firstly, Killer Robots enable the military to fulfil its duty of care towards its members. While it is not obligatory for the military to protect service personnel from all possible threats to life and limb, it is obliged to take reasonable steps to reduce risks to its soldiers. Thus, one could argue that, in theatres where the deployment of Killer Robots is a viable alternative to 'putting boots on the ground', it is, *ceteris paribus*, obligatory for the military to use robots. Secondly, the deployment of Killer Robots, it is claimed, enhances compliance with the rules of war. Hence Killer Robots serve an important humanitarian purpose, namely the reduction of suffering during armed conflict.[5] Given that belligerents are under a duty to use the most effective yet least harmful methods in pursuit of their aims, the deployment of Killer Robots is morally obligatory – or so one might argue.

Neither of these arguments is entirely new. Duties of care have featured extensively in the debate on uninhabited aerial vehicles ('drones'). Likewise, humanitarian defences of precision-guided weapons, including drones, are common. That said, the

humanitarian argument in favour of Killer Robots adds an interesting twist to existing debates on weapons technology. As the persistence of war crimes perpetrated by soldiers attests, human beings often fail to comply with the rules of war, especially in stressful situations.[6] They lose their cool and become trigger-happy. By contrast, robots, as the roboticist Ronald Arkin points out, are not susceptible to similar psychological pressures. They are 'motivated' by algorithms. If it is possible to programme robots in such a way that they do not violate the rules of war, their deployment, rather than the deployment of 'fallible' human soldiers, is morally desirable. It is safer, in other words, to let a robot, rather than a human, pull the trigger.

In this article, I challenge the humanitarian argument for Killer Robots. While I argue that, conceptually, Killer Robots are not technologically unprecedented because they exist on a continuum with other precision-guided weapons, the replacement of human agency in war with the 'artificial agency' of machines is deeply problematic. To develop this critique, I proceed as follows. In the first part of the article, I make some preliminary remarks about the relationship between just war theory and my analysis of the debate on Killer Robots. In the second part, I offer a brief discussion of conceptual issues. In the third and fourth parts, I turn to two normative criticisms of Killer Robots, respectively. In the fifth part of the article, I defend my main normative critique of Killer Robots against four potential criticisms.

1. Preliminary Remarks

This article approaches Killer Robots from a contemporary just war perspective. One recent debate in just war theory, which I will briefly touch upon later in this article, relates to the moral standing of combatants.[7] On the one hand, orthodox just war theorists claim that all combatants have equal rights to kill enemy combatants in war, regardless of the nature of their cause. Walzer calls this the moral equality of soldiers. On the other hand, neo-classical just war theorists deny the moral equality of soldiers. They argue that only soldiers who are fighting for a just cause are allowed to fight and kill. In a nutshell, orthodox theorists and their neo-classical critics disagree on the relationship between the two major normative frameworks that regulate the use of force, *jus ad bellum* and *jus in bello*. The former framework governs the declaration of war, while the latter regulates the conduct of hostilities. Neo-classical just war theorists argue that the permission to use force in war (*jus in bello*) depends on the justness of that war (*jus ad bellum*), a claim which orthodox theorists deny. Despite this fundamental disagreement on the ethics of killing in war, I argue below that there is potential for agreement between orthodox and neo-classical just war theorists on the issue of Killer Robots.

More generally, it is possible to approach the debate on Killer Robots from the perspective of *jus ad bellum* and *jus in bello*, respectively. Regarding the former, one could argue that the availability of new weapons, and weapons research in general, might translate into more *unjust* wars.[8] This line of reasoning is likely to be overly simplistic. While technology plays an important role in warfare, wars often have multiple causes. Nevertheless, if weapons research is morally impermissible for this or related reasons, it is unnecessary to engage with the debate on Killer Robots. It would simply be impermissible to develop Killer Robots precisely because they are weapons. Killer

Robots do not raise any distinctive issues here. In this article, I assume that the development of certain types of weapons is not morally impermissible. If the development and eventual deployment of Killer Robots turns out to be morally impermissible, the reasons for this must pertain to this specific type of weapon, rather than general reasons against weapons research.

From the perspective of *jus in bello*, on which I concentrate here, the question is whether new weapons technologies enhance or threaten compliance with its three core principles:

1. *Distinction*: Belligerents are morally obliged to distinguish between legitimate and illegitimate targets in war, especially with regard to human persons. In war, some human persons are legitimate targets, while others must not be harmed.[9]
2. *Proportionality of Means*: Belligerents are morally obliged to abstain from causing excessive harm in pursuit of military goals.
3. *Military Necessity*: Belligerents are morally obliged to abstain from causing unnecessary harm in pursuit of military goals.[10]

The question of compliance leads me to distinguish between two types of arguments, contingent and intrinsic ones. Contingent arguments focus on a particular state of technological development. Contingent arguments against Killer Robots, in particular, could point to the technological shortcomings of military robotics, which means that the deployment of Killer Robots is likely to violate the principles of *jus in bello*. However, future technological developments may render contingent arguments redundant. Intrinsic arguments against Killer Robots, by contrast, are not contingent upon a specific stage in technological development. One could imagine a technologically perfect Killer Robot, potentially capable of fully complying with the three principles of *jus in bello*, yet oppose this type of weapon. The main argument I shall raise against Killer Robots in the article is intrinsic in nature.

2. What is a Killer Robot?

Before we can analyse the ethical challenges raised by Killer Robots, let me briefly turn to conceptual issues. To begin, the term 'robot' is used to describe artificial devices that sense and purposefully interact with their environment in order to carry out useful tasks.[11] To carry out their tasks, robots have sensor suites that enable them to perceive the world around them. Their sensor suite is usually integrated into a body to which wings, wheels, mechanical legs or arms may be attached. These enable the robot to interact with its environment. Robotic behaviour emerges from the interaction of the machine with its environment. Naturally, robots need energy to carry out tasks, which they receive via batteries, solar panels, or mains. Most importantly, robots have microprocessors and a governing software that controls their behaviour. Without this software, robots would not be able to do anything. If one combines these different elements, it is possible to classify robots as artificial agents. Robots perceive their environment, interact with it, and cause changes within it.

The above definition of Killer Robots leaves open the kinds of useful tasks robots could carry out. Because robots are used for different purposes, it is important to be aware that the debate on Killer Robots is only concerned with a very small subset of

robots. It is not even concerned with military robots as such, that is, robots used by the military to carry out tasks useful to the military. For instance, the use of a bomb disposal robots to dismantle 'improvised explosive devices' appears morally unproblematic. Rather, the debate focuses on military robots that are (1) classifiable as weapons and (2) capable of applying kinetic force to a target by delivering a payload. In this article, I follow the current debate on Killer Robots by focusing on uses of lethal force by a robot.

Compared to other weapons, some argue that Killer Robots represent a distinctive form of military technology. This is because they (1) operate autonomously and (2) are capable of deciding when and how to apply lethal force to a target. This points us to the crucial concept of machine autonomy. In a basic sense, a robot, some roboticists would argue, is autonomous if it is capable, once programmed, of carrying out tasks without human assistance. Clearly, this falls short of what philosophers understand by the term autonomy. For Kant, an agent is *morally* autonomous in the sense that she acts for reasons she has given herself.[12] Yet, as noted above, robots are controlled by a governing software, which has been programmed by a human to enable robots to carry out specific tasks. This means that robots act, as Kant would put it, heteronomously. Robots, then, are better classified as what I call *operationally* autonomous. In the present context, this means that, once programmed, Killer Robots can carry out the five steps underlying what the military calls a Kill Chain without human assistance:[13]

(1) acquire information about potential targets via their sensors,
(2) analyse the resulting data,
(3) check whether a potential target may be engaged,
(4) execute a targeting decision, and
(5) assess the aftermath of the use of force.

Still, it is not clear in what sense Killer Robots 'make decisions'. To tackle this issue, I propose that we distinguish between executing a targeting decision ('execution model') and generating a targeting decision ('generation model'). According to the execution model, the relevant *jus in bello* criteria are applied by the programmer before the robot is being deployed. That is, the programmer determines which targets the robot is permitted to engage. The question, on this model, is whether it is possible to deploy Killer Robots in such a way that they do not violate any of the assessments made by the programmer. On the generation model, the robot must be able to interpret and apply the *jus in bello* criteria of distinction, proportionality, and necessity without human assistance. That is, it must be capable of assessing whether a particular object or person qualifies as a morally legitimate target, and whether a particular military act is morally permissible. On both models, machines make decisions, but these are qualitatively different.

According to the execution model, the robot is capable of carrying out the steps of the Kill Chain on its own. For instance, it is able to search for potential targets, analyse the information it receives via its sensors, and engage appropriate targets. But as we just saw, what counts as a morally legitimate target is determined by the programmer in what one may term a meta-targeting decision. A highly sophisticated robot could make further decisions during the execution process, but, crucially, these occur within the parameters set by the meta-targeting decision. Suppose that P has

programmed KB to attack enemy tanks. Once deployed, KB detects three enemy tanks. If $tank_3$ is easier to reach than $tank_1$, KB could execute P's meta-targeting decision by opting for $tank_3$, rather than $tank_1$. In this sense, KB made a decision. The morally relevant point, however, is that KB only decides between targets that have already been deemed morally legitimate by P. KB does not determine that enemy tanks, rather than enemy border posts, are morally legitimate targets.

The generation model, by contrast, assumes that robots can generate something like meta-targeting decisions. But I do not think that the generation model is, in the foreseeable future, a realistic possibility. The criteria of *jus in bello* pose major problems here. Beginning with the principle of distinction, it is difficult to see how Killer Robots could determine whether a human person is a legitimate target or not. From the perspective of a Killer Robot, for instance, a child with a toy gun, an illegitimate target, may look very similar to a legitimate target, such as a fully armed combatant. This point might also undermine the above claim that Killer Robots can be used to execute targeting decisions: they will not be able to identify the target they have been programmed to attack. True, the execution model is already incredibly technologically demanding. There is a real question whether robotics will ever be able to produce a Killer Robot that could function under the execution model. But it appears even more difficult to programme a robot to determine itself whether a human person is a legitimate target or not than to programme a robot with criteria that would enable it to attack a specific person.

Things look bleaker when one considers whether Killer Robots could apply the criteria of proportionality and necessity.[14] Regardless of whether one follows orthodox or neo-classical interpretations of *jus in bello*, any calculations of whether a particular course of action is likely to cause excessive harm would involve an assessment of the impact of military actions on legitimate human targets *as well as* those who must not be harmed, at least intentionally. The criterion of distinction, however one conceives it in detail, is thus directly relevant to the criteria of proportionality and necessity. But because Killer Robots will not be able to apply this principle in the first place, they will lack a sound basis for making appropriate judgements of proportionality and necessity.

Another problem arises from the inherent vagueness of the necessity and proportionality criteria.[15] Even if compliance with the principle of discrimination could be ensured, Killer Robots will find it hard to interpret what constitutes proportionate and necessary harm. This is because the application of the two criteria is highly context-dependent. To illustrate the point, imagine that, in order to halt Green's aggression, it is militarily necessary for Yellow to destroy a strategically important weapons factory in Green's territory (necessity). Suppose that, at t_1, Yellow is about to lose the war and ponders whether to attack the factory. If, at t_2, Yellow bombed the factory, one hundred civilians would die as a side-effect of the strike. It is necessary to destroy the factory, and given how badly things are going for Yellow, the one hundred civilian casualties may not be disproportionate. However, because of bad weather at t_2, Yellow has to call off the attack. At t_3, weather conditions have improved and Yellow considers attacking the factory again. Fortunately, at t_2, Yellow secured a major victory on a different front. It might still be militarily necessary to destroy the factory in order to prevent Green from producing more weapons, but because the importance of this target has declined due to gains at t_2, it may now be excessive and disproportionate to

cause one hundred civilian deaths at t_3. This example shows that one cannot assign a fixed value to potential targets. It is impossible to programme a general rule into Killer Robots that states that targeting *x*-category objects causing *z*-amount of harm would always be excessive, while causing *y*-amount of harm in targeting *x*-category objects would not be. Essentially, the application of the proportionality criterion involves making *moral judgements*, which machines are currently incapable of making. To avoid confusion, I should stress that these judgements do not just involve, as in the above example, tracking the relevant empirical information. This is something that machines could do. Rather, the real problem is that the ways in which lives are 'weighted' and 'balanced' in proportionality calculations is not fixed, but subject to a much broader change in circumstances on the ground. In the above example, Yellow, at t_3, may decide, due to its improved situation, to weigh the lives of Green's civilians more strongly than at t_1. This is hard to replicate in machines.

If this is sound, Killer Robots, for the time being, will be incapable of generating ethically sound targeting decisions, and it is impermissible, if not practically impossible, to develop and deploy such robots. But importantly, it may not be impermissible to develop and deploy Killer Robots to *execute*, rather than generate, targeting decisions. I critically assess this claim in the next two parts of the article.

3. Killer Robots and the Responsibility Gap

One general criticism of autonomous technology is that it creates so-called responsibility gaps.[16] These are situations in which no one is responsible for what a machine does. Operationally autonomous machines sometimes behave in an unpredictable manner. It is inappropriate, some critics argue, to hold programmers and designers responsible for the unforeseen consequences of operationally autonomous machine behaviour.[17] For Robert Sparrow, in particular, the possibility of responsibility gaps is a strong reason against the development of Killer Robots. I debunk the notion of the responsibility gap below.

Sparrow defends his version of the responsibility gap by utilising the following thought experiment. A Killer Robot commits a 'robo-war crime' by killing a group of surrendering combatants. The robot targets the combatants because it has determined that the costs of keeping them alive are too high. To be fair, the robot differs from operationally autonomous weapons that will realistically be developed over the next decades. It is a quasi-human. Nevertheless, Sparrow's conclusion is relevant here. Since the 'decision' to kill the surrendering combatants arose solely within the robot, no one is responsible for the robo-war crime.

But Sparrow perhaps proceeds too quickly here. From a technological perspective, the decision to kill the surrendering combatants could not have solely arisen within the robot. The designer of the robot must have developed a mechanism that enables the machine, quasi-human or not, to balance lives against costs. Further, the programmer deploying the weapon must have allowed the machine to relate the criteria of 'lives' and 'costs' to each other. Why should the designer or programmer not be responsible for the illegitimate use of force by the robot? In what follows, I focus on the programmer.

In her defence, the programmer could reply that she did not know that the Killer Robot would, by inappropriately relating 'lives' and 'costs', kill surrendering

combatants. But given the unpredictability of operationally autonomous technology, she should have considered that the Killer Robot may connect the two criteria in inappropriate ways. She should have also taken into account that, in the sensitive area of targeting, this could have potentially disastrous outcomes. And yet, she knowingly proceeded by giving the machine permission to balance lives against costs. It is hard to see why, in this case, there should be a responsibility gap. Rather, the question is whether the programmer violated a standard of care, and corresponding duties of care, by leaving calculations of lives and costs to the robot.

It is possible to arrive at the same conclusion via a different route. Sparrow draws a rough analogy between Killer Robots and child soldiers. Neither can be held responsible, he argues, for what they do in war.[18] The case of child soldiers has the potential to give rise to a responsibility gap because there could be circumstances in which those who deploy child soldiers cannot be held responsible for resulting wrongdoing by the children.

But this appears counterintuitive. Imagine that, in Green's aggressive war against Yellow, Commander Roberto, at t_1, orders a group of child soldiers to attack a group of Yellow's soldiers who are hiding in a village. Suffering high losses, the children, at t_2, eventually manage to overpower Yellow's soldiers. Enraged by the deaths of their friends, they vent their anger on the villagers by committing a terrible massacre at t_3. Could Commander Roberto claim that he is only responsible for the deaths of Yellow's soldiers at t_2 but not the massacre at t_3? After all, he did not order the children to kill the villagers. Moreover, once he had sent the children into the village at t_1, he did not have any control over their actions.

Roberto's attempt at justification is not successful. First, he should have known that it is risky to use children for military operations.[19] Considering the de-sensitisation that child soldiers undergo at the hands of their captors, they are likely to get out of control. Second, he provided them with the means – guns, alcohol, and drugs – that enabled them to carry out the massacre. Roberto's claim that he could not have stopped the children is irrelevant here. He voluntarily created a situation in which he ceded control to a group of unstable children. Consequently, it would be odd to claim that no one is morally responsible for the massacre. In fact, Roberto is not only morally responsible but culpable for ordering the children into battle. By imposing excessive risks on the villagers, he showed utter disregard for their safety. Contrary to what one would expect from a responsible military commander, Roberto wilfully ignored standards of due care.[20]

The above shows that notions of risk and standards of due care are central to debunking the responsibility gap. Because operationally autonomous machines are unpredictable, their use poses a risk. Just as Roberto is morally responsible for imposing risks on the villagers by deploying child soldiers, programmers of operationally autonomous weapons are morally responsible for imposing risks on individuals in conflict zones. True, the mere imposition of risk does not mean that programmers are automatically morally culpable when things go wrong. Roberto's use of child soldiers is unreasonably risky and shows a callous disregard for human life. Allowing an unpredictable robot to balance lives against costs may be equally risky. But there could be situations where the use of Killer Robots is not excessively risky and the programmer is not culpable but merely responsible for what the machine does. In any case, there is no responsibility gap.

If the analysis is correct, the implications of machine autonomy for moral responsibility are less severe than assumed. The action, in the debate on Killer Robots, should be on the site of risk, rather than responsibility. Critics of these weapons could simply argue that, at the current stage of technological development at least, it is not possible to formulate a sound standard of care for the deployment of Killer Robots.

Yet risk-based arguments against Killer Robots, though practically sensible, are philosophically rather weak. They amount to what I referred to earlier as a contingent argument. With further technological development, it might eventually become possible to formulate a sound standard of care for the deployment of Killer Robots. Naturally, this would render contingent, risk-based arguments against Killer Robots redundant. Thus, in the next part of the article, I discuss whether there are stronger, intrinsic arguments against Killer Robots.

4. Killing in War: Human Agency and Artificial Agency

In this part of the article, I outline an intrinsic argument against Killer Robots.[21] Some argue that the use of Killer Robots signals increasing moral disengagement in warfare.[22] According to this criticism, machines should not make decisions about life and death. This is because they have no appreciation of the value of life. The tension between the gravity of the decision to take human life and the morally disengaged way in which it is made may indeed explain why some feel uncomfortable with Killer Robots. Yet, as the above showed, the claim that machines 'make decisions' is not straightforward. According to one interpretation, which I referred to as the execution model, Killer Robots can potentially make decisions, but these always occur within the parameters of what I called a meta-targeting decision made by the programmer. What is morally significant about the execution model, I believe, is that, once the robot has been programmed, the programmer is at least one causal step removed from the execution of the meta-targeting decision: the actual point of force delivery, the act of killing. Put simply, no human is present when the robot pulls the trigger as humans have been taken out of the Kill Chain. As I observed at the beginning of the article, advocates of Killer Robots think that this is the best way to realise humanitarian objectives in armed conflict. In what follows, I argue that this removal of human agency from the Kill Chain in general, and from the point of force delivery in particular, amounts to partial moral disengagement. Contrary to the humanitarian argument in favour of Killer Robots, this counts as an intrinsic reason against operationally autonomous weapons technology.

In order to develop an intrinsic argument against Killer Robots, I operate with an ideal case in which all contingent worries have been resolved: Killer Robots can precisely navigate their way to their target, their deployment is not excessively risky, and a sound standard of care has been developed. Importantly, the orders they have been programmed with do not violate the principles of *jus in bello*. In particular, they do not violate the rights of potential human targets. This is because these human targets *qua* legitimate targets lack a right and a corresponding claim not to be attacked.[23] And Killer Robots will *only* attack legitimate human targets. If the deployment of Killer Robots does not wrong the target, and bystanders are not exposed to unreasonable levels of risk, why should it be problematic? Assuming that John is a legitimate target,

is there really a morally relevant difference between (1) ordering a soldier, Jack, to kill John and (2) programming a robot to kill John?

I think the answer lies in the issue of agency. As noted above, robots can be classified as artificial agents. Yet Killer Robots *qua* artificial agents lack a central component of human agency, the ability to do otherwise: *not* to shoot the target. Upon encountering John, the robot *will* attack John. True, according to the execution model, the robot may decide *how* and *when* to attack John. But it cannot decide *if* to attack John at all. After all, it has been programmed to do so. Compared to artificial agency, what makes human agency in warfare, and in ordinary life, valuable is the possibility of engaging in an alternative course of action. The ability of individuals to do otherwise is, of course, an important feature of the free will debate in philosophy.[24] For the purpose of this article, we can assume that for our ordinary practices of praising and blaming to work, and for us to be able to lead a social existence, we can presume that human persons are free and have the capacity to do otherwise.[25] This is true even in a situation of war, notwithstanding that armed conflict often pushes human agency to the limit. Jack, the soldier ordered to kill John, retains the capacity *not* to pull the trigger, even if he has been ordered to do so. The robot sent to kill John, however, does not.

This takes us back to the very beginning of the article. Michael Walzer may be right in arguing that wars are fought by states with human instruments. But these instruments are still human. By entering into war and fighting under the authority of the state, soldiers do not become automatons. If they did not retain the ability to engage or not engage in certain actions, Walzer's subsequent claim that soldiers are responsible for any atrocities they commit would not make sense.[26] War restricts human agency but does not extinguish it. The debate between orthodox and neo-classical just war theorists is essentially about the degree of human agency in war. The former are slightly more pessimistic in this regard than the latter, but both camps agree that human have at least some agency.[27]

But why should any of this matter? In our ideal scenario, programming the robot to kill John does not violate John's rights. John *qua* legitimate target lacks a claim not to be attacked. It seems also inappropriate to speak of moral disengagement. The programmer deploying the robot against John may have taken great care to minimise any risks and ensure that the robot only attacks John. The answer, I think, is that morality is not just about rights. This is why I speak of *partial* moral disengagement. Orthodox and neo-classical accounts of just war theory are predominantly concerned with rights, though they disagree about what rights individuals have in war. While rights are clearly important, they are not the *only* values that are morally relevant. Respect for rights may be necessary but insufficient for full moral engagement.

This argument harks back to broader critiques of right-based moralities. Right-based moralities are accused of neglecting other important moral considerations, and neither orthodox nor neo-classical just war theory may be different in this respect.[28] It is true that, in our ideal case, the rights of a legitimate target are, by definition, not violated when a Killer Robot executes a meta-targeting decision. But what about mercy and compassion?[29] It is harder to place these in a right-based moral framework, for sure – not shooting at a legitimate enemy target may, under certain circumstances, be morally desirable but it is not obligatory. The enemy *qua* legitimate target does not hold a claim against the attacking soldier not to kill him. But surely the human

capacity to have mercy with, feel pity for, or empathise with other humans, even if these belong to an opposed state, is morally relevant and worthy of protection.

In this respect, the issue of Killer Robots provides an interesting lens through which to view the current state of rights-based just war theory. Why is the focus on rights and rules unsatisfactory? Briefly, killing another person, even if doing so is morally permissible, is surely a grave act. The early just war theorists – perhaps because of their theological outlook – were deeply aware of this. The more recent language of rights does not quite manage to convey this. Because of the graveness of the decision to take human life, it seems desirable to retain the option not to kill, even if morally permissible. This point is corroborated by historical accounts of soldiers who refused to pull the trigger, even though (a) they had been ordered to do so and (b) it would have been permissible for them to do so.[30] Those soldiers who did not kill might have done so because they recognised the humanity of the enemy and realised the graveness of the decision to pull the trigger. Regardless of the wider political dynamic of armed conflict, the enemy about to be targeted is still a fellow human being with one life to live. Killer Robots, by contrast, do not relate to humans in the way humans relate to each other. They do not have the agency to put their gun down because 'they pity the poor sod' they have been programmed to target.[31] Notwithstanding centuries of human carnage, I think retaining human agency at the point of force delivery, thereby protecting the freedom not to pull the trigger, push the button, or throw a grenade, is essential for retaining our humanity in exactly the situation that challenges it the most: war.

Ultimately, what is bad about Killer Robots is that by removing the person deploying the robot at least one causal step from the point of force delivery, we remove an important and distinctive aspect of human agency (the ability to do otherwise) from a crucial moment in war (the act of killing). While this argument exposes a shortcoming of contemporary just war theory, orthodox and neo-classical just war theorists can both potentially agree to the rationale of the intrinsic argument and oppose Killer Robots on the grounds that they replace human agency. In particular, they could look for ways to complement their right-based frameworks with wider moral considerations of mercy, pity, and empathy outlined here. Indeed, Walzer's observations on the 'naked soldier' may be useful in this context.[32] The naked soldier is a legitimate target in the sense that killing him does not violate his rights. But, as Walzer points out, it may be inappropriate to kill him because he is defenceless, scared, or engaged in ordinary activities – smoking a cigarette – that represent a common bond, even between enemies.

This intrinsic argument does not have a technological fix. There are fundamental differences in the way in which robots and humans relate to each other. These differences are unlikely to be bridged via technological advances. In this sense, the intrinsic argument, as its name suggests, is not contingent on a particular stage of technological development. Moreover, even if differences between humans and machines could be transcended, it is questionable whether the military has an interest in doing so. Arguably, it is not in the interest of the military to create weapons that do not execute a legitimate targeting decision. To be sure, the same could be said about human soldiers. Military training seeks to desensitise soldiers to the act of killing. And yet, as history shows, it does not guarantee that soldiers will always pull the trigger. In general, any claim that militaries could, and should, turn soldiers into automatons that unquestioningly execute their orders should be met with scepticism. Firstly, it is

questionable whether such training is compatible with the moral rights of recruits. Soldiers may not have the same rights as ordinary civilians, and their agency may be more restricted than in other professions, but this does not mean that anything goes. Secondly, any attempt to turn soldiers into unquestioning automatons is likely to require extreme desensitisation to the act of killing. This raises serious doubts over whether soldiers, once entirely desensitised, would comply with the moral rules of war. For now, we can assume that differences between human and artificial agents remain sufficiently great so as to vindicate the intrinsic argument against Killer Robots.

5. Can, and Should, Human Agency Be Preserved in War?

The intrinsic argument is not uncontroversial. In the final part of the article, I outline and respond to four potential criticisms. The first criticism of the intrinsic argument is that it conflicts with duties of care.[33] As indicated in the introduction to the article, robotic weapons may be desirable because they enable the military to fulfil its duties of care towards their service personnel. Rather than sending soldiers into dangerous places, the military can send robots. By contrast, the intrinsic argument, in its attempt to preserve human agency, would demand that soldiers are sent into dangerous places, rather than robots. As a result, lives that could have otherwise have been saved are sacrificed on the altar of human agency.

While there is a tension between the intrinsic argument and duties of care this needs to be kept in perspective. Firstly, the intrinsic argument does not, in principle, oppose the development of remote-controlled weapons. Remote-controlled aerial vehicles ('drones'), for instance, enable militaries to remove soldiers from combat zones, reducing any physical risk to their lives. Given that drone pilots control the application of force to a target via remote-control, they retain, in principle, the capacity to exercise their agency by not 'pressing the button'.[34] This is not to suggest that 'killing by remote-control' does not create moral problems. But these have little to do with the intrinsic argument. The argument is opposed to 'killing by algorithm', not 'killing by remote-control'.

Secondly, the intrinsic argument does not oppose attempts to preserve human agency in robotic warfare.[35] For instance, it might be possible to give programmers the capacity to override the robot at the point of force delivery. In such a scenario, the programmer remains on standby and has the ability to intervene in the Kill Chain. Yet the question is whether this is a realistic prospect. In order for the programmer to change his mind, the robot would have to be equipped with a sensor suite comparable to those found in 'tele-operated' weapons, such as drones, and the programmer would need to be 'immersed' in the actual scenario. Such an immersion might not be possible since robotic warfare is likely to be conducted at speeds that transcend human (cognitive) capacities. Advocates of equipping Killer Robots with a standby/override function would have to show that the programmer's ability to stop the robot is a realistic prospect. There needs to be meaningful human control, and not just control on article.

These two responses may not satisfy those who cite duties of care as reasons in favour of Killer Robots. They could argue that while 'drones' protect their pilots from physical harm, they do not shield them from mental harm. Existing drones are

controlled via a video link.[36] This means that the operator sees the effects of the use of force. There have been reports that some operators develop post-traumatic stress disorders as a result.[37] The added benefit of Killer Robots is that they offer protection from mental harm. Once the robot has been programmed, it can be deployed without further human supervision. Unlike the drone operator, the programmer neither witnesses the use of force, nor its aftermath.

Such claims should give us pause for thought. If war is so awful that even professional soldiers, not just conscripts, must be granted *absolute* protection from *all* potential harms, one wonders on what grounds these very harms can legitimately be inflicted on others (potential enemies). To be sure, it does not necessarily follow that war is *never* morally permissible. But it suffices to say here that the justificatory scaffold will be fairly high – perhaps higher than just war theorists have traditionally assumed.[38] Moreover, one wonders whether duties of care requiring near absolute protection from harm – physical and mental – are defensible. For it is hard to see how, with such demanding duties of care, 'soldiers' could be required to 'fight': they would have to be completely insulated from any fighting activity. And yet, being a soldier consists in accepting some exposure to harm. Surely, this separates soldiers from civilians. Intuitively, such duties of care are over-demanding, and therefore unlikely to derail the intrinsic argument.

This leads us to the second criticism of the intrinsic argument. Critics could argue that the intrinsic argument has undesirable consequences: it leads to higher casualty rates than the deployment of Killer Robots. Soldiers have the ability to do otherwise, but there is no guarantee that they will put their weapons down. Far from it, they may be inclined to pull the trigger more often than Killer Robots. Would we really, in the name of human agency, reject a technology that reduces loss of life in armed conflict?[39]

To deal with this charge, it is useful to distinguish between 'absolutist' and 'threshold' conceptions of the intrinsic argument. An absolutist conception of the intrinsic argument would categorically deny that it is ever permissible to replace human with artificial agency in war, regardless of consequences. This position is vulnerable to the criticism mentioned in the preceding paragraph. A threshold conception of the intrinsic argument potentially allows trade-offs between human and artificial agency when a certain threshold of lives saved is met.[40] The threshold conception, to which I am sympathetic, cedes some ground to advocates of Killer Robots. But it also raises the bar for the case in favour of Killer Robots.

Firstly, given that human agency is valuable, the threshold conception would set the threshold for potential trade-offs fairly high – perhaps much higher than some supporters of Killer Robots may be comfortable with. It would need to be shown that the deployment of Killer Robots consistently leads to a dramatic reduction in casualties or wrongdoing in armed conflict. For a threshold to be reached we should expect nothing less than a transformation of warfare through this new form of technology.

Secondly, the threshold conception illustrates that, in the case for Killer Robots, it is not sufficient to point to the negative aspects of human agency, as some advocates in the debate frequently do. There is no doubt that human beings have done terrible things during armed conflict, but they are also capable of not pulling the trigger, out of compassion and mercy. This cannot simply be dismissed in the case *for* Killer Robots. The threshold for the deployment of Killer Robots must be high precisely

because the introduction of such weapons would not only quell (if successful) the negative aspects of human agency but also the positive ones. This shows that the case for Killer Robots is more complex than its advocates assume.

The threshold conception, then, cedes some ground. In a highly idealised scenario where the deployment of Killer Robots had a transformative impact on warfare, it is hard to insist on an absolutist version of the intrinsic argument. However, it is worthwhile pointing out that history indicates that humanitarian arguments in favour of certain weapons need to be viewed with a healthy dose of scepticism. Often the promise of new military technologies is that they will reduce causalities. Indeed, many precision-guided technologies have been welcomed for humanitarian reasons. And yet, casualties remain depressingly high, especially amongst those who are not classifiable as legitimate targets. For instance, during Israel's Operation Protective Edge in 2014, in which precision weapons were used, it is estimated that 551 Palestinian children were killed.[41] Over 3000 Palestinian children were injured, 1000 of which were left permanently disabled. Further, the use of drones, heralded by many as a major advance in precision-strike technology, during US counterterrorism operations in Pakistan is estimated to have led to over 1000 civilian casualties.[42] This is not the place to debate the rights and wrong of those conflicts. The two examples merely illustrate that military technology is no panacea.

If experience is anything to go by, then humanitarian arguments in favour of Killer Robots must be taken with a pinch of salt. In particular, there are two reasons why Killer Robots may not have the positive effect on casualty rates their advocates hope for. Firstly, the development and production costs of such robots will be immense. This means that the military will not be inclined to take risks that may lead to the loss of a Killer Robot. In order to avoid destruction, the robot might not only be programmed with a particular target but also a wide range of defensive measures. Consequently, it would take defensive steps, including the use of force, against those who resist it.

Secondly, and directly related to the preceding point, many discussions of Killer Robots are Western-centric. Perhaps unconsciously, they assume technological asymmetries between those states which, in the future, are potentially capable of deploying Killer Robots and their opponents. Humanitarian concerns may be met as long as there are such asymmetries. The problem, though, is that opponents could catch up by developing effective anti-robot weaponry. In such cases, it is unlikely that the programming of Killer Robots would be particularly restrictive. Killer Robots would probably be programmed to shoot immediately upon contact with enemy troops, otherwise the risk of destruction is too high.

Taken together, these two points raise doubts about whether Killer Robots, in reality, would lead to a considerable decline in casualties. If this is true, the replacement of human agency with artificial agency in warfare is unlikely to have the desired humanitarian effect, let alone meet a threshold where trade-offs between the two would be warranted.

This leads me to the third criticism. It contends that the intrinsic argument offers too little, too late. Even without the development of Killer Robots, modern military technology and warfare have already had a highly negative impact on human agency.[43] Just as the medieval concept of chivalry has become out-dated due to the rise of the modern nation state, considerations underpinning the intrinsic argument have lost

their relevance in contemporary armed conflict. Consider a pilot who flies a high-altitude bombing mission. Do we really think that, in the split second the pilot has to release his payload, he could have done otherwise? Or consider a young artillery operator who is ordered to fire an artillery shell at some coordinates. Does the young gunner realistically have the capacity not to fire the shell? Lastly, consider a drone pilot who 'tele-operates' an uninhabited airplane. As we just saw, the intrinsic argument does not oppose 'killing by remote-control', but one wonders how much agency a drone pilot really has. This is because drones enable much better monitoring of pilots by their commanders, thus restricting their ability not to push the button. The three examples reinforce the earlier point about control. Just how much meaningful human control is there in contemporary armed conflict?

Those who sympathise with the intrinsic argument can respond to this criticism in two ways. On the one hand, they could argue that proposals for the development of Killer Robots are symptomatic of a form of warfare in which humans have been reduced to cogs in a machine. The advent of Killer Robots illustrates that human agency is not valued anymore in armed conflict. Not surprisingly, the space for individuals to exercise agency in war becomes ever more restricted. If this is sound, a concern for human agency may necessitate an endorsement of something like contingent pacifism. According to such a position, there could, in principle, be wars in which individuals retain some agency, yet contemporary warfare falls short in this respect. Hence one should oppose contemporary war – or so one could argue.

On the other hand, it is possible to contend that, while it is true that human agency has been undermined in contemporary warfare, Killer Robots remain morally distinctive. In the above examples, the agency of the bomber pilot, gunner, and drone operator is restricted. But they retain the ability to do otherwise, though the chance they might do so is fairly small. By contrast, the distinctive feature of Killer Robots is that they eliminate all human agency from the Kill Chain. There is, one could argue, a morally relevant difference between restricting agency and eliminating it. Once agency is eliminated, there is zero chance of doing otherwise; alternative actions do not even remain a remote possibility. For this reason, the intrinsic argument is right to reject Killer Robots, though this does not mean that one could turn a blind eye to other technological developments in contemporary warfare. In light of some of the above trends, there is a lot to worry about.

The fourth criticism of the intrinsic argument is that it undermines military authority. This is because it allows soldiers to disobey *legitimate* orders.[44] Indeed, the intrinsic argument illustrates a tension between the hierarchical structure of the military and the conscience of the individual soldier. The legitimate order issued to James to kill a legitimate target, Pete, may conflict with James' sense of compassion or mercy towards Pete. In cases like these, the intrinsic argument contends that soldiers should follow their conscience, precisely because the taking of life, even if legitimate, remains a grave act. The question is whether this would undermine military authority, and ultimately the effective functioning of the military. If it did, it might be better to use robots, rather than soldiers: they get the job done without any qualms. In defence of the intrinsic argument, it is worthwhile pointing out that it does not imply that soldiers should always show pity or mercy. It may be inappropriate, for instance, to show pity or mercy towards members of a genocidal militia about to commit a massacre. The intrinsic argument only maintains that there are situations, such as Walzer's

aforementioned 'naked soldier' scenario, which are less clear-cut. It is because of these types of scenarios that we must not lose our sense of humanity, pity, or mercy in war, no matter how justified a particular order is. This is unlikely to undermine the effective functioning of the military.

In sum, the intrinsic argument can be defended against these four criticisms. The possibility that, at the point of force delivery, human agency may be replaced by artificial agency should make us deeply uneasy about design proposals for battlefield robots. Claims that robots can absolve us from confronting the hard, messy and essentially human choices that individuals have to make in war, and that we so often and regrettably get wrong, should be treated with deep suspicion. Killer Robots will not rescue us from the human condition.

Concluding Remarks

This article offers four contributions to the debate on Killer Robots, and contemporary just war theory more generally. Firstly, I introduce a distinction between contingent and intrinsic arguments against this type of weapons technology. This is important because it has a filtering effect, clarifying arguments currently raging in the debate. It also offers a lens through which we can assess the normative strength of particular arguments against Killer Robots. Contingent arguments are necessarily weaker than their intrinsic counterparts, because they can be overturned by future developments in technology. Intrinsic arguments, such as the one I presented relating to the removal of human agency at the point of force delivery, represent fundamental problems that are unlikely to be resolved technologically.

The second contribution is the distinction between generating and executing a targeting decision. If machines were capable of generating targeting decisions without assistance from an operator, Killer Robots would indeed be a game changer. But I refuted the claim that machines can generate targeting decisions by applying normative principles themselves. I argued that Killer Robots are best seen as highly technologically sophisticated executors of targeting decisions. However, this does not mean that the execution of a targeting decision by a Killer Robots is morally unproblematic.

Thirdly, I have debunked the claim that the assignment of moral responsibility is an overly problematic issue in robotic warfare. Those who deploy Killer Robots impose considerable risks on others, and they are responsible, though not necessarily culpable, for these risks. Rather than focusing on the issue of responsibility, the debate should pay greater attention to questions of risk in warfare. Killer Robots, like many other forms of technology, are not 100% fail safe. The question is whether the risks posed by them are so severe that their deployment would be unreasonably risky.

This takes me to the final and main contribution of the article, which is an intrinsic argument against Killer Robots. Even if we imagine an ideal scenario in which Killer Robots function perfectly – their deployment does not violate any rules of warfare and the risks arising from their use are contained within reasonable limits – the introduction of Killer Robots represents a normatively unwelcome shift in warfare. Crucially, Killer Robots remove human agency from the point of force delivery. In other words, they remove the possibility of *not* pulling the trigger. We need to leave space in warfare for pity, compassion and empathy, and the ability to put one's gun down; otherwise

we truly risk losing humanity in warfare. Sometimes, in war, *not* pulling the trigger might be the right, and humane, thing to do.

Alex Leveringhaus, Department of Politics, University of Sheffield, Elmfield Building, Sheffield, S10 2TU, UK. alexander.leveringhaus@manchester.ac.uk [Correction added on 8 September 2017, after first online publication: Author's email address is updated in this version.]

NOTES

1 Research for this article was financed by a grant from the Netherlands Organisation for Scientific Research (NWO), which is gratefully acknowledged. I also thank the journal editor and two anonymous referees for their helpful comments. Early versions of this article were presented at the universities of Liverpool, Surrey, Aarhus, Frankfurt a.M., as well as Goldsmiths College London, the London School of Economics and Political Science, Chatham House, Europa University–Viadrina/Frankfurt/Oder, Changing Character of War Programme/University of Oxford, and the Institute for Peace Research and Security Policy/Hamburg. I am grateful for the comments I received on these occasions. Special thanks go to Tjerk de Greef for our discussions of technological issues in military robotics. I am also immensely grateful to Nicola Leveringhaus, Bernhard Koch, Seumas Miller, and the members of the War Workshop at All Souls College Oxford for discussing earlier drafts of the article with me. This article draws upon some of my other writings on robotic weapons, most notably A. Leveringhaus & T. de Greef, 'Keeping the operator in-the-loop: A qualified defence of autonomous weapons' in T. Dyson, W. Aslam & M. Aaronson (eds) *Precision-strike Technology and Intervention: Strategic, Legal and Moral Implications* (London: Routledge, 2014), and Alex Leveringhaus, *Ethics and Autonomous Weapons* (Basingstoke: Palgrave Macmillan, forthcoming 2016).

2 Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations* (New York: Basic Books, 2006), p. 36.

3 To my knowledge, Robert Sparrow first used the term in philosophical debate in his seminal paper 'Killer Robots'. My debt to Sparrow's work is clear. See Robert Sparrow, 'Killer robots', *Journal of Applied Philosophy* 24,1 (2007): 62–77.

4 See Peter W. Singer, *Wired for War* (New York: Penguin, 2009).

5 Humanitarianism is a complex doctrine. In its basic form it entails a commitment to the reduction of human suffering. For some of the intricacies surrounding humanitarian ethics, see Hugo Slim, *Humanitarian Ethics: A Guide to the Morality of Aid in War and Disaster* (London: Hurst, 2015).

6 Ron Arkin, *Governing Lethal Behaviour in Autonomous Robots* (Boca Raton, FL: Taylor and Francis, 2009).

7 I shall not take sides in this debate here. On neo-classical just war theory, see Cécile Fabre, *Cosmopolitan War* (Oxford: Oxford University Press, 2012) and Jeff McMahan, *Killing in War* (Oxford: Oxford University Press, 2009). On orthodox just war theory, see Walzer op. cit. For an overview of the debate, see David Rodin & Henry Shue (eds), *Just and Unjust Warriors* (Oxford: Oxford University Press, 2008).

8 For a defence of the impermissibility of weapons research, see John Forge, *Designed to Kill: The Case Against Weapons Research* (Amsterdam: Springer, 2013). I am critical of this view, see A. Leveringhaus, 'Book review essay: *Designed to Kill* by John Forge', *Australasian Journal of Philosophy* 92,2 (2014): 394–397.

9 Just war theorists disagree about how to distinguish between legitimate and illegitimate targets, especially with regard to human persons. Orthodox theorists usually argue that combatants are legitimate targets, while civilians, or non-combatants, are illegitimate targets. Neo-classical theorists, by contrast, argue that only individuals who have rendered themselves liable to attack are legitimate targets. Contrary to orthodox theory, civilians may be liable to attack, while just combatants are not liable to attack. To avoid taking sides in the debate, I use the term legitimate (human) target throughout.

10 The relationship between proportionality and necessity is contested. Some traditional versions of just war theory treat the two criteria as separate. One could, however, argue that unnecessary military acts are always excessive and thus disproportionate. I do not take a view on the issue here. I follow the more traditional account of *jus in bello* because it is still widely used. For an excellent exploration of necessity and proportionality, see Larry May, *War Crimes and Just War* (Cambridge: Cambridge University Press).

11 I follow the British roboticist Alan Winfield here; see Alan Winfield, *Robotics: A Very Short Introduction* (Oxford: Oxford University Press, 2013). I shall not touch upon the issue of 'software robots' and their use in what some call 'cyber warfare' here.
12 Immanuel Kant, *Groundwork of the Metaphysics of Morals*, revised edition (Cambridge: Cambridge University Press, 2012).
13 I should emphasise the narrow focus on the use of force here. Killer Robots, of course, could be operationally autonomous with regard to functions not directly related to targeting, such as flying, refuelling, navigating etc. The philosophical and policy debates on robotic weapons are only concerned with operationally autonomous uses of lethal force by a robot. It is important to realise that this leaves plenty of scope for the use of operational autonomy in weapons systems for tasks other than targeting.
14 On the complexity of proportionality judgements, see Thomas Hurka, 'Proportionality in the morality of war, *Philosophy & Public Affairs* 33,1 (2005): 34–66.
15 For a longer version of this argument, see A. Leveringhaus & T. de Greef op. cit.
16 Sparrow op. cit; Andreas Matthias, 'The responsibility gap: Ascribing responsibility for the actions of learning automata', *Ethics and Information Technology* 6 (2004): 175–183.
17 It goes without saying that the machine is not morally responsible for what happens either, though Daniel Dennett seriously considers holding machines morally responsible. His argument, however, focuses on the fictional computer HAL from Stanley Kubrick's classic science fiction movie *2001 Space Odyssey*. The machines that will be developed over the next few decades will be nothing like HAL. Generally, it is not clear what would be gained by holding machines morally responsible. We could simply re-programme them, send them error messages, or switch them off. Needless to say, we can't simply re-programme or 'switch off' humans. This is one of the reasons why human societies have developed practices of praising and blaming. See Daniel Dennett, 'When HAL kills, who's to blame? Computer ethics' in D.G. Stork (ed.) *HAL's Legacy: 2001's Computer as Dream and Reality* (Cambridge, MA: MIT Press, 1997), pp. 351–365.
18 This is a controversial point, see McMahan op. cit., pp. 198–202.
19 To avoid misunderstanding, the main reason why it is wrong to use children in war is that turning children into soldiers harms them.
20 On the concept of due care, see Walzer op. cit., p. 156.
21 For a slightly different take on intrinsic arguments against autonomous see Bradley Strawser, Duncan Purves & Ryan Jenkins, 'Autonomous machines, moral judgements, and acting for the right reasons', *Ethical Theory and Moral Practice* (forthcoming).
22 See Noel Sharkey, 'Saying 'no!' to lethal autonomous targeting', *Journal of Military Ethics* 9,4 (2010): 369–383.
23 I follow Wesley Hohfeld's concept of claim-rights here. The right not to be attacked usually entails a correlative claim or obligation for another party not to attack the right holder. Legitimate targets in war lack such a claim. That is, they cannot hold another party under a duty not to attack them. Orthodox and neo-classical just war theorists disagree on which grounds a person becomes a legitimate target. The ideal case in which a robot is programmed to kill John can be conceived along neo-classical or orthodox lines. The general argument with regard to human agency holds, regardless of whether one follows the orthodox or neo-classical approach. On Hohfeld's theory, see Peter Jones, *Rights* (Basingstoke: Palgrave, Macmillan, 1994).
24 See Harry Frankfurt, 'Alternate possibilities and moral responsibility', reprinted in P. Russell & O. Deery (eds) *The Philosophy of Free Will: Essential Readings from the Contemporary Debates* (Oxford: Oxford University Press, 2013), pp. 139–148.
25 I follow Peter Strawson here; see 'Freedom and resentment' reprinted in P. Russell & O. Deery (eds) op. cit., pp. 63–83.
26 Walzer op. cit., p. 39.
27 The issue of agency, I contend, is at the heart of the dispute about the moral equality of soldiers, though this is not always clearly expressed in the relevant philosophical debate. Walzer defends the moral equality of soldiers because he thinks that soldiers cannot be held responsible for participation in war (just or unjust). The main reason for this is that their agency is compromised since the state coerces them into fighting. McMahan thinks that the moral equality of soldiers fails because individual agency and moral responsibility is unaffected by war. There is no difference, McMahan opines, between individual agency in war and individual agency in peacetime. As a result, for McMahan, the ethics of armed conflict is analytically reducible to the ethics of private self-defence. Walzer disagrees.

28 Joseph Raz, 'Right-based moralities' in J. Waldron (ed.) *Theories of Rights* (Oxford: Oxford University Press, 1984), pp. 182-200.
29 I thank Cécile Fabre for discussing this point with me.
30 Interestingly, in World War II, few soldiers actually pulled the trigger. Since then, the American and British militaries, in particular, have developed extensive training techniques in order to lower the inhibition against killing in their soldiers, to the effect that soldiers pull the trigger more often. However, while the rates may have changed in favour of pulling the trigger, some soldiers still refuse to do so. See Samuel (S.L.A.) Marshall, *Men Against Fire: The Problem of Battle Command*, new edn. (Norman, OK: University of Oklahoma Press, 2000, original edition 1947) and Dave Grossman, *On Killing: The Psychological Cost of Learning to Kill in War and Society*, revd. edn. (New York: Back Bay Books, 2009). I thank Thomas Simpson for discussing this point with me.
31 The phrase is associated with George Orwell, and used by Walzer in *Just and Unjust Wars*. See Walzer, op. cit., p. 36.
32 Walzer, see op. cit., pp. 138–144.
33 Discussion of this first criticism benefitted from a wider discussion held at Imperial College London in November 2015. I thank the participants.
34 Remotely piloted aeroplanes are only the tip of the iceberg when it comes to uninhabited weapons technologies. The push towards uninhabited weapons can be seen across all branches of the military. Unmanned underwater vehicles, for instance, are currently being developed for the navy.
35 The US military, for instance, distinguishes between in-the-loop, on-the-loop, and out-of-the-loop systems. I represent these distinctions here against the background of the execution model. This model is not endorsed by the US military.

    1. *In-the-loop*: The operator makes and executes a targeting decision.

    2. *On-the-loop*: The operator makes a targeting decision and the Killer Robot executes it, though the operator has the ability to override the execution process.

    3. *Out-of-the-loop*: The operator makes a targeting decision and the Killer Robot executes it, without giving the operator the opportunity to override the execution process.

    Strictly speaking, the intrinsic argument against Killer Robots is aimed at out-of-the-loop-systems. On-the-loop-systems might not be as problematic, since the operator has the ability to prevent the robot from pulling the trigger, at least in theory. For the US Army's definition of autonomous systems, see Department of Defense, *Unmanned Systems Integrated Roadmap FY 2011-2036* & *Unmanned Systems Integrated Roadmap FY 2013-2038*. Available online at http://www.defense.gov/pubs/DOD-USRM-2013.pdf (accessed 9 July 2014).
36 For a discussion of remote-control and drones, see A. Leveringhaus & T. de Greef, 'Tele-operated weapons systems: Safeguarding moral perception and responsibility' in M. Aaronson & A. Johnson (eds) *Hitting the Target: How New Capabilities are Shaping International Intervention* (London: Royal United Services Institute Whitehall Papers, 2013), pp. 57–63.
37 N. Abe, 'Dreams in infrared: The woes of an American drone operator, *Spiegel Online*, international edn. (2012). Available online at http://www.spiegel.de/international/world/pain-continues-after-war-for-american-drone-pilot-a-872726.html (accessed 9 July 2014).
38 For instance, one could argue that the use of force is only permissible to halt extreme atrocities, such as genocide. I thank one of the anonymous referees for bringing up this issue.
39 I thank William Boothby for raising this issue.
40 The threshold conception of the intrinsic argument draws, roughly, upon Frances Kamm's idea of 'threshold deontology'. Once a certain threshold is reached, deontological side-constraints may be relaxed. See F.M. Kamm, *Intricate Ethics* (Oxford: Oxford University Press, 2007).
41 Save the Children, *A Living Nightmare: Gaza – One Year On*. Available online at http://www.savethechildren.org.uk/sites/default/files/docs/A_Living_Nightmare.pdf (accessed 13 July 2015).
42 See *Guardian UK* (online edition). Available online at http://www.theguardian.com/us-news/2014/nov/24/-sp-us-drone-strikes-kill-1147 (accessed 13 July 2015).
43 I thank Laurence Lustgarten for pressing me on this point.
44 Since the Nuremberg Trials, soldiers are morally and legally required to disobey orders that do not violate the rules of war. The intrinsic argument goes further than this. For a philosophical treatment of the impact of Nuremberg on law and ethics, see Larry May, *Crimes against Humanity: A Normative Account* (Cambridge: Cambridge University Press, 2005), especially chapter 10.