Autonomous Weapons and the Ethics of Artificial Intelligence
Prof. Peter Asaro, *The New School*
Ethics of Artificial Intelligence, Oxford Press.

While "killer robots" have long been a staple of science fiction dystopias, they also represent a critical and central issue in the ethics of artificial intelligence (AI). More technically speaking, autonomous weapons are a real and emerging technology that has the potential to radically transform warfare, policing, and how we understand human rights in relation to the operations of machines, algorithms, and artificial intelligence. The issues raised by giving machines the capability, and more importantly, the *authority* to kill human beings raises a range of ethical issues, as well as legal, social and political issues. Many of the issues raised are of critical importance even if we consider only simple forms of automation, or *artificial stupidity*. Other issues arise if we consider the difficulty of properly gauging the capabilities and reliability of increasingly sophisticated forms of AI. And yet other issues arise if we consider the remote advent of some form of an artificial general intelligence (AGI), human-like AI, or super-intelligence. Because the issues raised by simple autonomous weapons are the most urgent, I will focus on these. But I will also consider some of the issues raised by increasingly capable systems, and reflect on the implications of highly capable future AI.

In considering a few of the most significant issues raised by autonomous weapons, I will seek to articulate them according to some major philosophical approaches to ethics. As such, I will not endorse any particular meta-ethical approach. Roughly speaking, my approach is that where there is broad agreement that moral rights and duties exist and are clear, they provide reasons that are more compelling than utilitarian reasons, while utilitarian reasons are useful in the absence of clear moral duties and rights. Further, I believe that moral virtues and sentiments reflect psychological and cultural norms and preferences and often function as heuristics in moral reasoning, especially when one must choose between competing duties and values, and when one reflects on the implications for one's own moral character when taking an action. While no single moral theory alone can fully explain our views of autonomous weapons, each of the leading Western moral theories–deontological and consequentialist–points to the immorality of autonomous weapons in its own way, and taken together the leading moral theories present a clear case that building and using autonomous weapons, and permitting or authorizing autonomous violence, is morally wrong.

1. Defining Autonomous Weapons

Let us start by considering a simple working definition of what constitutes an "autonomous weapon." Modern weapons and weapons platforms utilize a great deal of automation in the operation of various functions of the system, and at various levels of control. For instance, a guided missile contains feedback regulators over the direction of thrust so as to direct the missile toward some target. In a sense, the system is guided toward a "goal" or "target" according to some control mechanism and a sensor (heat, electro-magnetic, or optical), or some coordinate guidance system, such as satellite-based global-positioning systems (GPS). The missile is

assigned its specific "goal" by a human operator, who "locks" it on target using a laser or radar system, or provides GPS coordinates. The automation then follows some set of control parameters such that it advances towards its goal, and adjusts its controls to keep the missile on course until it strikes its designated target.[1]

Remotely-piloted vehicles, better known as drones, also contain a great deal of automation. With what amounts to sophisticated auto-pilot systems, these drones can be given preset flight paths, or a series of GPS way-points, and automatically fly to them. Along the way, they make numerous automated flight-surface and throttle control adjustments, to compensate for wind, thermals, aberrant sensor data and more. Some advanced drones are also capable of automated take-off and landings, automated aerial refueling, and researchers are regularly achieving increasingly sophisticated automated maneuvers. Of course, some drones carry the same missiles and bombs found on other military aircraft, which can themselves have complex automated guidance systems. They are thus considered "weapons platforms" for these weapons, containing multiple levels of control.

However, neither guided missiles nor remotely-operated drones are autonomous weapons in the relevant technical sense that concerns us. Following the working definition offered by the International Committee of the Red Cross, an autonomous weapon system is any system that automates the critical functions of targeting and engaging a weapon (ICRC 2016). This means that the targeting and use of force must be automated for the system to be considered as an autonomous weapon. Another way of looking at it is that autonomous weapons systems *lack meaningful human control over the critical functions of targeting and use of force*.

Whether such systems already exist depends on how one interprets *meaningful human control*, a topic to which I will return later. A variety of existing systems use some form of automated targeting. In particular there are mines and booby-traps that are victim-activated by pressure or proximity sensors, or even acoustic sensors; sentry guns that are similarly victim-activated by motion sensors; loitering munitions that search out specific radar signals over large areas; and a host of projectile intercept systems that automatically track and target incoming missiles and mortars and shoot them down. Most missile defense systems[2] are only autonomous for a few seconds at a time, however, and remain under the direct control of humans who can observe its operation, scrutinize its targets, and deactivate it at any moment. So they could be argued to be under meaningful human control (or *not,* if one further requires strict individual target authorization), while wide-ranging loitering munitions, mines and sentry guns systems are more problematic. This issue partly turns on the ethical aspects of such targeting, so we will return to it later.

[1]One should be careful in the use of such anthropomorphic terms as "goal" and verbs like "seeks." In humans these terms imply a kind of intentionality that, so far, machines are not capable of. This was the subject of much debate in early cybernetics, as well as in functionalist theories of the mind in philosophy (Asaro 2015).

[2]Examples of such systems include the U.S. Patriot missile system, Phalanx and Close-In-Weapons-Systems, and the Israeli Iron-Dome system.

In the case of remotely operated drones, a human (albeit from very far away) interprets the imagery data, identifies potential targets, verifies and selects a target, and then aims and engages a weapon. The drone operator may also consult human intelligence analysts, lawyers, and seek authorizations from superior officers. Typically, such armed drones employ a laser that they shine on the target–and automation helps to ensure that it stays on that target–which the sensors on the drone's missile can use to guide it on a path to the designated target. Thus, while such systems use automation, they are not autonomous weapons.

If, however, the drone utilized automated software to scan through its video feeds and sensor data, automatically identified targets, and then selected and fired on those targets, all without human intervention, supervision or control, then that would clearly be an autonomous weapon. From a causal or functional perspective, this may not seem like a large or significant difference–just a bit more automation, or automation in a different stage of the operation of the system. But, of course, the difference is one with ethical, as well as legal and political, significance. While there are different ways to frame the operations of these systems in moral terms, I will argue that automating the targeting of weapons and the use of violent force necessarily has ethical and moral significance, and should be recognized as such. Further, I will argue that as the development of such systems has become technically feasible, we should recognize existing moral and legal principles and establish new norms that clearly prohibit delegating the authority to kill to machines.

Another way of looking at such a norm is as a positive obligation to ensure meaningful human control over the use of weapons. Of course, the concept of "meaningful human control" will require some articulation, but the basic idea is that morally and legally responsible human agents must retain control over the functions of any system that directs and releases violent force. But it is also true that a robust conception of meaningful human control could find useful applications to other problems in AI ethics. Such applications include the relationship between self-driving cars and their occupants/operators, and in the application of algorithms to decisions with the potential to deny or deprive humans of their fundamental rights, from access to medical care, credit, educational and employment opportunities, to exercising their economic, cultural and political rights. In all of these cases, there is a potential to delegate an authority to a machine which might directly impact on the moral and legal rights of a human person.

The Moral Problems Raised by Autonomous Weapons

In providing a moral analysis, it will also be helpful to lay out the various types of issues that have been raised as problems with autonomous weapons. These represent a range of different concerns, and can potentially be characterized differently under different moral theories. Yet, each set of concerns also lends itself to one or more moral approaches. The concerns can be grouped together into some broader categories: harms to civilians, arms races and international instability, intrinsic unpredictability, hacking and cybersecurity threats, a new type of weapon of mass destruction, threats to responsibility and accountability, and threats to human rights and dignity. Along the way we will also consider arguments that autonomous weapons and their use might be morally superior to human-controlled weapons.

Harms to Civilians

By far the most commonly expressed concerns around autonomous weapons are that they will kill innocent civilians and destroy civilian infrastructure. Such a concern may seem quite simple and straightforward, but there are different ways to characterize this worry. Following the formulations of international humanitarian law, which requires military attacks to be discriminate and proportionate, one could argue that autonomous weapons will be indiscriminate in their targeting, failing to distinguish civilians from combatants. One could also argue that autonomous weapons might make disproportionate attacks, killing many civilians for a relatively low-value military objective. One could alternatively argue that autonomous weapons would lack aspects of human psychology that might make them more humane in warfare. They might thus be far more aggressive, or fail to show any compassion in situations where a human might be merciful. Worse, autonomous weapons could be easily designed, altered or manipulated to purposely harm civilians (*i.e.* given such a goal either explicitly or implicitly). Despots and tyrants might turn such weapons against their own people, or apply them to genocidal ends, or terrorists might use them to attack civilians. Despite the various ways autonomous weapons might cause negative impacts on civilians, it is possible to group these concerns together.

On initial consideration, this looks like a consequentialist concern–there will be significant negative consequences for civilians if autonomous weapons are deployed. Of course, it could also be viewed as violation of the rights of those individuals, and thus a deontological concern, that we will consider shortly. But in a consequentialist evaluation, whether the use of autonomous weapons is morally good or bad depends on numerous empirical facts about the actual impacts, and the probabilities of those impacts, which are difficult to assess before such weapons are used. Indeed, the proponents of developing autonomous weapons often argue on the same consequentialist grounds that autonomous weapons could be designed to be far better than humans at making targeting decisions and conducting attacks, thus reducing the risks of harm to civilians (Arkin 2013). From an engineering perspective, these negative consequences can also be viewed as "risks" and systems can be designed to try to minimize and eliminate such risks. This kind framing sets up the design of autonomous weapons as a form of safety design–maximize killing "bad guys" while minimizing the killing of "good guys." But while this might be reasonable as an argument for reducing unintended and undesired killing, it does not fully address the morality of the intended automated killing.

Upon further reflection, one can also look at this as a deontological issue–the negative impacts on civilians in these situations will be death, grave injuries, trauma, and displacement. While these are, of course, bad consequences, they are also the deprivation of fundamental human rights–the rights to life, bodily integrity, and dignity. Under this view, we have a moral duty to respect the rights of others, and to treat them as ends in themselves. But autonomous weapons could prevent us from performing our duty to respect others in war. Deontological ethics does not completely prohibit killing, particularly in war, but does require that there be a valid justification, such as self-defense, for killing. Similarly, international law permits the unintentional killing of civilians in war, provided there was a lawful (justified) military objective for an attack. But it is less clear what counts as "intentional" or "unintentional" when it comes to autonomous systems–the operator may have an idea of what an autonomous system will do, but

it may do many things they did not specifically consider, and kill people the operator did not intend to kill. In most circumstances we view such systems as accidents, and only hold people morally responsible if they were acting in a reckless or negligent manner. Autonomous weapons, particularly those involving advanced artificial intelligence, can serve to cloud the moral issue, however, insofar as the operator of a weapon might believe that the system is designed to only attack valid military targets, and to avoid and protect civilians. If this is a reasonable belief, then we might be inclined to attribute any harms to civilians as simple accidents or technical failures, and not the moral responsibility of the operator.

More critically, there is a question as to whether the use of an autonomous weapon might be a means of fulfilling one's moral duty to respect the rights of others, or actually precludes the ability to respect those rights. As Sparrow (2016) demonstrates, it could be argued that if there is a weapon system designed to protect civilians, and which actually works in that way, we may have a moral obligation to use it. While I see the consequentialist side of this argument, as discussed above with respect to Arkin (2013), I do not see the deontological side of it. In particular, in order to fulfill our duty to respect the human dignity of others, I believe it is required to recognize them as human, and to consider them as such when making the decision that it is justified to kill them, or put them at risk of death. Insofar as this consideration is not actually taken by the operator of the weapon, but they instead rely upon an automated process, then they are not really fulfilling their duty. Indeed, they do not necessarily even think about the individuals that may be killed, much less regard them as persons. And as we will see, neither does the automated system. This has implications for both the respect given to human rights and dignity, and to how we ascribe moral responsibility and legal accountability.

Arms Races, Rapid Proliferation & Instability

Another broad range of issues concerns the impact of the introduction of autonomous weapons in the context of international relations. Insofar as these weapons are seen as high-tech and prestigious, as well as providing tactical or strategic advantages over the capabilities of adversaries, or serve as an effective deterrent, there will be strong incentives for countries to develop or obtain such weapons. The same logic, of course, applies to their adversaries and competitors. This is the logical foundation of a competitive arms race wherein rivals expend large amounts of resources in an effort to gain military advantage over their competitors (Asaro 2019). Apart from being an expensive use, or waste, of economic, intellectual and natural resources, such arms races are tied to political and military instability.[3] Since significant military build ups and strategic advantages are viewed as threatening to neighbors and adversaries, some states may view pre-empting such advantages as better than allowing them to develop. As such arms races can raise tensions and create instability. Having access to new hi-tech weapons, especially ones untested in real conflicts, can also give leaders a sense of having superior

[3]There is some debate in political science as to whether arms races are the cause, or result, of political instability. I am inclined to view such arms races as involving positive feedback loops, wherein small instabilities lead to small arms build-ups, which lead to greater instability and great build-ups. See (Asaro 2019).

military capabilities, which in turn makes them more inclined to initiate or escalate military actions–whether or not their confidence is warranted. And insofar as the weapons themselves may behave or perform in unexpected ways due to artificial intelligence, they become less predictable as threats by adversaries leading to greater instability. Such arms races and rivalries can operate at regional levels between neighboring states, or at global levels between superpowers and groups of aligned states.

Again, from a consequentialist perspective, whether such arms races are good depends on one's evaluation of the outcomes, as well as how such rivalries play out. Some might argue that the Cold War was an arms race that ended in a stalemate of sorts, or *detente*, and was preferable to war. Others might argue that the Cold War led to numerous proxy wars, and that there were many other ways this rivalry might have played out, short of war, that did not require massive investments in nuclear arms and their delivery systems, and could have had much better outcomes. If we examine the wider set of outcomes and their probabilities, including the possibility of nuclear war and its history of near-misses, what we see is that instability is itself undesirable in international relations as it is a leading factor in many conflicts, including "low-intensity" and proxy conflicts. Instability makes it harder to predict how one's adversaries might act, as well as what the severity of threat, which makes it more likely that one takes a defensive or proactive stance. Insofar as such build-ups are viewed as intrinsically or implicitly hostile, the very existence of an arms race is a manifestation of hostilities between adversaries. In short, such instability increases the probability of violent conflicts occurring, which is bad on utilitarian grounds.

Closely related to concerns over the arms races between states, is the concern that autonomous weapons will proliferate rapidly. From the perspective of great power states, there is a concern that smaller states might rapidly acquire significant military capabilities to challenge larger wealthier militaries. Because autonomous weapons do not require the industrial and technical sophistication that nuclear weapons do, and simple ones can even be built with off-the-shelf technologies, many states will acquire them rapidly. We have seen this already with surveillance drones, and now increasingly with armed drones. As we have also seen with drone technology, autonomous weapons, including the advanced, sophisticated, and hardened types that major militaries would develop, would also find their way to non-state actors, terrorists, or even be acquired by police forces. At least with the more sophisticated versions, they are only likely to be developed by those with significant resources, and thus a strong stigmatizing norm against autonomous weapons could prevent such systems from being developed, or produced in significant numbers.

One could also consider the wider implications of the drain on resources such arms races will entail.[4] For all the discussion of "AI for Good" and "Socially Beneficial AI" in the AI

[4]In purely economic costs, the Cold war is estimated to have cost the United States an estimated $5.5 to $8 trillion in inflation-adjusted military expenditures. One can only wonder what else those funds could have been spent on. See:
https://www.nytimes.com/1999/05/20/opinion/l-cold-war-s-heavy-cost-770728.html
https://en.wikipedia.org/wiki/Effects_of_the_Cold_War

community, if many or most of the best AI engineers end up working on expensive military AI and robotics projects, they will not be working on those socially beneficial projects. It is difficult to measure the value of such missed opportunities, but it would be substantial.
(Amoroso and Tamburrini 2017)

Unpredictability

Another fear is that autonomous weapons could simply go "out of control" and do things that are completely unintended or highly unpredictable. While armed conflict is always unpredictable, such systems could add a whole new level of predictability. On the one hand, there is the possibility of such systems initiating or escalating a conflict without any human political or military decision to do so. While this can sometimes happen due to the unauthorized actions or mistakes of military personnel, humans are capable of recognizing the larger implications of their actions and can seek confirmation from superiors, while automated systems are not capable of this.[5]

The operator of an autonomous system may have a general idea of what the system is supposed to do, and may further have operational experience of how it operates in various specific contexts. But insofar as autonomous weapons are designed to operate over large geographic areas and time frames, and the possible interactions with the environment it may have grows exponentially, it will become increasingly difficult for even well trained operators to reliably predict what a system will actually do once deployed. Testing and reliability metrics can only offer confidence to operators when systems are deployed in situations and contexts that match those under which it has been previously tested, while increasing ranges and time frames imply that operators are less aware of the specific characteristics of the environment the system will encounter.

Further, there is much interest in developing large fleets and swarms of autonomous weapons systems. Such large ensembles of autonomous systems, even relatively simple ones, are known to be intrinsically unpredictable, from a mathematical perspective. But even a small number of autonomous systems interacting with each other, when we only know how some are programmed, are unpredictable because we do not know how an adversary's systems are programmed or what the net result of interactions between them will be. This issue is similar to that of various computer trading systems, whether for pricing products for on-line markets like Amazon, or for trading stocks. Both have manifested unexpected positive feedback loops resulting in million-dollar books being listed for sale on Amazon, and in major trading market crashes such as the one at the New York Stock Exchange in 2010, called a flash-crash, that lost 9% of the market's value in just a few a minutes.[6] However carefully programmed and tested autonomous weapons are, such catastrophic incidents will remain highly probable or inevitable

[5]Unless, of course, they are equipped with meaningful human control and are designed to seek human authorization for the use of force, and hence are not by definition autonomous weapons in that instance.

[6]See https://en.wikipedia.org/wiki/2010_Flash_Crash

with large numbers of autonomous systems deployed. The consequences, however could be far worse if those systems are controlling weapons, instead of trading stocks or books.

Lowering Thresholds of Conflict, Unintended Conflicts & Unattributable Attacks

Another set of concerns around autonomous weapons is that they will dramatically shift how political leaders think about armed conflict, how they make decisions about the use of military forces, and even how military leaders make strategic decisions. Because autonomous weapons promise to deliver military goals without putting human soldiers at risk, such weapons could lower the political thresholds for going to war. As we have already seen with armed drones, which create the possibility of military interventions without risks to either pilots or special forces commandos, leaders may be more likely to choose a military option offered by autonomous weapons when other options are too politically risky. If such situations are common, then the result will be more military operations, rather than seeking out political solutions.

There is also a set of concerns around the possibility of autonomous weapons acting or reacting in ways that initiate or escalate a conflict without any human political or military decisions. Imagine a border patrolled by autonomous combat aircraft from neighboring countries. One might be blown off course and into the airspace of the other, which could automatically initiate an attack, and the other could return fire, both could call in additional units, and very quickly a conflict could be initiated before humans were even alerted. Similarly, a low level military operation, such as a patrol, could rapidly escalate through a series of automated decisions into a major engagement, which could in turn escalate the nature of the overall conflict, or lead to the commitment of greater resources and more extreme forms of violence, or lead to the involvement of previously neutral parties. This concern is also related to the unpredictability concern of the previous section.

Autonomous weapons, even more so than remote operated weapons, offer the possibility of unattributable attacks. This is a phenomenon usually discussed in cyberwarfare, where it can be very difficult to determine the source of an attack with enough certainty to justify economic, political or military retaliation. But insofar as there is plausible deniability, or genuine uncertainty, as to who built and deployed a weapon, and its purpose, it will be difficult to definitively attribute an attack to its author. This could lead to widespread use of assassination by states against perceived foes, or seemingly random and chaotic attacks meant to destabilize and confuse civilians or political leaders. The possibility that systems, including those with known owners, could be hacked and hijacked and turned against third parties could also cause significant problems, to which we now turn.

Vulnerabilities to Hacking, Spoofing & Cyberattacks

It is possible to create autonomous weapons that do not use programmed computers, and we could even consider landmines as the “stupidest autonomous weapons” on the basis of their lack of discerning sensors or computational functions. However, it is much more likely that we will

see computational technologies involved in the decision processes of most autonomous weapons, as well as a variety of sophisticated sensors providing inputs to those decision functions. And given the nature of armed conflict, it is very likely that adversaries will attempt to interfere with autonomous weapons directly through cyberattacks which impair, disable, or take control of those systems, or indirectly by fooling or "spoofing" those systems through their sensors and what is known about how they process information.

Spoofing is a form of tricking an automated system to do what you want it to by manipulating its sensor data. One could do this by attacking its sensors, or simply manipulating what those sensors capture. A well-known example of this is spoofing GPS geolocation sensors. These sensors respond to signals from GPS satellites in space, and compute their location from the signals of multiple satellites. It is possible to bombard these sensors with signals that imitate the satellite signals but are much stronger. If, *e.g.* an autonomous drone aircraft is attempting to fly to a certain coordinate, but one systematically manipulates its GPS inputs, it is possible to force it to fly wherever you want.[7] It is not unreasonable that this, and many other means might be deployed to spoof autonomous weapons, including baiting them to attack the wrong targets, expend their ammunition, or even turn them against civilians or the military that fielded them.

Indeed, recent research in machine learning has demonstrated that because the dataspaces over which deep learning algorithms learn is so vast, it is possible to develop what are called generative adversarial neural networks which can systematically deceive a trained neural network so as to trigger any desired output. Moreover, it can do this with, *e.g.*, visual images that appear to the casual human observer as identical to an image that normally have a very different output. For example, two images of a "STOP" sign might appear identical to human observers, yet one could cause a self-driving car to stop as it should, while the other could be designed by an adversarial network that figured out a few select elements of the image that when altered will trigger the neural network to instead recognize this as a "SPEED LIMIT 55" sign. This is a major and fundamental problem within machine learning with no apparent solution. As long as it remains unsolved, any autonomous weapon that employs such machine learning techniques would be susceptible to manipulation, including making enemy combatants look like civilians, and friendly forces look like threatening adversaries.

Of course, because autonomous weapons will be primarily computer-controlled, and likely networked, they will be subject to most of the same vulnerabilities currently faced by computers and computer networks, namely, that hackers will be able to launch cyberattacks against them, and potentially gain control of them. While that is possible to some extent with advanced weapons systems that employ computer controls, insofar as those systems require human operators to engage the weapons, or pull the trigger, these functions cannot be commandeered by hackers. Autonomous weapons will extend the power of hackers and the kinds of effects they can have.

[7]This is most likely how the Iranian government forced down and captured a top secret United States RQ-170 surveillance drone flying over its territory in 2011 (see https://en.wikipedia.org/wiki/Iran%E2%80%93U.S._RQ-170_incident).

As discussed above, it can often be difficult to attribute cyberattacks to their source. Insofar as unattributable cyberattacks can gain control of weapons systems, then attacks from those weapons systems will also be unattributable. Thus, hackers could commandeer the weapons of one country and use them against another country–and it could turn out that neither country is able to determine the source of the attack. Alternatively, one country could simply claim that its systems had been commandeered and launch an attack. It will become increasingly uncertain, and more difficult to establish attribution, resulting in more plausible denials and highly unstable political situations, leading to greater international instability.

A New Kind of Weapon of Mass Destruction

Another concern raised around autonomous weapons is that they could constitute a new form of a Weapon of Mass Destruction (WMD). Historically, this term referred to nuclear weapons and later to chemical and biological weapons which could have massively devastating effects from a single use, similar to that of a nuclear weapon. But a more technical way of looking at, or defining a WMD, would be to say that it is a weapon that allows a single individual or small group to cause mass casualties. Conventional guns and bombs can cause mass casualties, but only to a degree far lower than what a nuclear bomb or poisoning of a water supply for a major city might. Because autonomous weapons do not need individual operators, it seems likely that a single individual or small group will be able to deploy vast fleets or mass swarms of such weapons. Unlike the indiscriminate nature of previous WMDs, such swarms of autonomous weapons could be designed to be highly discriminate, *e.g.* killing everyone over a certain height, or whose face matches a database. The point is that small groups could unleash mass devastation, and this is worrying because it could further empower terrorists, tyrants, and others who would wreak such devastation and sow chaos and fear in order to enhance their own position or power, thereby serving to destabilize the peace and security of the world.

Arguments for Moral Desirability of Autonomous Weapons

There are those who have argued that autonomous weapons are not only morally permissible, but care even morally desirable or morally required. While I disagree with this view, it is important to understand the basis of this argument in terms of moral reasoning. That basis is ultimately a consequentialist, *i.e.* utilitarian, one, and while compelling when considered as a singular decision regarding immediate consequences made in isolation, it fails to take seriously any of the broader consequences of permitting autonomous weapons.

The basic argument, as articulated by the roboticist Ron Arkin (2010, 2013), is that many of the civilian casualties in war are the result of human errors–soldiers making lethal decisions when they are exhausted, afraid, angry, or even vengeful. The "plight of the non-combatant," as Arkin (2013) puts it, is not simply being at the wrong place at the wrong time and getting caught in the crossfire, but rather to face death owing to the human failings of the soldiers who are entrusted to pull the triggers. If this is indeed the plight of non-combatants in warfare, then introducing automation that could "correct" those mistakes would greatly reduce civilian casualties. Given

that there is an obligation to protect civilians in combat, then it follows that it is desirable and perhaps even morally required to automate lethal decisions so as to eliminate such human errors.

Apart from the lack of evidence for its empirical claims,[8] this argument takes a very narrow view of morality with respect to killing. It imagines that we can directly substitute a human targeting decision with a computational process that will perform better than a human with respect to identifying civilians. Even if such a computational identification system existed, this by itself does not seem to require the elimination of the human decision maker. The computational system could be an advisory or recommendation system, allowing the human to make the decision and take action, but also provide warnings about potential errors in judgement, or the risks of alternative actions. We could even go further and design the system to not permit humans to target civilians, or put them at risk, at all. Like an automated "safety" that prevents a weapon from firing on civilians, Arkin *et al.* (2009) call this mechanism "the ethical governor" after the steam-engine governor of James Watt. Again, the ability of such computational processes to increase accuracy and precision in distinguishing civilians, or to better predict the risks of certain attacks to those civilians does not directly argue in favor of completely eliminating the human element. The only reasons in favor of that are reasons of efficiency, or expediency, and perhaps the military advantages of deciding and thus acting more quickly. Similarly, one could argue that humans are too slow, and suffer from fatigue or psychological pressures which machines will not. Those may be good reasons for accounting for human failings, and in certain situations where response time is critical, such as in the heat of battle. But in terms of overall military operations, live combat is a relative rarity, time is not always of the essence, and sometimes patience is rewarded with more favorable conditions for completing a mission or reducing risks to civilians. Indeed, much thinking around nuclear arms control is concerned with increasing the amount of time to make any potential launch decision.

While these arguments do not lead necessarily to the conclusion that humans will always make better decisions than machines, it does significantly weaken the intuition that replacing human decision with high-performing machine decisions will necessarily be good or even better than human decisions, or decisions reached by humans augmented with advisory information systems. Again, the basic structure of Arkin's argument, while appealing to the rights of civilians and duties of soldiers to protect civilians, takes a consequentialist form in which whatever reduces civilian casualties the most is the most desirable approach. But for this to be true, one

[8]Arkin offers only anecdotal evidence and citations to civilian casualties at U.S. checkpoints in Iraq to support this claim. However, we lack any substantial statistics as to how many casualties are due to such "errors" as opposed to other causes. It is also unclear how many civilian casualties are the result of what soldiers or officers consider acceptable proportionality considerations, or acceptable levels of collateral damage. It is also unclear how many civilian deaths in conflict zones result from non-weapons, such as loss of clean water, disease, starvation, and dislocation–the plight of those civilians would not necessarily be improved by automated targeting. And, of course, since the technologies in question are still hypothetical, it is impossible to evaluate their actual performance, or compare it to human performance. Moreover, they are likely to have their own types of errors and mistakes.

must consider a broader view of the consequences of permitting autonomous targeting and use of violent force without meaningful human control.

From the perspective of international relations and global security, the push towards greater autonomy in weapons has been regarded by some to be a revolution in military affairs with the potential to change the nature of warfare and the balance of power between those who have such weapons and those who do not. As such, there are clear risks to regional and global political stability, and precarious balances of military power, owing to potential arms races and proiferation in the domain of autonomous weapons as described earlier. Moreover, as these weapons go into mass production, their cost and availability will go down making them much easier to obtain by non-state actors and terrorist organizations.

The kinds of potential problems, risks and harms, described in more detail in the previous sections, are primarily consequentialist in nature. As such they depend on how the development deployment and use of autonomous weapons actually plays out in the real world. At this point, we can identify the most likely uses, and risks, posed by the technology. There are, of course, other possibilities for how the technology might unfold, paths it might take that we do not expect. Despite this, it seems unlikely that those paths would lead to more restrained development, more restrained uses of force, or lower risks to civilians and combatants. What seems far more likely is that the development and use of autonomous weapons will lead to more conflicts of increasing intensity, and an overall rise in political instability within countries and internationally. Thus even if we were to accept the utilitarian advantages of increased targeting precision of autonomous weapons, the overall consequences for civilians in war, and humanity in general, could be negative, and appears very likely to be so. But our ethical and moral consideration of autonomous weapons need not be limited to consequentialist analysis, or assessments of the value and likelihood of various possible outcomes. Rather we can look to the impact the development of autonomy in weapons systems will have on key moral principles of responsibility and accountability, human rights and human dignity, to which we now turn.

Threats to Responsibility & Accountability

Unlike previous technological advances, the advance of autonomy in machines presents some unique moral challenges. This is because machine autonomy intercedes upon human agency, both redistributing it and rearranging it, and in some ways confounding the norms we have long relied upon for ascribing moral responsibility and holding people and institutions accountable.

In particular, the delegation of targeting decisions to machines poses a specific threat to ascribing responsibility to the operators or commanders of an autonomous machine. Generally, when a human is “in control” of a weapon system, and directs that system at a target and engages it, we expect that the operator has an intention to destroy or disable the target. The target must be recognized as valid, or legal, and destroying must be recognized as fulfilling a military necessity, and the use of force in the situation must be morally justified, legitimate or permissible.

However, if the operator has delegated the targeting decision to an autonomous function of a machine, then the targets are determined by algorithms applied to data. While the operator may have a general intention in mind, such as "destroy enemy vehicles in this area," unless the operator inspects and confirms each of the targets selected by the automated process, they do not know if each is, in fact and in this particular circumstance, a lawful target. As such, neither the operator, nor the system designers, have access to the justification for designating a target as lawful. And if the system were to make mistakes, these might not be viewed as culpable crimes, but rather as mistakes or simply technical errors.

Moreover, the system itself, and its algorithms, are not legal or moral agents who can be held morally responsible or legally accountable for their choices and actions. While there is a certain intention behind the design of an algorithm, many assumptions must be made about the context and circumstances in which the algorithm will operate, and the kinds of sensor data it will receive. As such, algorithm designers are not making judgements based on the actual circumstances and situation in which those decisions will be made, but simply crafting clever rules that they expect will approximate such judgements given various assumptions. While there may be a certain degree of accountability for the designers of systems, particularly if they are negligent in their designs, it is quite natural to excuse mistakes owing to unforeseen circumstances.

Researchers in machine ethics have suggested that we can model or simulate legal and moral reasoning in a machine. But even if we try to represent International Humanitarian Law in a computational system,[9] and provide a means to reason out whether a particular target is lawful or not, this kind of simulation is not sufficient to justify killing. If the system were to kill a civilian, how would we hold it accountable or responsible for that death? We might be able to ask for its justification–the chain of reasoning that led it to make the incorrect decision. Or we might be able to diagnose the failure in its sensors, or logic, or the features of the environment which led to the error. And we might even be able to correct these failures. But the system itself would not be responsible, and could not be punished (Asaro 2011). And since we cannot really hold the operators or designers responsible either (Sparrow 2007), there would seem to be a responsibility gap that has been created by the introduction of the autonomous system.

In civilian cases, there is a body of liability and tort law in place to address the nature of responsibility in unintentional and accidental harms (Asaro 2016). However, in warfare, such laws do not apply to combatants. Some have argued that there are indeed war torts (Crootof 2016) yet this applies mainly to the liability of military suppliers and subcontractors to the military, not the liability of soldiers to their accidental victims.

Others have argued that states will always be responsible for the armed forces they command, and thus for any autonomous weapons they deploy. However, there is a very different psychological process involved in a soldier who is making a judgement to use lethal force for which they will be responsible, and deploying a system for which one does not expect to be held

[9]There are, of course, good reasons to believe that the rules of law are not easily encoded as computer programs, see (Asaro 2009).

accountable. As we have seen in other areas where automation has been deployed, the lack of potential responsibility creates greater levels of risk-taking and recklessness. Consider the vast array of financial instruments designed to limit risk and liability, and the high-risk markets they have spawned.[10]

Since machines and automated functions are not moral and legal agents, it is inappropriate to delegate moral and legal authorities to such systems (Asaro 2012). In the case of autonomous weapons, it is immoral, and should be illegal, to delegate the authority to kill, or to select and engage targets with violent force, to such systems. The legal consequence of such delegation is to create the responsibility gap, which undermines the moral and legal reposnsibilty of the individuals involved in armed conflict. And further, insofar as it becomes more difficult to apply international law to individuals, it also serves to undermine the international law framework itself.

From a deontological perspective, one can delegate the performance of certain obligations to other moral agents who take responsibility for fulfilling those obligations. But those agents must be moral agents capable of taking that responsibility. It is irresponsible, and thus immoral, to delegate obligations to entities that cannot take on those responsibilities.[11] From a consequentialist perspective, such delegation might look appealing if believe that the amoral agent might act to increase overall utility. However, in calculating the balance of utility, it should be noted that there is a moral hazard, itself a negative effect, in allowing individuals to wrongfully delegate their obligations because they will be less likely to take their obligations seriously, or feel responsible for their moral failings, or act to fulfill their obligations. Similarly, from a legal perspective, the inability to hold individuals accountable or responsible for their actions, or failures to fulfill their obligations, makes legal enforcement difficult or impossible. This, in turn, is likely to lead to the flouting of those laws, as well as a more general loss of respect for all laws and the legal framework itself. Thus the abrogation of duties through wrongful delegation is both deontologically wrong, and has serious negative consequences.

The other way to view this moral requirement is that humans need to maintain control over weapons systems to the extent that they can ensure the targeting of humans is lawful and the use of violent force against it is justified. Another way of stating this is that all weapons systems require *meaningful human control*. Before considering just what this might mean, we turn first to a discussion of the nature of human rights and human dignity which are threatened by the lack of meaningful human control of autonomous weapons.

---

[10]For example, collateralized debt obligations (CDOs) where a means of packaging high-risk subprime mortgages in ways that obscured their risk and diminished or eliminated the responsibility of both the issuers of the original debts and those who re-packaged them by appealing to third-party accreditations. The bubble created by these instruments eventually burst causing the global economic crisis of 2008, for which nearly no individuals or institutions were held liable. See https://en.wikipedia.org/wiki/Financial_crisis_of_2007%E2%80%932008

[11]Consider the case of an adult leaving a small child in charge of a fire, or to operate dangerous machinery. Here we would call the adult irresponsible, and immoral, not the small child who is not capable of taking on such responsibilities.

Threats to Human Rights and Human Dignity

The issue of autonomous weapons was first raised at the United Nations by Christof Heyns in his report on LAWS to the Human Rights Council as Special Rapporteur on Arbitrary Summary and Extrajudicial Execution (Heyns 2013). He argued subsequently that the fundamental ethical question is one of rights and dignity, and the ICRC has since reached the same conclusion (Heyns 2017, ICRC). Many states have also acknowledged the difference between consequentialist arguments for and against LAWS, and deontological arguments over the impact on rights and dignity. Of course, there is also a great deal of misunderstanding regarding deontological ethics and arguments–which some people find intuitively compelling and others do not.

The right to life is widely recognized as a fundamental human right, the loss of which is irrevocable, and upon which nearly all other rights depend. One cannot exercise one's right of free speech if one is dead, nor can one's life be restored if it is taken in error. As such, the right to life is highly valuable, and any decision to deprive someone of their life is of great significance and requires compelling justification. While accidents deprive people of their lives with some frequency, these are not intentional acts and thus we do not expect them to have justification. If we allow autonomous systems to target and engage violent and lethal force, however, we must ensure that the intentional killing that results is legally and morally justified. And, as discussed in the previous section, insofar as artificial systems are not capable of legal and moral agency, or of appropriate legal and moral deliberation, they cannot understand whether a particular killing is justified or be held responsible for such a judgement.

Human dignity is concept that often appears in discussions of human rights, but is rarely considered in detail in arms control. It is sometimes described as a sort of property that attaches to a person, and which can be stripped from them. But this metaphor only captures a part of what constitutes human dignity, and does little to help us understand its importance in armed conflict. Indeed, it is often said that there can be little dignity in war, or in dying in war, and certainly the manner in which many people are killed in wars by flames, explosions, shrapnel, bullets, etc. is lacking in dignity. But this is not what is really meant by human dignity or what it means to respect it. It is not the physical means of death that determines whether a death is dignified, any more than the manner of death justifies whether the death is lawful or moral. The human right to dignity, much like the human right to life, inheres not in a property and its loss, or the physical-causal means of its loss, but in the reasons for that right being violated or overridden. And it is fundamentally the right to be recognized as a human and accorded the respect and rights of all humans.

The human right to dignity, like the human right to life, is an intrinsic right that is realized in relations between humans–and duties of humans to respect the rights to dignity and life of other humans. And such rights are never lost, merely overridden by the rights of others, *e.g.* to self-defense. Accordingly, there is no "right to kill" or "license to kill" even in war. Rather, the right to defend one's self individually, or to defend one's nation collectively, is recognized as the right

to life of one party overriding the same rights of another. It is also recognized as limited, and requiring the proper relation between individuals. Thus, soldiers can be murdered by civilians in war, just as soldiers can murder civilians, or even fellow soldiers, in war, yet the killing of enemy combatants is not considered murder. But it is only when the proper relation between exists, *i.e.* the each are enemy combatants in a state of war, that killing is morally and legally acceptable. And further, establishing that relationship depends upon *reasons*–primarily the right to self-defense, but also satisfying the conditions that limit killing in war–discrimination to ensure that only those in the proper relation are killed, proportionality to ensure that killing is not disproportionate to its justification, and military necessity to ensure that the killing is really necessary for the ostensive purposes of collective self-defense. When killing lacks these properties it can be considered a war crime, or it can be an arbitrary, summary or extra-judicial execution in which there are not sufficient moral or legal reasons, reasoning, or legal process to justify the killing. For killing to be non-arbitrary, a morally responsible agent must have legitimate reasons for depriving someone of their life.

When it comes to the question of autonomous weapons, it may seem easy to argue that "it does not matter how one is killed in war." But clearly it does, from both a moral and legal perspective. What, then, is required to ensure that killing in war does not violate human dignity? As Heyns (2017) has argued, it must not be arbitrary, which is also to say that it must be justified. The question for autonomous weapons is whether a calculated machine decision–computations based in sensor data–can understand and act on legitimate reasons for killing.

I have argued (Asaro 2012) that computational systems are not moral and legal agents and thus cannot legitimately determine when it is appropriate to take human life. While computational systems may be able to accurately and reliably apply a computational rule to a set of data, in so doing they are not thereby respecting human dignity. In order to make a moral judgment to take a life, while respecting human dignity, it is minimally required that a moral agent can 1) recognize a human being as a human, not just distinct from other types of objects and things but as a being with rights that deserve respect, 2) understand the value of life and the significance of its loss, and 3) reflect upon the the reasons for taking life and reach a rational conclusion that killing is justified in a particular situation. Currently, only humans are capable of meeting these criteria, which is why it is morally and legally required that humans take responsibility for decisions to use lethal force, and should continue to be in the future.

Distinguishing a "target" in a field of data is not recognizing a human person as someone with rights. Nor is discriminating between combatants and non-combatants sufficient for recognizing someone as a human being with rights to dignity and life. When it comes to making proportionality decisions, the value of human life is not quantifiable in any deep sense. As human beings, who have experienced loss and are ourselves mortal, we have access to the qualitative value of human life. And finally, machines are not capable of deciding questions of military necessity–whether a state of war exists, whether a person in a given situational context can be justifiably killed, or what reasons justify the necessity of destroying a military objective.

There are two types of objections to this framing of the autonomous weapons and human dignity. One objection is that machines might be programmed to perform better than humans in some

sense. On the one hand they might be better at discrimination, proportionality or even military necessity calculations. However, this would depend on a consequentialist analysis, whereby "better performance" consists of making automated choices more accurately or reliably than a human. But this overlooks the reasons and justifications for the decisions, and respect for the underlying duties and rights, as opposed to the consequences, of choices. When it comes to human dignity, what is crucial is both the manner in which the decision is made and the legitimacy of who is making the decision, not simply the final outcome of the decision. The second objection is some form of skepticism of human dignity–either that it does not really exist, or that it is reducible to the formal respect of other rights (e.g. life), or that it is a spiritual, mystic, or ephemeral quality that can never rally be adequately respected and thus can be ignored. While it is notoriously difficult to argue against skepticism, the fact that human dignity has been articulated in various world philosophies and religions, has been integral to legal theory and practice, and has been codified in the constitutions of countries, including Article 1 of the German constitution, and the Preamble of the United Nations Universal Declaration of Human Rights, provides strong support that human dignity has both a defined structure and is broadly recognized as integral to law and morality.

Meaningful Human Control

Given the moral obligation to ensure that weapons are only used against justified and lawful targets, what is the best way to realize this obligation? The discussions by States at the United Nations Convention on Conventional Weapons has repeatedly asserted, with consensus, a few key points in this regard. First, that International Humanitarian Law applies in all cases of armed conflict, and to all weapon systems. Accordingly, the Geneva Conventions which require commanders to take all precautions to protect civilians in every attack, as well as to review all "new weapons, means and methods of warfare" for their compliance with the law under Article 36 of Additional Protocol 1 of the Geneva Conventions also applies to any autonomous or highly automated weapons. But how to actually ensure these obligations are met actually requires that there is space for human reasoning, and moral consideration within decisions to use violent force.

It has been proposed at the CCW discussions that what is needed for this is a requirement for "meaningful human control" over the targeting and engagement of all weapons. This could be viewed as the positive obligation which mirrors the negative obligation of not delegating the authority or responsibility for lethal decisions to machines or automated processes. But this term serves other functions and deserves a bit of unpacking. While "control" does most of the work in terms of moral responsibility, the "human" element is clearly the one that stands out as a requirement for the non-delegation of certain authorities. There is a sense in which software, and automated systems in general, are authored and created by humans, and could be seen as a form of human control. While this is acceptable in certain situations, such as automatic doors and thermostats, the decision to use violent force and take human life requires a human capable of assessing the situation, determining the necessity to engage a weapon on a target, who has access to the moral and legal justification for the use of violent force, and who can take moral and legal

responsibility for the consequences of that decision. Together these elements add up the something we can describe as "meaningful."

The "control" element implies that systems are acting at the direction and under the control of some specific and identifiable person. This is important both in terms of accountability for the consequences of the use of such a system, but also that there is a human who can actively intervene on a system should it act erratically, unpredictably, or create undesirable effects. This is control both in the engineering sense and in the legal and moral sense of there being a human who is controlling a system to achieve their intended results. If the system can act without human intention, or against those intentions, then it is not really in control, or is completely out of control.

The "human" element of the term is meant to carry the full burden and responsibilities of moral and legal agency. By virtue of being a morally responsible human, one has an understanding of the value of human life and human dignity that cannot really be represented in a calculation. It is the human condition, and our particular relationship with mortality and life, that informs and grounds our morality. It is the human, or group of humans, who control a system who are responsible for it and the consequences of the actions taken by the system. Responsibility here operates in two directions. On the *post hoc* side of things, responsibility means that we can hold somebody responsible for what an autonomous system does, after the fact. For this to be fair, the responsible person must actually be in control of the system. If a system goes haywire, acts in unpredicted and unintended ways, this would tend to diminish the responsibility of the human. Indeed, it would be unfair to hold someone responsible for the complex actions of a system that the individual could not possibly have foreseen. But we can hold them responsible if they should have known, or if simply activating a system was itself reckless or negligent.

Responsibility also applies *before* a system is deployed or engaged, in terms of acting psychologically on the human who is in control. That is, one crucial way that moral norms function to regulate human behavior is to encourage people to reflect upon their actions before taking them, and to avoid immoral or reckless acts. In psychological terms, the person should consider the moral implications and consequences of their action, and take moral responsibility for choosing to act in certain ways. A key moral issue with automating all sorts of decisions is that it removes human conscience, and moral deliberation, from the decision process. If an act is morally questionable, or morally wrong, it should be hard to commit.

This comes to the "meaningful" portion of the concept. Human acts are social and cultural acts that have meaning within social and cultural systems. War itself is a cultural phenomenon, imbued with complex layers of meaning. For it to be meaningful at all, humans must be capable of engaging with it in meaningful ways. While we do not wish to encourage warfare, or to glorify it, we also should not permit violent acts of armed conflict which are untethered to human understanding and morality. Killing should only be done with good reasons, and humans should always be deliberating whether their reasons are good. And when they are, there should be a human willing to take responsibility for acting on those reasons.

It could be argued that we might be able to build artificial agents who are in fact capable of moral agency If it were possible, would it be acceptable for such entities to make decisions and take actions to kill humans? I am skeptical that we understand the nature of moral agency sufficiently to automate it. While we can create models of moral decision and teach machines to follow them, or even simulate aspects of human psychology, these are not really the same thing as being a moral agent and taking responsibility for ones own actions. While we cannot prove in principles that it is impossible to replicate the capacities in artificial systems, it would be extremely challenging, and would itself be immoral (Bryson 2018). Furthermore, even if we succeed in creating artificial moral agents, we might still not consider them human persons, but perhaps a more alien form of being, deserving of respect perhaps but not necessarily entitled to judge when it is acceptable to take human life. At any rate, for the foreseeable future we should work to ensure that autonomous systems are not given the authority to use lethal or violent force against humans.

Citations

Arkin, Ronald C. "The Case for Ethical Autonomy in Unmanned Systems," *Journal of Military Ethics*, 9 no. 4 (2010).

Arkin, Ronald C., "Lethal Autonomous Systems and the Plight of the Non-Combatant," *AISB Quarterly,* 137 (2013).

Arkin, Ronald C., Patrick Ulam, and Brittany Duncan, "An Ethical Governor for Constraining Lethal Action in an Autonomous System," Georgia Tech Technical Report GIT-GVU-09-02 (2009).

Amoroso, Daniele and Guglielmo Tamburrini, "The Ethical and Legal Case Against Autonomy in Weapons Systems," *Global Jurist* 17(3), January 2017.

Asaro, Peter, "Modeling the Moral User: Designing Ethical Interfaces for Tele-Operation," *IEEE Technology & Society*, 28 (1), 20-24. (2009).

Asaro, Peter, "A Body to Kick, But Still No Soul to Damn: Legal Perspectives on Robotics," in Patrick Lin, Keith Abney, and George Bekey (eds.) *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press, pp. 169-186. (2011).

Asaro, Peter, "On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making," *International Review of the Red Cross,* 94, no. 886 (Summer), pp. 687-709. (2012).

Asaro, Peter, "Roberto Cordeschi on Cybernetics and Autonomous Weapons: Reflections and Responses," *Paradigmi: Rivista di critica filosofica*, Anno XXXIII, no. 3, Settembre-Dicembre, 2015, pp. 83-107, (2015).

Asaro, Peter, "The Liability Problem for Autonomous Artificial Agents,"AAAI Symposium on Ethical and Moral Considerations in Non-Human Agents, Stanford University, Stanford, CA, March 21-23, (2016).

Asaro, Peter, "What is an 'AI Arms Race' Anyway?," *I/S: A Journal of Law for the Information Society*, Vol. 15 No. 1-2 (Spring 2019), pp. 45-64.

Bryson, Johanna J., "Patiency is not a virtue: the design of intelligent systems and systems of ethics," *Ethics and Information Technology* (2018) 20:15–2

Crootof, Rebecca, "War Torts: Accountability for Autonomous Weapons," 164 *University of Pennsylvania Law Review*, 1347 (2016).

Heyns, Christof, "Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns," A/HRC/23/47 (2013), https://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf

Heyns, Christof, "Autonomous weapons in armed conflict and the right to a dignified life: an African perspective," *South African Journal on Human Rights*, Volume 33, 2017 - Issue 1, pp. 46-71.

International Committee of the Red Cross, "Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons," ICRC Report, September 1, 2016. https://www.icrc.org/en/publication/4283-autonomous-weapons-systems

International Committee of the Red Cross, "Towards limits on autonomy in weapon systems," ICRC Statement, April 9, 2018. https://www.icrc.org/en/document/towards-limits-autonomous-weapons

Matthias, Andreas, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata," *Ethics and Information Technology*, 5 no. 3 (2004).

Roff, Heather M., "Killing in War: Responsibility, Liability, and Lethal Autonomous Robots," in Fritz Allhoff, Nicholas G. Evans, and Adam Henschke, eds., *Routledge Handbook of Ethics and War: Just War Theory in the Twenty-First Century* (Milton Park, Oxon: Routledge, 2013);

Sparrow, Robert, "Killer Robots," *Journal of Applied Philosophy*, 24 no.1 (2007).

Sparrow, Robert, "Robots and Respect: Assessing the Case Against Autonomous Weapon Systems," *Ethics & International Affairs*, Volume 30, Issue 1, Spring 2016 , pp. 93-116.