Robots and Respect: Assessing the Case Against Autonomous Weapon Systems

*Robert Sparrow**

The prospect of "killer robots" may sound like science fiction. However, the attention given to the operations of remotely piloted drones in recent years has also spotlighted the amount of research that is being conducted on weaponized robots that can select and attack targets without the direct oversight of a human operator. Where the armed services of the major industrialized countries were once quick to distance themselves from the use of autonomous weapons, there is increasing speculation within military and policy circles—and within the U.S. military in particular—that the future of armed conflict is likely to include extensive deployment of autonomous weapon systems (AWS).[1] A critical 2012 report on AWS by Human Rights Watch[2] and the 2013 launch of an NGO-led campaign for a treaty prohibiting their development and use[3] has intensified the ongoing ethical debate about them.[4]

My aim in this article is twofold. First, I will argue that the ethical case for allowing autonomous targeting, at least in specific restricted domains, is stronger than critics have typically acknowledged.[5] Second, I will attempt to defend the intuition that, even if this is so, there is something ethically problematic about such targeting. Given the extent of my ambitions, the dialectic that follows is somewhat complicated and for this reason it will be useful to briefly sketch an outline of the argument here.

My argument proceeds in three parts. In the first section I introduce a working definition of "autonomous" weapons and describe the military dynamics driving

*Thanks are due to Ron Arkin, Daniel Brunstetter, Ryan Jenkins, Mark Gubrud, Duncan Purves, Heather Roff, B. J. Strawser, Michael Schmitt, Ryan Tonkens, several anonymous referees, and the editors of *Ethics & International Affairs* for comments on drafts of this manuscript. Mark Howard ably assisted me with sources and with preparing the article for publication.

Ethics & International Affairs, 30, no. 1 (2016), pp. 93–116.

doi:10.1017/S0892679415000647

the development of these systems. In the second section I survey and evaluate the existing literature on the ethics of AWS. The bulk of this discussion is framed as an account of two "rounds" of debate between an influential advocate for AWS, Ron Arkin, and his critics. In the third and final section I turn to a deeper investigation of the philosophical foundations of the just war doctrine of *jus in bello* in order to develop a new account of the origins and force of the intuition that the use of killer robots would necessarily be morally problematic. I conclude that although the theoretical foundations of the idea that AWS are weapons that are evil in themselves are weaker than critics have sometimes maintained, they are nonetheless strong enough to support the demand for a prohibition of the development and deployment of such weapons.

The Military Case for Autonomy

Defining Autonomous Weapon Systems

Any sensible discussion of autonomous weapons must begin by clarifying what the author understands by "autonomy." The difficulties involved in providing a definition of autonomy broad enough to capture what people take to be (alternatively) exciting and/or problematic about these systems, without begging central questions in the debate about the ethics of their use, go a long way toward explaining why the literature on the ethics of AWS is so vexed.[6] A minimal definition of autonomy vis-à-vis a weapon or weapon system is that it must be capable of some significant operation without direct human oversight. Perhaps the strongest definition would posit that a system must be *morally* autonomous—that is to say, be a moral agent with free will and be responsible for its own actions—for it to truly possess autonomy.[7]

Thus, as a number of authors have suggested, it is helpful to think about lethal autonomous operations as situated on a spectrum, with, for instance, antipersonnel mines—which "decide" when to explode on the basis of input from a pressure sensor—at one end, and human beings or (theoretical) strong artificial intelligence at the other.[8] Of course, if one models the operations of autonomous weapons on the assumption that they are merely sophisticated landmines, then it may be hard to see what all the fuss is about. Alternatively, if a weapon must have the capacities of a human being to be autonomous, then it may appear that we have nothing to worry about, as we are a long way from knowing how to design such systems.

Where the debate about lethal autonomous operations gets interesting is somewhere in the middle, wherein the weapon system possesses a complexity that problematizes understanding it as merely a complex sort of landmine, but where it is not so sophisticated as to require strong artificial intelligence (AI). Questions about the appropriate way to allocate responsibility for the consequences of the operations of such systems arise at the upper end of this range of autonomy.[9] However, even where these questions do not arise, many people have the intuition that there is something morally problematic about robots killing people.

For the purpose of this article, and in order to avoid prejudicing my discussion of the larger literature by insisting on a more precise—and, therefore, inevitably more controversial—definition, I understand an "autonomous" weapon as one that is capable of being tasked with identifying possible targets and choosing which to attack without human oversight, and that is sufficiently complex such that, even when it is functioning perfectly, there remains some uncertainty about which objects and/or persons it will attack and why. This admittedly rough-and-ready definition represents my attempt to single out an interesting category of systems while avoiding entering into an extended and difficult argument about the precise nature of machine autonomy.[10] While the first part of the foregoing accords with the U.S. Department of Defense's influential definition of autonomy in weapon systems,[11] the second part is intended to help distinguish between automatic systems, such as the Phalanx Close-in Weapon System,[12] and potentially more complex systems that may require us to hypothesize on a case-by-case basis about the "reasons" for "its" actions.[13] Perhaps the paradigmatic case of the latter would be an autonomous weapon wherein genetic algorithms or machine learning played a central role in determining its behavior.[14] However, I also intend this definition to capture robots of sufficient complexity that do not rely on these mechanisms.

An Arms Race to Autonomous Operations?

Many speculate that the perceived success of remotely piloted drones and other unmanned systems in recent military conflicts means that the development of AWS is more or less inevitable.[15] There are at least three military and/or technological logics that drive powerfully in the direction of the development of weapon systems that can operate—and kill people—autonomously.

First, the communications infrastructure that enables the operation of existing remotely piloted weapons, such as the United States' Predator and Reaper drones, places significant limitations on the operations of these systems. The operation of long-range Unmanned Aerial Vehicles (UAVs) requires the transmission of large amounts of data by military satellite and radio systems. This places an upper limit on the number of UAVs that can be fielded at any point in any given theater of operations. It also restricts the capacity to field long-range UAVs to those few states that have the ability to launch and operate communication satellites (or that are able to access bandwidth provided by their allies). The need to sustain regular communication with human operators also effectively rules out a major role for remotely operated submersibles in future naval combat, given that submarines must operate in a communications blackout in order to avoid detection and destruction by enemy forces. The communication systems necessary to allow remote piloting of unmanned systems are also vulnerable to electronic countermeasures and/or kinetic attacks on the physical infrastructure that sustains them. In any future large-scale conflict involving major industrialized powers, military communication satellites would be among the first targets of attack. Developing and deploying AWS would therefore allow more weapons to be fielded and for the systems to be more survivable.[16]

Second, a number of technological factors have combined to greatly increase the tempo of battle over recent decades, especially in air-to-air combat. In conflicts involving modern high-technology military systems, victory may depend upon decisions that must be made in a matter of seconds and that require integrating information from multiple sources. The demands of modern combat already push the limits of what the human nervous system is capable of. In the future only AWS may be capable of reacting within the time frame necessary to facilitate survival in a hostile environment.[17]

Finally, a number of other features of the operations of UAVs and other unmanned weapons systems suggest that it would be preferable to remove human beings from their operations. For example, the routine operation of UAVs such as Predator and Reaper drones is, on all accounts, extremely boring for the vast majority of the time they are in theater. Consequently, pilot fatigue and error remain a significant cause of accidents involving these systems. Autonomous systems might be less prone to mishaps at launch and recovery, and while traveling to the battlespace, than those controlled by human operators. Moreover, training costs, salaries, and medical benefits for the operators of remote systems are significant. Just as unmanned systems have been touted as cheaper than the manned systems they replace, autonomous systems may eventually become less expensive to operate than remotely piloted ones.[18]

The Ethical Case for Autonomy

The development of autonomous weapons might be desirable on grounds unrelated to these military and technological logics. Alternatively, even if the development of these systems *is* more or less inevitable, it may still be the case that we should resist them on ethical grounds. Indeed, given that military competition between states is driving the rise of AWS, a global arms control treaty prohibiting autonomous weapons may represent the only way to prevent their being developed and fielded.[19] We must therefore consider the ethical case for (and, later, against) the development and deployment of autonomous weapons.

Arkin and His Critics: Round I

Ronald Arkin is perhaps the most vocal and enthusiastic advocate for developing AWS writing today. He is also actively involved in their development.[20] In his influential essay "The Case for Ethical Autonomy in Unmanned Systems," Arkin adduces a number of arguments in favor of autonomous operations.[21] He argues that, in the future, AWS may be better able to meet the requirements of ethical conduct in war than humans because: robots can be designed to accept higher risks in the pursuit of confidence in targeting decisions; will have better sensors; will not be swayed by emotions, such as fear or anger, which often prompt humans to act unethically; need not suffer from cognitive biases that afflict human decision-making; and will be better able to integrate information quickly from a wide variety of sources.[22] As I will discuss further below, his identification of the relevant standard of distinction against which the ethical use of AWS should be measured as that achieved by human warfighters is also a crucial intellectual move in the debate about the ethics of autonomous weapons.

Difficulties with Discrimination

Critics of Arkin's proposal have been quick to point out just how far existing robots are from being able to outperform human beings when it comes to adherence to the requirements of *jus in bello*.[23] In particular, Arkin systematically underestimates the extent of the challenges involved in designing robots that can reliably distinguish legitimate from illegitimate targets in war.

Despite many decades of research—and much progress in recent years—perception remains one of the "hard problems" of engineering. It is notoriously difficult for a computer to reliably identify objects of interest within a given environment and to distinguish different classes of objects. This is even more the case in crowded and complex unstructured environments and when the environment and the sensor are in motion relative to each other. In order for AWS to be able to identify, track, and target armed men, for instance, they would need to be able to distinguish between a person carrying an assault rifle and a person carrying a metal tube or a folded umbrella. Moreover, in order to be able to assess the likelihood of collateral damage and thus the extent to which a particular attack would satisfy the *jus in bello* requirement of proportionality, autonomous weapons will need to be able to identify and enumerate civilian targets reliably, as well as potential military targets. Thus, it will not be sufficient for AWS simply to identify and track armed persons (by recognizing the LIDAR[24] signature of an AK-47, for instance)—they must also be able to identify and track unarmed persons, including children, in order to refrain from attacks on military targets that would involve an unacceptably high number of civilian casualties. Weapons intended to destroy armored vehicles must be capable of distinguishing them from among the almost countless different cars and trucks manufactured around the world; autonomous submarines must be able to distinguish warships from merchant vessels, and so on. Moreover, AWS must be capable of achieving these tasks while their sensors are in motion, from a wide range of viewing angles in visually cluttered environments and in a variety of lighting conditions.

These problems may be more tractable in some domains than others, but they are all formidable challenges to the development of AWS. In fact, the problem of discriminating between legitimate and illegitimate targets is even more difficult than the foregoing demonstrates. For instance, not every person carrying a weapon is directly engaged in armed conflict (in many parts of the world carrying a weapon is a matter of male honor); with prior approval, foreign warships can pass through the territorial waters of another state; neutral troops or peacekeeping forces are sometimes present in areas in which legitimate targets are located; and children sometimes climb on decommissioned tanks placed in playgrounds. Thus, in order to discriminate between combatants and noncombatants, it is not sufficient to be able to detect whether someone (or something) is carrying a weapon. Discrimination is a matter of context, and often of political context. It will be extremely difficult to program robots to be able to make this kind of judgment.[25]

Even if a weapon system could reliably distinguish combatants from noncombatants, this is not the same as being able to distinguish between legitimate and illegitimate *targets*. According to *jus in bello* conventions, attacks on combatants may be illegitimate in at least three sorts of circumstances: first, where such attacks may be expected to cause a disproportionate number of civilian casualties (Additional Protocol I to the Geneva Conventions, Article 57);[26] second, where they would constitute an unnecessarily destructive and excessive use of force;[27] and third, where the target has indicated a desire to surrender or is otherwise *hors de combat* (Additional Protocol I to the Geneva Conventions, Article 41).[28] Before it would be ethical to deploy AWS, then, the systems will need to be capable of making *these* sorts of discriminations, all of which involve reasoning at a high level of abstraction.

Thus, for instance, how many noncombatant deaths it would be permissible to knowingly cause in the course of an attack on a legitimate military target depends on the military advantage that the destruction of the target is intended to serve; the availability of alternative means of attacking the target; the consequences of not attacking the target at that time (which in turn is partially a function of the likelihood that an opportunity to attack the target will arise again); the availability of alternative means of achieving the desired military objective; and the weaponry available to conduct the attack. Similarly, whether an attack would constitute an unnecessarily destructive use of force (which it may, even where there is *no* risk of killing noncombatants) is a function of the nature of the military object being targeted; the extent of the military advantage the attack is intended to secure; and the availability of alternative, less destructive, means of achieving this advantage.

Assessing these matters requires extensive knowledge and understanding of the world, including the capacity to interpret and predict the actions of human beings. In particular, assessing the extent to which an attack will achieve a definite military advantage requires an understanding of the balance and disposition of forces in the battlespace, the capacity to anticipate the probable responses of the enemy to various threats and circumstances, and an awareness of wider strategic and political considerations.[29] It is difficult to imagine how any computer could make these sorts of judgments short of the development of a human-level general intelligence—that is, “strong” AI.[30]

Identifying when enemy forces have surrendered or are otherwise *hors de combat* is also a profound challenge for any autonomous system.[31] Perhaps it will be possible to program AWS to recognize the white flag of surrender or to promulgate a convention that all combatants will carry a "surrender beacon" that indicates when they are no longer participating in hostilities.[32] Yet these measures would not resolve the problem of identifying those who are *hors de combat*. A gravely wounded soldier separated from his comrades is not a legitimate target even if he has not indicated the desire to surrender (indeed, he may have had no opportunity to do so), but it may be extremely hard for a robot to distinguish such a person from one lying in ambush. Similarly, a ship that has had its guns destroyed or that has been holed below the water so that all hands are required just to remain afloat—and is therefore no military threat—will not always have a different radar or infrared profile from a functioning warship. Human beings can often—if not always—recognize such situations with reference to context and expectations about how people will behave in various circumstances. Again, short of possessing a human-level general intelligence, it is difficult to imagine how a computer could make these discriminations.

Possible Solutions? "Ethical" Robots and Human Oversight

Arkin has offered two responses to these sorts of criticisms. I believe both are inadequate.

First, Arkin has suggested that it should be possible to build into the weapon system the capacity to comply with the relevant ethical imperatives through what he calls an "ethical governor."[33] This will not, of course, address the problems of identifying and classifying objects in complex environments, although it is possible that improvements in computer vision technology will reduce these problems to a manageable level. More fundamentally, it presumes an impoverished account of ethics as a system of clearly defined rules with a clearly defined hierarchy for resolving clashes between them.

The sketches of deontological or utilitarian systems of ethics that philosophers have developed are just that—sketches. The task of ethical theory is to try to explain and systematize the ethical intuitions that properly situated and adequately informed persons evince when confronted with various ethical dilemmas. These intuitions are extremely complex and context dependent, which is why philosophers are still arguing about whether they are primarily deontological or consequentialist or perhaps virtue-theoretical. It is these—still poorly understood and often highly contested—intuitions that a machine would need to be capable of replicating in order for it to "do" ethics. Moreover, even the schematized accounts of some subsets of these intuitions that philosophers have developed require agents to reason at a high level of abstraction and to be able to make complex contextual judgments for their application. For instance, consequentialists must be capable of predicting the effects of our actions in the real world, making a judgment about when this attempt to track consequences—which are, after all, essentially infinite—may reasonably be curtailed, and assessing the relative value of different states of the world. It is unclear whether even human beings can do this reliably (which itself is a reason to be cautious about embracing consequentialism), but it seems highly unlikely that, short of achieving human-level general intelligence, machines will ever be able to do so. Similarly, Kantian ethics requires agents to identify the moral principles relevant to their circumstances and resolve any clashes between them—again a task that requires a high degree of critical intelligence.[34]

However, the most fundamental barrier to building an "ethical robot" is that ethics is a realm of meanings. That is to say, understanding the nature of our actions—what they mean—is fundamental to ethical reasoning and behavior.[35] For instance, most of the time intentionally killing a human being is murder—but not during a declared armed conflict, when both the killer and the victim are combatants; or in situations of self-defense; or when it has been mandated by the state after a fair criminal trial. Thus, in order to be able to judge whether a particular killing is murder or not, one must be able to track reliably the application of concepts like intention, rights, legitimacy, and justice—a task that seems likely to remain well beyond the capacity of any computer for the foreseeable future. Perhaps more importantly, the *meaning* of murder—why it is a great evil—is not captured by any set of rules that distinguishes murder from other forms of killing, but only by its place within a wider network of moral and emotional responses. The idea that a properly programmed machine could behave ethically, short of becoming a full moral agent, only makes sense in the context of a deep-seated behaviorism of the sort that has haunted computer science and cognitive science for decades.

Arkin's second suggestion is that weaponized robots could be designed to allow a human operator to monitor the ethical reasoning of the robot. The operator could then intervene whenever she anticipates that the robot is about to do something unethical.[36] Other authors have suggested that AWS could be designed to contact and await instruction from a human operator whenever they encounter a situation their own programming is unable to resolve.[37]

This is problematic for two reasons. First, the need to "phone home" for ethical reassurance would mitigate two of the main military advantages of autonomous weapons: their capacity to make decisions more rapidly than human beings,[38] and their ability to operate in environments where it is difficult to establish and maintain reliable communications with a human pilot.[39] If an "autonomous" weapon has to rely on human supervision to attack targets in complex environments, it would be, at most, *semi*-autonomous.[40] Second, it presumes that the problem of accurately identifying the ethical questions at stake and/or determining when the ethics of an attack is uncertain is more tractable than resolving uncertainty about the ethics of a given action. However, the capacity of AWS to assess their own ability to answer an ethical question would *itself* require the capacity for ethical deliberation at the same level of complexity needed to answer the original ethical question. Thus, if we cannot trust a machine to make ethical judgments reliably, we cannot trust it to identify when its judgments themselves might be unreliable.

Arkin and His Critics: Round II

For these reasons, I believe that Arkin's critics are correct in arguing that the difficulties of reliably distinguishing between legitimate and illegitimate targets in complex environments probably rules out the ethical use of AWS in many roles for the foreseeable future.[41] However, Arkin does have available to him two replies to these sorts of criticisms that are more compelling. First, the problem of discriminating between legitimate and illegitimate targets is much more tractable in specific, restricted domains than Arkin's critics—and the arguments above—suggest. Second, Arkin has argued that the relevant standard of reliability in discriminating between legitimate and illegitimate targets, which robots would need to attain in order for their use to be ethical, is that achieved by human warfighters, which is much lower than might first appear. If AWS would kill fewer noncombatants than human troops, this establishes a strong consequentialist case for their deployment, regardless of other ethical concerns about them. As I will discuss below, a number of other advocates for autonomous weapons have also made a nonconsequentialist argument for the ethical use of autonomous weapons from their putative reliability compared to human warfighters.

One possible solution to the problems discussed above would be to constrain the spatial domain of the operations of AWS and/or the sorts of systems they are tasked with destroying.[42] How difficult it is to distinguish between a military and nonmilitary object depends on their specific features as well as the sensors available to the AWS; how difficult it is to avoid collateral damage depends upon the relative number of legitimate and illegitimate targets within an area of operations. In anti-submarine warfare, for instance, there are few civilian targets.[43] Similarly, in air-to-air combat, counter-artillery operations, or the suppression of enemy air defenses it is relatively straightforward to distinguish military from nonmilitary systems.[44] Tanks and mechanized artillery, and—to a lesser extent—naval assets, also have, for the most part, distinctive visual silhouettes and radar and infrared signatures that distinguish them from the nonmilitary objects (cars, trucks, merchant ships, and so on) among which they might be found. When potential targets are mechanized and combat is confined to a distinct theater of operations, it is much more plausible to hold that autonomous weapons will be capable of reliably identifying potential military targets and distinguishing combatants from noncombatants. Indeed, existing target identification systems are already capable of reliably distinguishing between military and civilian systems in these domains.[45] The claim that autonomous weapons will never be capable of reliably distinguishing between military and nonmilitary targets therefore appears incorrect.

Nevertheless, not every military target is a legitimate one. Even in these restricted domains, then, the challenge of discriminating between legitimate and illegitimate targets is harder than first appears. In order to be able to avoid causing disproportionate civilian casualties, AWS must be capable not only of identifying potential military targets but also of detecting the presence of civilians in the target area; in addition, proportionality requires them to be able to identify the civilian objects (churches, hospitals, and such) that are relevant to this calculation. They must also be able to determine when attacks on military targets are justified by the principle of necessity and will secure a definite military advantage.[46] Yet when combat is occurring in a discrete geographical area, especially in the air, in outer space, or underwater—or, more controversially, when civilians have been given sufficient warning to vacate an area in which hostilities are about to commence—*and* when victory in this context would advance an important military objective, it might prove possible to guarantee that the destruction of any of the military objects present would be justified. It is less clear, however, that the problem of identifying forces that have surrendered or are otherwise *hors de combat* is any more tractable simply because AWS will be restricted to a specific geographical area. The idea of "surrender beacons" is, perhaps, more practicable when the forces engaged are military assets rather than personnel. Yet the problem of identifying when enemy forces are illegitimate targets by virtue of being so incapacitated by wounds or damage that it is no longer reasonable to consider them as constituting a military threat remains profound. Nevertheless, it seems likely that by confining the operations of AWS to a carefully delineated "kill box," it might be possible to greatly reduce the risk of attacks on illegitimate targets.

The Consequentialist Case for Autonomy

At this point, Arkin has one further counterargument. That is, while he concedes that the task of designing AWS capable of distinguishing between legitimate and illegitimate targets is difficult, he claims that it is conceivable that the judgment of AWS may someday be superior to that of humans.[47] Indeed, by highlighting the real world attitudes and behaviors of U.S. soldiers deployed in Iraq, Arkin has argued persuasively that human warfighters are actually quite bad at behaving ethically during wartime.[48]

However, this is arguably the wrong standard to set when considering whether the use of a weapon system would be ethical.[49] Given that what is at stake is the value of an innocent human life, when it comes to protecting noncombatants from deliberate (or negligent) attack it might be argued that the relevant ethical standard is perfection. For instance, one could not justify deliberately or negligently killing two civilians for every ten combatants by pointing out that other warfighters typically kill three civilians for every ten combatants in similar circumstances. It is reasonable to expect human warfighters not to deliberately target noncombatants or use disproportionate force because they have both the power and the freedom not to. Thus, it is reasonable to expect perfect ethical compliance from humans, even if they seldom achieve it.[50] Putting AWS into combat when it would be *un*reasonable to expect that they *will not* violate the requirements of distinction and proportionality could only be defensible if one believes that the only relevant consideration is the eventual number of civilian casualties. Arkin's argument about the benefits of autonomous targeting therefore depends on adopting a consequentialist ethical framework that is concerned only with the reduction of civilian casualties—a controversial position, especially in the ethics of warfare.

There is, however, another version of this argument, which buttresses the claim about the relative effectiveness of weaponized robots with an appeal to the agency of those their operations might threaten. Thus, Brian Williams, for instance, has argued that the civilian population in the autonomous tribal areas of Pakistan actually prefer operations against al-Qaeda militants to be conducted via drone attacks because the alternative—antiterrorist operations by the Pakistani armed forces—is so much more destructive.[51] By appealing to the consent of the civilian population, this argument in support of AWS appears to mobilize powerful deontological intuitions.[52] Yet, on closer inspection, it is a red herring. Consider the nature of the circumstances in which civilians in Pakistan or Sudan might say that they prefer the operations of AWS to the deployment of human warfighters in the areas where they live. These civilians likely bear no responsibility for the events that led to the current conflict there, yet they face the choice of being threatened with death as a result of the operations of poorly trained and often terrified human beings or by autonomous weapons. This is a less than ideal circumstance in which to be trying to secure a meaningful consent, to say the least. Indeed, in many ways one would have to say that this "consent" is coerced. It is like saying to the civilian population in the theater of operations, "Let us risk your life with AWS; otherwise we will threaten you with our (human) armed forces." While it may be rational for them to prefer AWS to human warfighters, the fact that they do so hardly justifies their use.

The Prospects for Ethical Autonomous Targeting Thus Far

The prospects for ethical autonomous targeting are, therefore, according to my investigation here, mixed. Critics of AWS are correct in holding that the difficulties involved in operating in accordance with the principles of *jus in bello* are profound and unlikely to be resolvable in urban environments for the foreseeable future. On the other hand, in specific limited domains—and, in particular, in operations against naval assets, tanks, self-propelled artillery, and/or aircraft in a given geographical area—it may be possible for robots to distinguish between legitimate and illegitimate targets with a high degree of reliability. Indeed, in this context AWS might prove even *more* reliable than human beings, as Arkin has argued. At the very least, the possibility of deploying AWS in this fashion establishes that they are not, as some have suggested, inherently indiscriminate weapons.

At this stage, then, it would be premature to conclude that any of the ethical arguments I have surveyed thus far stand as an insurmountable barrier to the ethical operations of AWS. If we are to explain the widespread ethical intuition that there is something profoundly disturbing about the prospect of "killer robots" we must delve deeper into the philosophical foundations of just war theory.

There is, as I have argued elsewhere, a case to be made against developing and deploying robotic weapons in general—both tele-operated and autonomous weapon systems—on the basis of the doctrine of *jus ad bellum*.[53] The fact that robotic weapons hold out the prospect of the use of force without risk to one's troops and the likelihood that such systems will be used in more aggressive postures during peace time—again, due to the lack of threat to the life of the "pilot"—suggests that these systems will lower the threshold of conflict and make war more likely.[54] Furthermore, as Paul Kahn has argued, the pursuit of risk-free warfare problematizes the justification of wars of humanitarian intervention by juxtaposing the high value based on the lives of our own military personnel against the lower value placed on the lives of those in the theater of conflict, whose rights and welfare are supposed to justify the intervention, but who are placed at higher risk of death as a result of the use of robotic weapons.[55] However, these sorts of concerns are not specific to AWS and have force against a wider range of means of long-distance war fighting.[56]

AWS and Jus in Bello

If there is going to be anything uniquely morally problematic about AWS, then, the explanation will need to be located within the doctrine of *jus in bello*. In an influential article on the moral foundations of the principles of *jus in bello*, Thomas Nagel argues that the force of these principles can only be explained by the idea that they are founded in absolutist moral reasoning.[57] Nagel develops an essentially Kantian account of the key injunctions of *jus in bello* by way of a principle of respect for the moral humanity of those involved in war. He argues that even during wartime it is essential that we acknowledge the personhood of those with whom we interact and that

> whatever one does to another person intentionally must be aimed at him as a subject, with the intention that he receive it as a subject. It should manifest an attitude to him rather than just to the situation, and he should be able to recognize it and identify himself as its object.[58]

Another way of putting this is that we must maintain an "interpersonal" relationship with other human beings, even during wartime. Obviously, if this principle is to serve as a guide to the ethics of war—rather than as a prohibition against war—the decision to take another person's life must be compatible with such a relationship.[59]

Thus, on Nagel's account, applying the principles of distinction and proportionality involves establishing this interpersonal relationship with those who are the targets of a lethal attack—or who might be killed as a result of an attack targeting another—and acknowledging the morally relevant features that render them combatants or otherwise legitimately subjected to a risk of being killed. In particular, in granting the possibility that they might have a right not to be subject to direct attack by virtue of being a noncombatant, one is acknowledging their humanity.[60] This relationship is fundamentally a relationship between agents—indeed, between members of the Kantian "kingdom of ends." Immediately, then, we can see why AWS might be thought to be morally problematic, regardless of how reliable they might be at distinguishing between legitimate and illegitimate targets.[61] When AWS decide to launch an attack the relevant interpersonal relationship is missing.[62] Indeed, in some fundamental sense there is no one who decides whether the target of the attack should live or die. The absence of human intention here appears profoundly disrespectful.

"*Killer Robots*" *or* "*Robots for Killing*"?

I believe this intuition is central to popular concerns about "killer robots." However, importantly, this way of understanding the ethics of AWS treats a robot as though "it" were doing the killing. Short of the development of artificial intelligences that are actually moral agents, this seems problematic. We might equally well think of a robot as a tool by which one person attempts to kill another—albeit an indeterminate other.[63] The relevant interpersonal relationship would then be that between the officer who authorizes the release of the weapon and those the officer intends to kill. Neither the fact that the person who authorizes the launch does not know precisely who she is killing when she sends an AWS into action nor the fact that the identity of those persons may be objectively indeterminate at the point of launch, seems to rule out the possibility of the appropriate sort of relationship of respect.

When a missile officer launches a cruise missile to strike a set of GPS coordinates 1,000 kilometers away, it is highly unlikely that she knows the identity of those she intends to kill.[64] Similarly, mines and improvised explosive devices (IEDs) kill anyone who happens to trigger them and thus attack persons whose identity is actually indeterminate and not merely contingently unknown. If an interpersonal relationship is possible while using *these* weapons, it is not clear why there could not be an interpersonal relationship between the commanding officer launching AWS and the people these weapons kill. Thus, neither of these features of AWS would appear to function as an absolute barrier to the existence of the appropriate relationship of respect. That said, however, it is important to note that this comparison is not entirely favorable to either AWS or these other sorts of weapons. People often *do* feel uneasy about the ethics of anonymous long-range killing and also—perhaps especially—about landmines and IEDs.[65] Highlighting the analogies with AWS might even render people *more* uncomfortable with these more familiar weapons. Nevertheless, insofar as contemporary thinking about *jus in bello* has yet decisively to reject other sorts of weapons that kill persons whose identity is unknown or actually indeterminate without risk to the user, it might appear illogical to reject AWS on these grounds.

It is also worth noting that the language of machine autonomy sits uneasily alongside the claim that autonomous systems are properly thought of merely as tools to realize the intentions of those who wield them.[66] The more advocates of robotic weapons laud their capacity to make complex decisions without input from a human operator, the more difficult it is to believe that AWS connect the killer and the killed directly enough to sustain the interpersonal relationship that Nagel argues is essential to the principle of distinction. That is to say, even if the machine is not a full moral agent, it is tempting to think that it might be an "artificial agent" with sufficient agency, or a simulacrum of such, to problematize the "transmission" of intention. This is why I have argued elsewhere that the use of such systems may render the attribution of responsibility for the actions of AWS to their operators problematic.[67] As Heather Roff has put it,[68] drawing on the work of Andreas Matthias,[69] the use of autonomous weapons seems to risk a "responsibility gap"; and where this gap exists, it will not be plausible to hold that when a commander sends AWS into action he or she is acknowledging the humanity of those the machines eventually kill.

However, this argument about responsibility has been controversial and ultimately, I suspect, turns upon an understanding of autonomy that is richer and more demanding than that which I have assumed here.[70] At least some of the "autonomous" weapons currently in early development seem likely to possess no agency whatsoever and thus arguably *should* be thought of as transmitting the intentions of those who command their use.

What the Use of AWS Says About Our Attitude Toward Our Enemies

Yet this is not the end of an investigation into the implications of a concern for respect for the ethics of AWS. As Nagel acknowledges, there is a conventional element to our understanding of the requirements of respect.[71] What counts as the humane or inhumane treatment of a prisoner, for instance, or as the desecration of a corpse, is partially a function of contemporary social understandings. Thus, certain restrictions on the treatment of enemy combatants during wartime have ethical force simply by virtue of being widely shared. Moreover, there is ample evidence that existing social understandings concerning the respectful treatment of human beings argue against the use of AWS being ethical. A recent public opinion survey, for example, found high levels of hostility to the prospect of robots being licensed to kill.[72] Most people already feel strongly that sending a robot to kill would express a profound disrespect of the value of an individual human life.[73]

Evidence that what we express when we treat our enemies in a certain way is sometimes crucial to the morality of warfare is provided by how widely shared is the intuition that the mutilation and mistreatment of corpses is a war crime. Such desecration does not inflict "unnecessary suffering" on the enemy; rather, it is wrong precisely because and insofar as it expresses a profound disrespect for their humanity. Importantly, while the content of what counts as a "mistreatment" or "mutilation" is conventional and may change over time, the intuition that we are obligated to treat even the corpses of our enemies with respect is deeper and much less susceptible to revision.

The ethical principles of *jus in bello* allow that we may permissibly attempt to kill our enemy, even using means that will inevitably leave them dying horribly. Yet these principles also place restrictions on the means we may use and on our treatment of the enemy more generally. I have argued—following Nagel—that this treatment should be compatible with respect for the humanity of our enemy and that the content of this concept is partially determined by shared social understandings regarding what counts as respectful treatment. Furthermore, I have suggested that widespread public revulsion at the idea of autonomous weapons should be interpreted as conveying the belief that the use of AWS is incompatible with such respect. If I am correct in this, then even if an interpersonal relationship may be held to exist between the commanding officer who orders the launch of an autonomous weapon system and the individuals killed by that system, it should be characterized as one of *dis*respect.

Interestingly, conceiving of AWS simply as *the means* whereby the person who authorizes the launch of the robot attempts to kill the intended targets vitiates an influential criticism of the campaign to ban these systems.[74] Defenders of AWS have suggested that robotic weapons could not be morally problematic "in themselves" because those who might be killed by robots would die as a result of the effects of weapons that are—understood in a more narrow sense—identical to those that a human might use. In conventional military terminology, Predator drones—and by extension, perhaps, future AWS—would ordinarily be understood as platforms from which a weapon (such as a Hellfire missile) may be delivered.[75] Correspondingly, defenders of AWS claim that it could make no difference to the suffering or the nature of the death of those killed whether the Hellfire missile was fired from an AWS, from a (remotely piloted) Predator drone, or from a (manned) Apache helicopter. Yet if, when it comes to the question of the presence or absence of an "interpersonal" relationship, we are going to understand AWS as a means of attacking targets, we must also understand them as the means the user employs to kill others when it comes to the evaluation of the nature of that means. Indeed, it is quite clear that a combatant who launches AWS is *not* herself launching Hellfire missiles. Consequently, there is nothing especially problematic with the idea that AWS might be an illegitimate means of killing by virtue of being profoundly disrespectful of the humanity of our enemy.

The Case for Banning AWS

I believe that the contemporary campaign to ban autonomous weapons should be understood as an attempt to entrench a powerful intuitive objection to the prospect of a disturbing new class of weapons in international law: AWS should be acknowledged as *mala in se* by virtue of the extent to which they violate the requirement of respect for the humanity of our enemies, which underlies the principles of *jus in bello*.[76] That the boundaries of such respect are sometimes—as in this case—determined by convention (in the sense of shared social understandings rather than formal rules) does not detract from the fact that it is fundamental to the ethics of war.

A number of critics of the campaign to ban AWS have objected that this proposal is premature and that until we have seen robot weapons in action, we cannot judge whether they would be any better or worse, morally speaking, than existing weapons systems.[77] Yet insofar as a ban on AWS is intended to acknowledge that *the use* (rather than the effects) of robotic weapons disrespects the humanity of their targets, this objection has little force.

There is, of course, something more than a little intellectually unsettling about the attempt to place a class of weapons in the category of *mala in se* through legislation or (legal) convention: what is *mala in se* should ideally be recognized independently of positive law. Yet if we are honest about the matter, we will admit that there has always been controversy about the extent of this class of weapons, and that some weapons now held to be evil in themselves were once widely believed to be legitimate means of waging war. Only after a period of contestation and moral argument were technologies such as chemical and nuclear weapons acknowledged as prohibited.[78] The current situation regarding the campaign against AWS is therefore analogous to the way in which the campaigns against the use of chemical weapons at the beginning of the twentieth century and against the use of cluster munitions in the 1990s proceeded.[79] Should this campaign ultimately prove successful, we will understand it to have recognized truths about these weapons that existed independently of—and prior to—the resulting prohibition.[80] In the meantime, the strength and popular currency of the intuition that the use of AWS would profoundly disrespect the humanity of those they are tasked to kill is sufficient justification to try to establish such a prohibition.

Conclusion

The prospect of AWS being capable of meeting the *jus in bello* requirements of distinction and proportionality in the context of counterinsurgency warfare and/or complex urban environments remains remote. However, in particular limited domains, the major barrier to AWS being able to reliably distinguish legitimate from illegitimate targets would appear to be their capacity to detect when enemy forces have surrendered or are otherwise *hors de combat*. If these difficulties can be overcome, then concerns about the capacity of AWS to identify and attack only the appropriate targets are unlikely to rule out the ethical use of these systems.

The strength of the case for autonomous weapons will also depend on how we assess the relative weight of consequentialist and deontological considerations in the ethics of war. If our main consideration is to reduce the number of noncombatant deaths, it becomes easier to imagine AWS being ethical: they would simply have to be better than human beings at distinguishing between legitimate and illegitimate targets in some given domain.[81] However, if we are concerned with what we owe noncombatants and others who are not legitimately subject to lethal force, then the merely statistical form of discrimination achievable by robots may be insufficient.

The deeper issue regarding the ethics of AWS, though, concerns whether the use of these weapons is compatible with the requirement of respect for the humanity of our enemies, which underpins the principles of *jus in bello*. If we understand AWS as "artificial agents" that choose which targets to attack and when, it is likely that the necessary relationship of respect is absent and, therefore, that their use would be unethical. Yet in many cases it may in fact be more plausible to consider AWS as the means whereby the person who is responsible for their launch kills those that the AWS are tasked to attack. However, this means of killing may itself be unethical insofar as it expresses a profound disrespect for the humanity of our enemies. Because this argument relies on a reference to conventions—that is, to social expectations that acquire normative force simply by virtue of being widely shared—to settle the question of what respect requires, the case against AWS is much weaker than critics might prefer. Nevertheless, the line of argument I have developed here is still equal to the task of justifying an international treaty prohibiting the development and deployment of AWS on the grounds that such weapons are "evil in themselves."

There are, of course, further questions about whether it is realistic to imagine such a prohibition coming into force, let alone being effective in preventing (or at least significantly delaying) the deployment of AWS.[82] States that have the capacity to develop or field such weapons will also have to confront the question of whether the ethical case for any such treaty is worth whatever sacrifice of military advantage that might result from signing it.[83] These are matters for further discussion and argument—and where, moreover, philosophers may have little to contribute.[84] What I have shown here is that there is an ethical case to be made for working toward such a treaty.

NOTES

[1] Kenneth Anderson and Matthew C. Waxman, "Law and Ethics for Robot Soldiers," *Policy Review* 176 (2012); Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots* (Boca Raton Fla.: CRC Press, 2009); Gary E. Marchant et al., "International Governance of Autonomous Military Robots," *Columbia Science and Technology Law Review* 12 (2011); Department of Defense, *Unmanned Systems Integrated Roadmap: FY2011–2036* (Washington, D.C.: Department of Defense, 2012); Michael N. Schmitt and Jeffrey S. Thurnher, "'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict," *Harvard National Security Journal* 4, no. 2 (2013); Peter W. Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (New York: Penguin Press,

2009); Robert O. Work and Shawn Brimley, *20YY: Preparing for War in the Robotic Age* (Centre for a New American Security, 2014). Elsewhere in the literature, AWS are sometimes referred to as Lethal Autonomous Robots (LARs).

[2] Human Rights Watch, *Losing Humanity: The Case against Killer Robots* (2012), www.hrw.org/reports/2012/11/19/losing-humanity-0.

[3] Campaign to Stop Killer Robots website, "Campaign to Stop Killer Robots: About Us," www.stopkillerrobots.org/about-us/.

[4] Charli Carpenter, "Beware the Killer Robots: Inside the Debate over Autonomous Weapons," *Foreign Affairs* online, July 3, 2013, www.foreignaffairs.com/articles/139554/charli-carpenter/beware-the-killer-robots#.

[5] I will leave the task of determining the *legality* of AWS under international humanitarian law to those better qualified to address it. However, insofar as the just war theory doctrine of *jus in bello* is developed and expressed in both legal and philosophical texts, I will occasionally refer to the relevant legal standards in the course of my argument, which concerns the *ethics* of autonomous targeting.

[6] For the argument that this problem also impacts on the science and engineering of AWS, see the U.S. Department of Defense, Defense Science Board, *The Role of Autonomy in DoD Systems* (Washington, D.C.: Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, 2012), pp. 23–24.

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

[8] Human Rights Watch, *Losing Humanity*, pp. 6–20; Schmitt and Thurnher, "Out of the Loop"; Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Farnham, U.K.: Ashgate Publishing, 2009), pp. 43–45.

[9] Heather M. Roff, "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, "Killer Robots."

[10] I have discussed this question at more length in Sparrow, "Killer Robots."

[11] U.S. Department of Defense, "DoD Directive 3000.09: Autonomy in Weapon Systems," Washington, D.C., November 21, 2012.

[12] The Phalanx Close-In Weapon System is fitted to U.S. (and U.S. ally) ships to defend them from missiles and aircraft. It uses a sophisticated radar to identify, track, and target incoming threats and a high-speed gatling gun to attempt to destroy them. While the system allows for a manual override, it is designed to engage threats automatically due to the high speed at which engagements must occur.

[13] Daniel C. Dennett, *The Intentional Stance* (Cambridge, Mass.: MIT Press, 1987).

[14] Andreas Matthias, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata," *Ethics and Information Technology* 6, no. 3 (2004).

[15] Anderson and Waxman, "Law and Ethics for Robot Soldiers"; Arkin, *Governing Lethal Behavior in Autonomous Robots*; Marchant et al., "International Governance of Autonomous Military Robots"; Work and Brimley, *20YY*. The United States Department of Defense clarified its own policy in relation to the development and use of AWS in "DoD Directive 3000.09: Autonomy in Weapon Systems" (2012), which some (see, for instance, Spencer Ackerman, "Pentagon: A Human Will Always Decide When a Robot Kills You," *Wired*, November 26, 2012) have read as prohibiting the use of AWS armed with lethal weapons against human targets. However, see Mark Gubrud, "US Killer Robot Policy: Full Speed Ahead," *Bulletin of the Atomic Scientists*, September 20, 2013.

[16] Kenneth Anderson and Matthew C. Waxman, "Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can," *Hoover Institution, Jean Perkins Task Force on National Security and Law Essay Series* (2013), p. 7; Schmitt and Thurnher, "Out of the Loop," p. 238.

[17] Thomas K. Adams, "Future Warfare and the Decline of Human Decisionmaking," *Parameters* 31, no. 4 (2001/2002).

[18] Schmitt and Thurnher, "Out of the Loop." The costs of developing and fielding AWS are another matter entirely. Complex information technology systems are notoriously prone to running over budget and delivering less than was promised. Developing the computer software required for autonomy and debugging it effectively may be very expensive indeed.

[19] Mark Gubrud, "Stopping Killer Robots," *Bulletin of the Atomic Scientists* 70, no. 1 (2014); Human Rights Watch, *Losing Humanity*; International Committee for Robot Arms Control website, *Mission Statement*, icrac.net/statements/; Robert Sparrow, "Predators or Plowshares? Arms Control of Robotic Weapons," *IEEE Technology and Society* 28, no. 1 (2009).

[20] Ronald C. Arkin, "On the Ethical Quandaries of a Practicing Roboticist: A First-Hand Look," in Adam Briggle, Katinka Waelbers, and Philip Brey, eds., *Current Issues in Computing and Philosophy* (Amsterdam: IOS Press, 2008); Arkin, *Governing Lethal Behavior in Autonomous Robots.*

[21] Ronald C. Arkin, "The Case for Ethical Autonomy in Unmanned Systems," *Journal of Military Ethics* 9, no. 4 (2010). See also Marchant et al., "International Governance of Autonomous Military Robots," pp. 279–81; Department of Defense, *Unmanned Systems Integrated Roadmap: FY2011–2036*, pp. 43–51.

[22] For a critical evaluation of these claims, see Ryan Tonkens, "The Case against Robotic Warfare: A Response to Arkin," *Journal of Military Ethics* 11, no. 2 (2012).

[23] Human Rights Watch, *Losing Humanity*; Noel E. Sharkey, "Autonomous Robots and the Automation of Warfare," *International Humanitarian Law Magazine*, no. 2 (2012); Noel E. Sharkey, "The Evitability of Autonomous Robot Warfare," *International Review of the Red Cross* 94, no. 886 (2012).

[24] Light Detection And Ranging (LIDAR).

[25] For a useful discussion of the ways in which applying the principle of distinction requires assessment of intention and of just how hard this problem is likely to be for a machine, see Marcello Guarini and Paul Bello, "Robotic Warfare: Some Challenges in Moving from Noncivilian to Civilian Theaters," in Patrick Lin, Keith Abney, and George A. Bekey, eds., *Robot Ethics: The Ethical and Social Implications of Robotics* (Cambridge, Mass.: MIT Press, 2012). Again, as Guarini and Bello concede and I will discuss further below, in some—very restricted—circumstances it *may* be reasonable to treat every person carrying a weapon and every weaponized system, within a narrowly defined geographical area, as a combatant.

[26] Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), adopted at Geneva on June 8, 1977, www.icrc.org/Ihl.nsf/INTRO/470?OpenDocument.

[27] Geoffrey S. Corn et al., *The Law of Armed Conflict: An Operational Approach* (New York: Wolters Kluwer Law & Business, 2012), pp. 115–17.

[28] Ibid., pp. 165–66; Additional Protocol I (see note 27).

[29] In the course of writing this article, I was fortunate enough to read a draft of a manuscript by Heather Roff addressing the prospects for autonomous weapons meeting the requirements of the principle of distinction. My discussion here has undoubtedly been influenced by her insightful treatment of the topic. Schmitt and Thurnher, "Out of the Loop" also contains a useful discussion of this question.

[30] Schmitt and Thurnher, "Out of the Loop," pp. 265–66; Markus Wagner, "Taking Humans Out of the Loop: Implications for International Humanitarian Law," *Journal of Law Information and Science* 21, no. 2 (2011).

[31] Robert Sparrow, "Twenty Seconds to Comply: Autonomous Weapon Systems and the Recognition of Surrender," *International Law Studies* 91 (2015).

[32] That importance of this requirement is noted in Marchant et al., "International Governance of Autonomous Military Robots," p. 282.

[33] Arkin, *Governing Lethal Behavior in Autonomous Robots*.

[34] For an account of what would be required to produce "ethical robots" that is more sympathetic to the idea than I am here, see Wendell Wallach and Colin Allen, *Moral Machines: Teaching Robots Right from Wrong* (New York: Oxford University Press, 2009).

[35] Raimond Gaita, *Good and Evil: An Absolute Conception*, 2nd ed. (Abingdon, U.K.: Routledge, 2004), pp. 264–82.

[36] Arkin, *Governing Lethal Behavior in Autonomous Robots*, pp. 203–209.

[37] Donald P. Brutzman et al., "Run-Time Ethics Checking for Autonomous Unmanned Vehicles: Developing a Practical Approach" (paper presented at the 18th International Symposium on Unmanned Untethered Submersible Technology, Portsmouth, New Hampshire, 2013); Alex Leveringhaus and Tjerk de Greef, "Keeping the Human 'in-the-Loop': A Qualified Defence of Autonomous Weapons," in Mike Aaronson et al., eds., *Precision Strike Warfare and International Intervention: Strategic, Ethico-Legal, and Decisional Implications* (Abingdon, U.K.: Routledge, 2015).

[38] Adams, "Future Warfare and the Decline of Human Decisionmaking."

[39] Anderson and Waxman, "Law and Ethics for Autonomous Weapon Systems," p. 7; Schmitt and Thurnher, "Out of the Loop," p. 238; Work and Brimley, *20YY*, p. 24.

[40] See also Brutzman et al., "Run-Time Ethics Checking for Autonomous Unmanned Vehicles." This is not to deny that there would be some military advantages associated with the development of such systems, as long as the communications infrastructure necessary to allow contact with a human operator as required was in place. For instance, by removing the need for direct human supervision it would multiply the number of systems that could operate in the context of a given amount of bandwidth and also make it possible for one human operator to oversee the activities of a number of robots.

[41] Guarini and Bello, "Robotic Warfare"; Human Rights Watch, *Losing Humanity*; Sharkey, "The Evitability of Autonomous Robot Warfare." Indeed, I have argued this myself elsewhere. See Robert Sparrow, "Robotic Weapons and the Future of War," in Jessica Wolfendale and Paolo Tripodi, eds., *New Wars and New Soldiers: Military Ethics in the Contemporary World* (Surrey, U.K.: Ashgate, 2011).

[42] Michael N. Schmitt, "Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics," *Harvard National Security Journal* online, February 5, 2013.

[43] Brutzman et al., "Run-Time Ethics Checking for Autonomous Unmanned Vehicles."

[44] Guarini and Bello, "Robotic Warfare."

[45] Leveringhaus and De Greef, "Keeping the Human 'in-the-Loop.'"

[46] To the extent that this requirement is not recognized by the Law of Armed Conflict, the *legal* barriers to the ethical use of AWS in sufficiently restricted domains will be correspondingly lower.

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

[48] Arkin, "The Case for Ethical Autonomy in Unmanned Systems."

[49] The argument in this paragraph owes much to remarks made by Daniel Brunstetter in a session on drone warfare at the International Studies Association Annual Convention in San Francisco in April 2013. See also Megan Braun and Daniel R. Brunstetter, "Rethinking the Criterion for Assessing CIA-Targeted Killings: Drones, Proportionality and *Jus Ad Vim*," *Journal of Military Ethics* 12, no. 4 (2013). George Lucas subsequently brought it to my attention that he had in fact rehearsed this argument in a paper in 2011. See George R. Lucas Jr, "Industrial Challenges of Military Robotics," *Journal of Military Ethics* 10, no. 4 (2011).

[50] Lucas Jr, "Industrial Challenges of Military Robotics."

[51] Brian G. Williams, *Predators: The CIA's Drone War on Al Qaeda* (Washington, D.C.: Potomac Books, 2013).

[52] Williams's argument proceeds by providing evidence of actual consent but in most cases the argument will need to proceed by way of reference to "hypothetical consent"—that is, what civilians in the area of operations *would* prefer.

[53] Sparrow, "Robotic Weapons and the Future of War."

[54] Singer, *Wired for War*, p. 319; Sparrow, "Predators or Plowshares?"

[55] Paul W. Kahn, "The Paradox of Riskless Warfare," *Philosophy & Public Policy Quarterly* 22, no. 3 (2002).

[56] Bradley J. Strawser, "Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles," *Journal of Military Ethics* 9, no. 4 (2010).

[57] Thomas Nagel, "War and Massacre," *Philosophy & Public Affairs* 1, no. 2 (1972).

[58] Ibid., p. 136.

[59] Ibid., p. 138.

[60] Peter Asaro, "On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making," *International Review of the Red Cross* 94, no. 886 (2012), p. 701.

[61] Asaro, "On Banning Autonomous Weapon Systems."

[62] Mary Ellen O'Connell, "Banning Autonomous Killing: The Legal and Ethical Requirement That Humans Make Near-Time Lethal Decisions," in Matthew Evangelista and Henry Shue, eds., *The American Way of Bombing: Changing Ethical and Legal Norms, from Flying Fortresses to Drones* (Ithaca, N.Y.: Cornell University Press, 2014).

[63] Arkin, "Lethal Autonomous Systems and the Plight of the Non-Combatant"; Vik Kanwar, "Post-Human Humanitarian Law: The Law of War in the Age of Robotic Weapons," *Harvard National Security Journal* 2, no. 2 (2011), pp. 619–20; Schmitt and Thurnher, "Out of the Loop," p. 268.

[64] She should, of course, be appropriately confident that the persons at the facility or location they are attacking are legitimate targets.

[65] Indeed, the stockpiling and use of antipersonnel mines at least is prohibited by the Ottawa Treaty.

[66] Defense Science Board, *The Role of Autonomy*, pp. 1–2.

[67] Sparrow, "Killer Robots."

[68] Roff, "Killing in War."

[69] Matthias, "The Responsibility Gap."

[70] Thomas Hellström, "On the Moral Responsibility of Military Robots," *Ethics and Information Technology* 15, no. 2 (2013); Leveringhaus and De Greef, "Keeping the Human 'in-the-Loop'"; Gert-Jan Lokhorst and Jeroen van den Hoven, "Responsibility for Military Robots," in Lin, Abney, and Bekey, *Robot Ethics*; Sparrow, "Killer Robots."

[71] Nagel, "War and Massacre," p. 135, note 7.

[72] Charli Carpenter, "US Public Opinion on Autonomous Weapons" (University of Massachusetts, 2013), www.duckofminerva.com/wp-content/uploads/2013/06/UMass-Survey_Public-Opinion-on-Autonomous-Weapons.pdf.

[73] Gubrud, "Stopping Killer Robots," p. 40; Aaron M. Johnson and Sidney Axinn, "The Morality of Autonomous Robots," *Journal of Military Ethics* 12, no. 2 (2013); Sparrow, "Robotic Weapons and the Future of War."

[74] Asaro, "On Banning Autonomous Weapon Systems."

[75] Schmitt and Thurnher, "Out of the Loop," p. 10.

[76] Gubrud, "Stopping Killer Robots"; Human Rights Watch, *Losing Humanity*; Wendell Wallach, "Terminating the Terminator: What to Do About Autonomous Weapons," *Science Progress*, January 29, 2013. The legal basis for doing so might be found in the "Martens Clause" in the Hague Convention, which prohibits weapons that are "contrary to the dictates of the public conscience." Human Rights Watch, *Losing Humanity*.

[77] Arkin, "Lethal Autonomous Systems and the Plight of the Non-Combatant"; Anderson and Waxman, "Law and Ethics for Robot Soldiers"; Anderson and Waxman, "Law and Ethics for Autonomous Weapon Systems"; Schmitt and Thurnher, "Out of the Loop."

[78] The moral status of nuclear weapons remains controversial in some quarters. However, the last two decades of wars justified with reference to states' possession of "weapons of mass destruction" suggests that there is an emerging international consensus that such weapons are *mala in se*.

[79] Johnson and Axinn, "The Morality of Autonomous Robots," p. 137.

[80] Should this campaign fail, it is possible that public revulsion at sending robots to kill people will be eroded as AWS come into use and become a familiar feature of war—as has occurred with a number of weapons, including artillery and submarines, in the past. In that case, the argument that such killing disrespects the humanity of our enemies will eventually lapse as the social conventions around respect for the humanity of combatants are transformed. It might be argued that even if a prohibition on AWS is achieved, conventional understandings of the appropriate relations between humans and robots may shift in the future as people become more familiar with robots in civilian life. While this cannot be ruled out a priori, I suspect that it is more likely that outrage at robots being allowed to kill humans will only intensify as a result of the social and psychological incentives to maintain the distinction between "us" and "them."

[81] Arkin, "Lethal Autonomous Systems and the Plight of the Non-Combatant."

[82] For cynicism about the prospect of such, see Anderson and Waxman, "Law and Ethics for Autonomous Weapon Systems"; Arkin, "Lethal Autonomous Systems and the Plight of the Non-Combatant"; Marchant et al., "International Governance of Autonomous Military Robots." For a countervailing perspective, see O'Connell, "Banning Autonomous Killing."

[83] Wendell Wallach and Colin Allen, "Framing Robot Arms Control," *Ethics and Information Technology* 15, no. 2 (2013); Schmitt and Thurnher, "Out of the Loop."

[84] For an important contribution to this project, see Jürgen Altmann, "Arms Control for Armed Uninhabited Vehicles: An Ethical Issue," *Ethics and Information Technology* 15, no. 2 (2013).