

Can Lethal Autonomous Weapons Be Just?

Noreen Herzfeld

IN 2018 THE UNITED STATES DEPARTMENT of Defense (DoD) created a new Joint Artificial Intelligence Center to study the adoption of AI by the military. Their strategy, outlined in a document entitled “Harnessing AI to Advance Our Security and Prosperity,” proposes to accelerate the adoption of AI by fostering “a culture of experimentation and calculated risk taking,” noting that AI will soon “change the character of the future battlefield and the pace of threats we must face.”¹ The report cautions that Russia and China are investing in AI for military purposes. While it is the DoD’s intention to keep up in the AI arms race, the report states that we will “undertake research and adopt policies as necessary to ensure that AI systems are used responsibly and ethically.”² How will we determine what is responsible or ethical in this new technological age? Just as the advent of nuclear weapons caused twentieth-century theologians to reevaluate the justness of war, the advent of lethal autonomous weapon systems (LAWS) presents an even more urgent call to twenty-first century theologians to do likewise. For while nuclear weapons required human decision-makers, autonomous systems may not. While nuclear bombs are extremely costly and difficult to produce, intelligent weapons are becoming smaller and cheaper and can be used across the spectrum of conflict, from the highest to the lowest level.

Weapons of war have long had a certain degree of autonomy. Heat seeking missiles can change their course. Defensive systems include automatic modes that target inbound projectiles independent of any human decision.³ With its potential to bring autonomy to a new level, AI forces us to consider just how independent we want our weapons to be and what difference this heightened autonomy makes to the ethical conduct of war. While bombs, land mines, missiles, and drones do not always involve a direct human decision as to whom they target, nor when exactly they wreak their havoc, they do not make decisions. They cannot decide *not* to explode when triggered. Nor can they target

¹ US Department of Defense, “Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity,” (2018), §4, media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

² US Department of Defense, “Artificial Intelligence Strategy.”

³ See for example en.wikipedia.org/wiki/Phalanx_CIWS.

with any degree of specificity. A landmine targets whoever steps on it; a bomb maims or kills whoever is within range. The use of artificially intelligent weapons, especially when combined with capabilities such as facial recognition, inaugurates a new era in weaponry, one which differs from what has preceded it in kind rather than merely degree. Autonomous weapons take over not just the physical, but many of the mental decisions of the battlefield. Just as the necessity for the physical presence of soldiers limited the destruction of war prior to the twentieth century, so the mental limitations of human decision making have continued to function as a limiting principle. Autonomous weapons risk moving humankind into an era of warfare that moves with unprecedented speed, precision, and unexpected consequences.

Ideally, such weapons should be banned, and both individuals and international bodies have indeed called for such a ban. However, given the number of countries already engaging in their development, and the advantages possession of autonomous weapons could confer, an international ban is extremely unlikely. Lacking such a ban, we must at least examine the challenges such weapons will bring to the initiation, execution, and ending of wars and develop procedures and guidelines that ensure these weapons are used in a restrained and just manner. The tradition of just war theory, an amorphous set of rules and justifications rooted in Christian thought and consisting of “articulated norms, customs, professional codes, legal precepts, religious and philosophical principles, and reciprocal arrangements that shape our judgments of military conduct,” has provided justification for both when to go to war and acceptable conduct within a war in the past.⁴ These precepts illuminate some of the ethical quandaries of warfare that LAWS will exacerbate and highlight areas in which we will need to develop guidelines and policies we currently lack.

TRADITIONAL PRINCIPLES OF JUST WARFARE

There is no one set of principles of just warfare. We find an early rule of conduct for warfare in Deuteronomy 20:19–20 which forbids cutting down an enemy’s fruit trees during a siege. This verse counsels both restraint against needless environmental destruction while also hinting at a principle of distinction, for destruction of an enemy’s means of producing food harms combatants and noncombatants alike. A Christian tradition of just warfare can be traced back to the pre-Christian thought of Aristotle’s *Politics* and expanded upon by Augustine in *The City of God* and *Against Faustus the Manichaeon*. It is, however, Thomas Aquinas, in his *Summa Theologiae*, who first lays out the general outline of what we now regard as traditional just war

⁴ Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, fifth edition (New York: Basic Books, 2015), 44.

theory. These principles were adjusted and universalized to include conduct toward non-Christians by later Scholastics including Vitoria, Suarez, Grotius and Wolff⁵ in light of European colonial expansion, and further rethought in the twentieth century by Walzer, Nagel, Norman, and others in light of the changing face of warfare brought about by the atomic bomb and rise of international terrorism.⁶ The current field of just war theory is contentious, in that how a state should regulate war and what an individual may be morally obligated to do can conflict. Autonomous weapons add to this confusion by introducing a new potential actor beyond the state and individual, namely, the weapon itself.

Michael Walzer, in his highly regarded *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, lays out a set of traditional principles for when it is permissible to fight and how to engage ethically in battle. *Jus ad bellum*, when it is justifiable to fight, includes fighting for a just cause, either to resist attack, protect innocent life in imminent danger or, as phrased by the US Catholic bishops, to “correct a grave, public evil” such as genocide or a massive violation of a group’s basic human rights. Hostilities should be declared by a proper authority, be a last resort, fought only for the purpose that initiated it, have a reasonable chance of success, and be proportional, in that the anticipated benefits of waging war outweigh the expected harm. The calculation of the costs of war should include those paid by one’s enemy as well as one’s own, both tangible and intangible. These principles act together to militate against going to war, for as Pope Pius XII stated on the eve of World War II, “Nothing is to be lost with peace; everything can be lost with war.”⁷

Jus in bello, or proper conduct once engaged in hostilities, includes distinguishing between combatants and non-combatants, limiting the level of force to the minimum necessary to attain one’s ends, and acting solely with the intention of righting the wrong. Acts of vengeance and indiscriminate violence are forbidden as are intrinsically evil methods, such as mass rape, forcing enemy combatants to fight against

⁵ See Francisco de Vitoria, *Relectiones theologiae* (Lyons: Apud Jacobum Boyerium, 1557); Francisco Suárez, *De triplici virtute theologia* (London: Clarendon, 1944); Hugo Grotius, *The Rights of War and Peace: Including the Law of Nature and of Nations* (New York: M. W. Dunne, 1901); Christian Freiherr von Wolff, *The Law of Nations Treated According to the Scientific Method* (Carmel, IN: Liberty Fund, 2017).

⁶ Michael Walzer, *Just and Unjust Wars*; Thomas Nagel, “War and Massacre,” *Philosophy & Public Affairs* 1, no. 2 (1972): 123–44; Richard Norman, *Ethics, Killing, and War* (Cambridge: Cambridge University Press, 1995); G. E. M. Anscombe, “War and Murder,” in *Nuclear Weapons: A Catholic Response*, ed. Walter Stein (London: Sheed and Ward, 1961), 44–52.

⁷ Quoted by Rev. Diarmuid Martin, United States Council of Catholic Bishops, “Theological and Moral Perspectives on Today’s Challenge of Peace” (2011), www.usccb.org/issues-and-action/human-life-and-dignity/september-11/theological-and-moral-perspectives-on-todays-challenge-of-peace.cfm.

or betray their own side, or the use of weapons whose effects are uncontrollable, such as chemical or biological weapons. Civilians are to be safeguarded as much as possible.

Later scholars, such as Canadian philosopher Brian Orend and ethicists Mark Allman and Tobias Winright, have added to these two categories principles of conduct following the cessation of hostilities (*jus post bellum*).⁸ After the cessation of hostilities, the victorious party has the responsibility to lay down arms, enter into relevant treaties, and remove soldiers and weapons from the field of battle. They should also aid in the political and economic reconstruction of the defeated community or state. Both parties should be held responsible for war crimes or atrocities committed during hostilities and the victor is responsible for seeing that suitable restitution or reparations are made, especially in the case of genocide.

A final category, *jus ad vim* (introduced by Michael Walzer in the fourth edition of his seminal text), examines use of force in conditions falling short of full warfare. This category remains controversial due to the difficulty in determining where the line falls between acts just short of warfare and acts of terrorism, criminality, or policing. Many countries have seen the adoption of military-grade hardware by police departments and civilian branches of government. Autonomous weapons might in the future be used domestically in efforts to fight crime or, more sinisterly, to target political opponents, and internationally in efforts to stem terrorism or combat drug and weapons markets. LAWS hold the potential to make assassination or small targeted strikes easier and hence, more likely. Such uses demand a consideration of their potential to escalate a hostile situation and of their effect on the hearts and minds of a populace.

The rapid rise of new technologies of warfare in the twentieth and twenty-first centuries has sparked a commensurate rise in interest in and public debate of the just-war tradition. These precepts have become a part of political deliberations on the use of force and in military training in several countries. The US Council of Catholic bishops sees them as not just “a set of ideas, but as a system of effective social constraints on the use of force.” Their application is neither straightforward nor easy, but for the bishops, the “increasing violence of our society, its growing insensitivity to the sacredness of life and the glorification of the technology of destruction in popular culture” call for their use. The speed, precision, and lethality of modern weapons, all characteristics enhanced by AI, make it all the more important that our

⁸ See Brian Orend, “Jus Post Bellum,” *Journal of Social Philosophy* 31 (2000): 117–37; Mark Allman and Tobias Winright, *After the Smoke Clears: The Just War Tradition and Post War Justice* (Maryknoll, NY: Orbis, 2010).

decisions regarding the use of lethal force at least attempt to pass the “hard tests set by the just-war tradition.”⁹

LETHAL AUTONOMOUS WEAPONS AND THE “HARD TESTS” OF JUST WAR THEORY

Computer assisted weapons of war exist along a scale of autonomy. At one end we have traditional “fire and forget” weapons, such as some guided missiles where a human operator selects a target and launches the missile, which then uses sensors and algorithms to complete the task. These clearly depend on a human operator being “in the loop.” US Department of Defense Directive 3000.09 notes two other levels of autonomy. “Human on the loop” are weapons that act autonomously but under human supervision, where an operator can monitor the weapon and halt or alter its engagement. Fully autonomous weapons, where the human is “out of the loop” are “weapon system[s] that, once activated, can select and engage targets without further intervention by a human operator.”¹⁰ Philosopher Robert Sparrow adds that, to be seen as truly autonomous, such a system must be sufficiently complex, “such that, even when it is functioning perfectly, there remains some uncertainty about which targets it will attack and why.”¹¹

Lethal autonomous weapon systems have moved in the past decade from the semi-autonomy of “human in the loop” to being able to autonomously identify a target and decide, without synchronous human control, whether to attack or destroy it. As such, they potentially have not only a high degree of agency, acting without the direct control of another to achieve a chosen result but, since their targets are often human beings, they act as moral agents, making judgments and decisions that carry the potential for life or death. The advent of flight inaugurated a new era of warfare, releasing armies from physical presence on the field of battle. Fully autonomous weapons will inaugurate a third era, releasing soldiers from the mental decisions of the battlefield as well. According to legal scholar Rebecca Crootof, their “capacity for self-determined action makes them uniquely effective and uniquely unpredictable.”¹² Former USAF Major General Robert Latiff

⁹ National Conference of Catholic Bishops, *The Harvest of Justice is Sown in Peace* (Washington, DC: US Catholic Conference, 1994), C. The Centrality of Conscience.

¹⁰ Congressional Research Service, “Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems,” *In Focus*, December 1, 2020, fas.org/sgp/crs/natsec/IF11150.pdf.

¹¹ Robert Sparrow, “Killer Robots,” *Journal of Applied Philosophy* 24, no. 1 (2007): 64.

¹² Rebecca Crootof, “War Torts: Accountability for Autonomous Weapons,” *University of Pennsylvania Law Review* 164, no. 6 (2016): 1349.

has called this transformation the crossing of a new “moral Rubicon.”¹³

The urgency of considering the ethics of LAWS is exacerbated by their evolution in the last few decades from large and costly systems to weapons that are, in the words of Marine Col. James Jenkins, “small, smart, cheap, and abundant.”¹⁴ An early semi-autonomous system, the AEGIS naval air defense has increased in autonomy over the last fifty years. Currently it is able to search in the air, on the surface, and underwater, track and guide missiles, and decide autonomously when and where to fire. It can function both fully autonomously or in “human on the loop” mode with operators having the option to override its decisions.¹⁵ It is a big and costly system and currently only being updated by the US and Japan. The HAROP loitering missile, a smaller and more autonomous system developed by Israel and first used by Turkey in 2005, can once launched be controlled either via a two-way data link for “human in the loop” operation or programmed to autonomously recognize and attack high-value targets.¹⁶ Even smaller, the Kargu-2 is a 15-pound multi-copter drone that can be controlled directly or operate autonomously to track, identify (via facial recognition), and engage targets. These drones can work autonomously in swarms of up to 20, either with a human operated drone leading the swarm or fully independently. Turkey has ordered 500 of these for military surveillance and possible attack capabilities.¹⁷ Unlike missiles, drones can be sent to hunt down enemy targets and return, weapons unspent, if none are found. Large swarms of drones can be launched, and if only a small percentage find their target, you still have a win. At least sixteen countries possess armed drones. So far, they operate with humans in the loop, but this could easily change as facial recognition and AI decision making improve. Unlike the nuclear club, limited to a handful of nations, LAWS will proliferate much more easily and widely.

These and other autonomous weapons present military commanders with a variety of incentives for use. They can process vast amounts of data and operate at speeds and levels of precision far beyond human

¹³ Robert Latiff and Patrick McCloskey, “With Drone Warfare, America Approaches the Robo-Rubicon,” *Wall Street Journal*, March 14, 2013, www.wsj.com/news/articles/SB1000142412788732412850457834633246145590.

¹⁴ Jon Harper, “Navy, Marine Corps Officials Worried about Cost-Effectiveness of Unmanned Systems,” *National Defense*, April 5, 2017, www.nationaldefensemagazine.org/articles/2017/4/5/navy-marine-corps-officials-worried-about-costeffectiveness-of-unmanned-systems.

¹⁵ “AEGIS Weapon System,” *US Navy Fact File*, www.navy.mil/navydata/fact_display.asp?cid=2100&tid=200&ct=2.

¹⁶ “HAROP Loitering Munitions System,” www.iai.co.il/p/harop.

¹⁷ David Hambling, “Turkish Military to Receive 500 Swarming Kamikaze Drones,” *Forbes*, June 17, 2020. www.forbes.com/sites/davidhambling/2020/06/17/turkish-military-to-receive-500-swarming-kamikaze-drones/#4887e2c5251a.

capabilities, including making rapid decisions in changing circumstances. They can operate in harsh and difficult environments, such as underwater. They are less expensive than human troops and can work long hours without tiring. They can carry out orders with fewer mistakes. Most important, they keep soldiers out of physically and psychologically dangerous or deadly environments. However, these advantages do not come without costs. Just as twentieth-century ethicists and theologians were forced to reevaluate the justness of war in the light of nuclear weapons, so now must we reevaluate the morality of war in light of autonomous weapons. In what ways does the advent of these weapons affect our decisions on when to wage war, how to wage war, and who is responsible for the acts of war? We will examine one question raised by each category of just war theory to provide a brief and partial answer to this question.

Jus ad Bellum: Would LAWS Make War Too Easy?

Theologian Brian Stiltner argues that LAWS could make war too easy. He recalls an episode of the original *Star Trek* entitled “A Taste of Armageddon,” in which two planets, Eminar and Vendikar, have completely computerized warfare. Attacks are simulated by computers and those unlucky enough to have been “victims” of the simulated attack are required to report to disintegration chambers. Both planets claim to have found a way of maintaining their infrastructure despite being at war, removing the brutality and destruction. Captain Kirk destroys one planet’s computers, stating that war is inherently brutal and messy, which keeps us from going to war lightly or perpetuating it too long.¹⁸

Kenneth Payne, British scholar of international affairs, argues that like the computers on Eminar and Vendikar, LAWS will not only remove too many of the psychological barriers to war, but will further privilege offense over defense.¹⁹ Consider an attack by a swarm of drones such as the Kargu-2. They can attack *en masse* and then rapidly disperse, leaving little target for a counterattack. Autonomous weapons can also function as “suicide bombers” with no loss of life for the attacker. The direction of the attack can be various and loitering missiles can bide their time or come and go in unpredictable ways. The attacking AI will be expert in observing and analyzing terrain, removing that advantage from the defender, and will be unaffected by fatigue, mental strain, or emotional compunctions.

¹⁸ See Brian Stiltner, “A Taste of Armageddon: When Warring is Done by Drones and Robots,” in *Can War Be Just in the 21st Century?*, eds. Tobias Winright and Laurie Johnston (Maryknoll, NY: Orbis, 2015).

¹⁹ Kenneth Payne, “Artificial Intelligence: A Revolution in Strategic Affairs?,” *Vex Machina*, September 18, 2018, www.tandfonline.com/doi/full/10.1080/00396338.2018.1518374.

Second, use of LAWS removes the constraint of soldiers' lives being put at risk, significantly lowering the cost of an attack. As an editorial in *The Economist* points out, "A president who sends someone's son or daughter into battle has to justify it publicly, as does the congress responsible for appropriations and a declaration of war. But if no one has children in danger, is it a war?"²⁰ Swarms of intelligent drones would moreover add little material cost to an attack. LAWS thus prove to be a serious risk to the just war precept of last resort. They privilege offense, remove the psychological barrier to putting "boots on the ground," and could significantly lower the cost of an attack, making war seem a more desirable option than it would have been if waged with conventional soldiers and weapons.

Jus in Bello: Can a Robot Act Ethically?

Once at war, the principles of *jus in bello* demand that one act with restraint, that one refrains from gratuitous killing of civilians and excess destruction, and that soldiers conduct themselves with virtue and propriety. Georgia Tech roboticist Ron Arkin has argued that LAWS have the potential to act more virtuously than humans. Arkin cites a report from the Surgeon General's Office assessing the battlefield ethics of US soldiers and marines in which ten percent reported mistreating noncombatants and roughly thirty percent reported facing ethical situations in which they did not know how to respond.²¹ Soldiers, under pressure, react emotionally: "Fear and hysteria are always latent in combat, often real, and they press us toward fearful measures and criminal behavior."²² An analysis of civilian casualties in the second Iraqi war found that most were either the result of ethnic cleansing or caused by indiscriminate suppressive fire between sides. LAWS have no emotions, so will not react out of panic or vengeance. Similarly, they have no need for self-protection. They would follow orders more exactly and can integrate information regarding a changing battle scenario faster before responding with lethal force, thus acting with more precision and fewer mistakes. Arkin believes AIs could better discriminate between combatants and noncombatants, thus committing fewer war crimes and reducing civilian casualties.²³

Not everyone agrees. John Kaag and Whitley Kaufman argue that moral judgment is inherently ambiguous. Were the laws of war reducible to a set of simple rules, "It is likely that we would have discovered

²⁰ "Drones and Democracy," *The Economist*, October 1, 2010, www.economist.com/babbage/2010/10/01/drones-and-democracy.

²¹ Ronald Arkin, "Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture," *Technical Report GIT-GVU-07-11*, www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf.

²² Walzer, *Just and Unjust Wars*, 251.

²³ Arkin, "Governing Lethal Behavior," 6–8.

many or most of these rules long ago.”²⁴ Programming a robot to discriminate between a combatant and a civilian might be easy enough, using facial recognition, in the case of an individual assassination but remarkably difficult in the general context of a counterinsurgency. The US has used “signature strikes” in Pakistan and Afghanistan, authorizing the use of force against any who fit certain behavioral profiles, such as transporting weapons or congregating as large groups of young men. This has, unfortunately, resulted in the targeting of wedding parties in a part of the world where the shooting of rifles is part of the traditional celebration and gender exclusivity separates male and female wedding participants.²⁵ Discrimination requires a high level of context sensitivity, one that would be complex to program.

Arkin notes that soldiers often violate the principle of right intention, acting out of fear or anger. However, the emotions and consciences of human soldiers also act, at times, as a check on unjustifiable commands or illegal orders. Could a future autonomous weapon have a conscience or true moral agency? Current AIs do exactly what we tell them to do, even when their instructions or the sequence of their learning might be so complex that we cannot anticipate the result. A machine with moral agency would have a further ability to reason independently and unpredictably change course, should it consider the actions it is programmed to take unethical or in violation of an overarching value or intention.²⁶ Michael and Susan Anderson go a step further. They consider a robot or program to be a moral agent if it fits three criteria. First, it is not “under the direct control of any other agent or user.” Second, its interaction with its environment is “seemingly deliberate and calculated.” Third, it fulfills “some social role that carries with it some assumed responsibilities.”²⁷ This third criterion points to the relational nature of what we call conscience. We learn our social responsibilities gradually from parents, peers, our faith traditions, and, for a soldier, from his or her fellow soldiers, commanding officers, and basic training. Most soldiers report that the greatest motivating factor for their actions on the battlefield is their sense of solidarity with and responsibility for their fellow soldiers.

Could such social awareness—indeed, obligation—be instilled in AI? Isaac Asimov envisioned such a need and developed what we now call the Three Laws of Robotics:

²⁴ John Kaag and Whittel Kaufman, “Military Frameworks: Technological Know-how and the Legitimization of Warfare,” *Cambridge Review of International Affairs* 22, no. 4 (2009): 601.

²⁵ For more on US use of drones and signature strikes, see Kenneth R. Himes, *Drones and the Ethics of Targeted Killing* (Lanham, MD: Rowman and Littlefield, 2015).

²⁶ Joseph Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation* (New York: W. H. Freeman, 1976), 74.

²⁷ Michael Anderson and Susan Leigh Anderson, *Machine Ethics* (Cambridge: Cambridge University Press, 2011), 158.

- A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
- A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

While these laws sound good on first encounter, Asimov's two collections of short stories—*I, Robot* and *The Rest of the Robots*—present a thorough exploration of the many situations in which these laws are destined to fail, as they are overly broad, lack contextual adaptation, and can be self-conflicting. The robot is often helpless to act, hardly what any military would want in an autonomous weapon. Instead, we wish to empower such weapons to adapt to changing conditions and make “smart” choices. But on what basis? Human soldiers bring years of social experience that hones awareness of both their role and their responsibilities *vis-à-vis* other humans. As yet we have no idea how to instill this into a machine.

Expecting LAWS to engage in just warfare assumes not only that its precepts are codifiable or learnable by AI, but that they are what would actually be programmed or taught. Moreover, an AI's original programming might be altered or hacked by a bad actor. Should autonomous weapons become “small, smart, cheap, and abundant,” it will be hard to keep them out of the hands of terrorists and non-state actors, who would not have the same interest in following the rules of warfare. Even lacking these scenarios, how likely is a military to prioritize ethics over victory? I fear that a nod to ethical principles could easily degrade into mere “window dressing” for the public while the true goal in programming LAWS would be, and many would argue rightly so, to win at all costs.

Finally, sometimes moral behavior means breaking the rules. Most disputes between Jesus and the Pharisees recounted in the Gospels hinged precisely on Jesus or his disciples breaking a rule or religious convention. The spirit of the law does not always match the letter. Former Army Ranger Paul Scharre recounts a situation in the Iraq war in which the Mahdi Militia used a child as a forward observer. US forces did not shoot the child, even though the conventions of war would have allowed it.²⁸ Would AI be programmed with sufficient nuance to make this judgment call? This is certainly beyond our programming capabilities at this time. Hence, I must disagree with Arkin; for now, it seems unlikely that an autonomous weapon would behave more justly than a human being.

²⁸ Michael Anderson and Susan Leigh Anderson, *Machine Ethics*, 600.

Jus post Bellum: Who Was Responsible?

In *The City of God*, Augustine notes that the goal of every war is a final state of “peaceful order”: “It is an established fact that peace is the desired end of war. For every man is in quest of peace, even in waging war, whereas no one is in quest of war when making peace.”²⁹ While not a part of traditional just war theory, the aftermath of war may be as important in reestablishing order and right relationship between the warring parties as the war itself. One might think LAWS, being weapons, would need no further consideration once the shooting stops. However, one necessary process in the period immediately following a war is the determination and execution of retributive and/or restorative justice. Despite the best planning injustices, atrocities, accidents, and war crimes will occur. When these are not addressed, grievances may fester.

Eventually, an autonomous weapon will be involved in an accident or atrocity that seriously violates international law or Christian ethics. When this occurs, who is responsible? Recall that one of the stipulations for a weapon to be considered autonomous is that its choices and decisions must carry a certain degree of unpredictability. A computer program that is entirely predictable is completely determined by its programmer. While the actions of an autonomous weapon may be foreseeable in most circumstances, they will not always perform as expected. Indeed, systems that depend on machine learning can be “opaque even to the system’s designers.”³⁰ While the designers or programmers carry a certain degree of responsibility for creating such a machine, can they be held responsible for any particular decision?

If not the programmer, can the machine itself be held responsible? Rebecca Crootoof argues that there is no sense in this. To call something a war crime it must have been “willfully” committed. At this point, and in the foreseeable future, we cannot say a machine behaves either intentionally or recklessly. Nor would it make any sense to punish a machine that can feel neither emotional nor physical pain. The machine can certainly be decommissioned, but it will feel no sense of responsibility. According to Crootoof, “Traditional justifications for individual liability in criminal law—deterrence, retribution, restoration, incapacitation, and rehabilitation—do not map well from human beings to robots.”³¹ Ultimately, we expect another human to carry the mantle of responsible agent.

That leaves the one who deploys LAWS responsible. Under current military law, a commanding officer can be held indirectly responsible for the actions of those under his or her command if those actions

²⁹ Augustine, *Concerning the City of God against the Pagans*, trans. Henry Bettenson (London: Penguin, 1984), 866.

³⁰ Crootoof, “War Torts,” 1373.

³¹ Crootoof, “War Torts,” 1377.

could have been in any way foreseen or prevented. A panel at Harvard Law School noted the challenge to this raised by LAWS' inherent unpredictability: "Would fully autonomous weapons be predictable enough to provide commanders with the requisite notice of potential risk? Would liability depend on a particular commander's individual understanding of the complexities of programming and autonomy?"³²

Since none of the other options work, the responsibility to see that LAWS are designed to act justly must therefore remain, for now, with the state that deploys them, as does any responsibility for restorative or retributive justice in the event of a moral breach.³³

Jus ad Vim: Use of Autonomous Weapons Outside of War

LAWS need not be confined to the battlefield. They might also become a weapon of choice against internal enemies. In a 2015 statement, Amnesty International cited their concern regarding the possible deployment of weapons with facial recognition or selection abilities based on other physical traits against minorities or political targets.³⁴ In a special report to the Council on Human Rights of the UN, Christof Heyns similarly raised this concern: "On the domestic front, [LAWS] could be used by States to suppress domestic enemies and to terrorize the population at large, suppress demonstrations and fight 'wars' against drugs."³⁵

On the foreign front, LAWS could and likely will replace unmanned drones as the weapon of choice in combatting terrorism and effecting assassinations. We have already seen how drones have increased the incidence of attacks within the boundaries of states with whom we are not at war, specifically in Pakistan and Yemen in recent years. Use of LAWS presents the danger of automatically escalating hostilities and, more importantly, hardening the hearts and minds of the civilian populations under anonymous surveillance and ongoing threat of attack. The low costs and lack of danger to one's own soldiers presented by LAWS could easily tip calculations toward, rather than away from, escalating a conflict.

Berkeley AI researcher and activist Stuart Russell notes: "I'm not too worried about vast autonomous swarms of battle tanks...there are far cheaper ways to flatten a city and/or kill all of its inhabitants." Instead, Russell fears exactly those weapons that would be used off the battlefield—cheap, small, lethal drones that could be used by police,

³² Crotoft, "War Torts," 1381.

³³ Crotoft, "War Torts," 1390.

³⁴ Amnesty International, "UN: Ban Killer Robots before Their Use in Policing Puts Lives at Risk," April 16, 2015, www.amnesty.org/en/latest/news/2015/04/ban-killer-robots-before-their-use-in-policing-puts-lives-at-risk.

³⁵ Christof Heyns, "Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions," April 9, 2013, www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf.

terrorists, or indeed anyone with a grudge. “My UAV colleagues tell me they could build a weapon that could go into a building, find an individual, and kill them as a class project.”³⁶ This is a frightening scenario, and perhaps the strongest reason many have for an international ban on such weapons.

SHOULD LAWS BE BANNED?

We have only begun to include LAWS in deliberations on arms control. The International Committee of the Red Cross has published advisory guidance on the use of autonomous weapons, but there are no formal international agreements.³⁷ War must be waged by a responsible and legitimate authority, at all levels. Robert Sparrow argues that, since no human can ultimately be held responsible for their actions, LAWS are profoundly and irremediably unethical.³⁸ Several international groups agree. The Campaign to Stop Killer Robots, a coalition of 166 non-governmental organizations in sixty-six countries, has called for a ban on fully autonomous weapons. They maintain that the use of such weapons “crosses a moral threshold” because machines “lack the inherently human characteristics such as compassion that are necessary to make complex ethical choices.”³⁹ They fear that such weapons would lower the barriers against going to war and further shift the burden of warfare onto civilians. Amnesty International has also called for a ban, fearing the human rights implications of such weapons.

Others fear a new arms race. An open letter, signed by such notables as Elon Musk and Stephen Hawking as well as over 4500 other researchers in AI and robotics, published by The Future of Life Institute supports a total ban lest autonomous weapons “become the Kalashnikovs of tomorrow.” They note that because autonomous weapons are not nearly as costly or difficult to produce as nuclear weapons it may “only be a matter of time until they appear on the black market and in the hands of terrorists, dictators wishing to better control their populace, warlords wishing to perpetrate ethnic cleansing, etc. Autonomous weapons are ideal for tasks such as assassinations, destabilizing nations, subduing populations and selectively killing a particular

³⁶ Sydney Freedberg, “Genocide Swarms & Assassin Drones: The Case for Banning Lethal AI,” *Breaking Defense*, March 8, 2019, breakingdefense.com/2019/03/genocide-swarms-assassin-drones-the-case-for-banning-lethal-ai.

³⁷ International Committee of the Red Cross, “Autonomous Weapon Systems: Is It Morally Acceptable for a Machine to Make Life and Death Decisions?,” April 13, 2015, www.icrc.org/en/document/lethal-autonomous-weapons-systems-LAWS.

³⁸ Sparrow, “Killer Robots,” 64.

³⁹ Campaign to Stop Killer Robots, “Statement to the Informal Discussions on Autonomous Weapon Systems,” June 29, 2021, www.stopkillerrobots.org/wp-content/uploads/2021/09/CSKR-Statement-to-the-informal-discussions.docx.pdf.

ethnic group.”⁴⁰ In light of the work of these groups and many others, in 2018 the European Parliament called upon the United Nations General Assembly to “work towards an international ban on weapon systems that lack human control over the use of force” and “to urgently develop and adopt a common position on autonomous weapon systems.”⁴¹

The Vatican agrees, finding it an affront to human dignity to be killed by a machine that cannot make “intentional, rational, and deliberate decisions from a moral and ethical standpoint.”⁴² Speaking as Vatican observer to the UN in Geneva, Archbishop Ivan Jurkovic stated that the development of autonomous weapons would provide “the capacity of altering irreversibly the nature of warfare, becoming more detached from human agency, putting in question the humanity of our societies.”⁴³ According to Jurkovic, each new trend in weaponry “contributes to increasing awareness that the cruelty of conflicts must be done away with in order to resolve tensions by dialogue and negotiation.”⁴⁴

Former lieutenant colonel David Grossman suggests that killing has, in the past, not come easily to human soldiers. He cites a study conducted by the US Army stating that only 15 to 20 percent of soldiers fired their weapons in combat in World War II. Fewer fired to kill.⁴⁵ These percentages rose in subsequent wars; Grossman cites two factors that work together to overcome our resistance to killing one another. The first is dehumanization of the enemy. The more the enemy is seen to be “like us,” the harder it is to kill that person. The second is distance from the target. Hans Morgenthau has suggested that the increasing automation of war overcomes both these factors, bringing us close to “push-button war,” war that is “anonymously fought by people who have never seen their enemy alive or dead and

⁴⁰ “Autonomous Weapons: An Open Letter from AI & Robotics Researchers,” futureoflife.org/open-letter-autonomous-weapons.

⁴¹ “European Parliament Recommendation to the Council on the 73rd session of the United Nations General Assembly (2018/2040[INI]),” www.europarl.europa.eu/doceo/document/A-8-2018-0230_EN.html?redirect#title1.

⁴² Liam McIntyre, “Autonomous Weapons Systems Threaten Peace, Says Vatican Official,” *Crux*, March 29, 2019, cruxnow.com/vatican/2019/03/autonomous-weapons-systems-threaten-peace-says-vatican-official.

⁴³ Catholic News Service, “Vatican Official: Prohibit ‘Killer Robots’ Now before They Become Reality,” *The Tablet*, November 28, 2018, www.the-tablet.co.uk/news/11072/vatican-official-prohibit-killer-robots-now-before-they-become-reality.

⁴⁴ Catholic News Service, “Holy See Renews Appeal to Ban Killer Robots,” November 28, 2018, www.catholicnewsagency.com/news/holy-see-renews-appeal-to-ban-killer-robots-74479.

⁴⁵ S. L. A. Marshall, *Men against Fire: The Problem of Battle Command* (Norman, OK: University of Oklahoma Press, 2000).

who will never know whom they have killed.”⁴⁶ Distance and dehumanization go hand in hand. In a 2013 “Resolution against Drone Warfare,” the Church of the Brethren noted: “Jesus, as the Word incarnate, came to dwell among us (John 1:14) in order to reconcile humanity to God and bring about peace and healing.... We find the efforts of the United States to distance the act of killing from the site of violence to be in direct conflict to [this] witness of Christ Jesus.”⁴⁷

International treaties have banned or limited chemical and nuclear weapons. While such bans have not been honored by all countries and regimes, they have served to keep these weapons out of many national arsenals. While ideal, a total ban on autonomous weapons is unlikely. At the level of the United Nations, it is likely that the US, Russia, and China, each with a vigorous LAWS program, would veto such a ban. The pace of technological development far outstrips that of diplomacy. Lacking a total ban, the rules of just war theory become that much more important. In their recognition that warfare is at times regrettably unavoidable, they provide a system for mitigating war’s reach and effects. Scharre suggests that we use just war traditions to establish “rules of the road” for autonomous weapons, rules that would at minimum serve to reduce civilian casualties, militate against escalation, and promote transparency.⁴⁸

CONCLUSION: THE ROLE OF HUMAN JUDGMENT

We have noted that lethal autonomous weapons could make war too easy, act without morals, muddy the attribution of responsibility, and lead to more acts of violence outside the already established norms and conventions of warfare. While the optimal solution would be a ban, that remains unlikely. Thus, we now need to establish new norms and conventions, ideally based on just war traditions, to ensure these weapons are used in a restrained and responsible manner.

This calls for human judgment, which has been implicit but rarely specified in the rules of warfare.⁴⁹ As we design and build autonomous weapons there are two directions we can take. The first is to continue designing weapons that wage war in our stead. Alternatively, we can design for human-machine symbiosis, leveraging the distinctive strengths of the computer to work together with human beings “to

⁴⁶ Hans Morgenthau, *Politics among Nations* (New York: McGraw-Hill, 2006), 250.

⁴⁷ Church of the Brethren Ministry and Mission Board, “Resolution against Drone Warfare,” March 2013, www.brethren.org/about/statements/2013-resolution-against-drones.pdf.

⁴⁸ Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York: Norton, 2018), 356–57.

⁴⁹ Scharre, *Army of None*, 357.

empower, not replace, those who serve.”⁵⁰ AI should not take over tasks from humans but use differing talents and opportunities to work together to complete those tasks. We must retain a role for human judgment.

Is this, like an international ban on LAWS, also unlikely? While most commanders express a desire for autonomous weapons to have humans in or at least on the loop (having, at minimum, the ability to veto the machine’s decisions), how much control can they actually have if decisions in the field are made at a speed humans are unable to follow? The tempo of war has steadily accelerated, increasing dramatically in recent years.⁵¹ LAWS will only further this acceleration. At what point might things move so quickly that we would be forced to cede all decision making to machines? As AI moves from tactical to strategic decision making, this could eviscerate any meaning from the concept of “mission command.”⁵²

We can still choose not to go down that road. At a recent workshop sponsored by the law faculty at Penn State, I was heartened to hear military commanders, both active and retired, express their personal distaste for LAWS. I join with these commanders in hoping that we choose to never reach the point where LAWS outstrip human commanders’s ability to control them. Whether on the field of battle or in the workplace, human dignity depends on our working with our tools rather than letting them supplant us, and this is most important in matters that involve questions of life and death. After watching the first test of a nuclear bomb in 1945, Harry Truman wrote: “Machines are ahead of morals by some centuries, and when morals catch up perhaps there’ll be no reason for any of it.”⁵³ International cooperation, treaties, and overwhelming abhorrence have kept nuclear weapons from being used since Truman’s time. Let us hope that seventy-five years from now we will have similarly limited the usage of autonomous weapons. It is past time for our morals to catch up with our technology.

M

Noreen Herzfeld is the Nicholas and Bernice Reuter Professor of Science and Religion at St. John’s University and the College of St. Benedict and a

⁵⁰ Rebecca Slayton, “The Promise and Peril of Artificial Intelligence: A Brief History,” *War on the Rocks*, June 8, 2020, warontherocks.com/2020/06/the-promise-and-risks-of-artificial-intelligence-a-brief-history.

⁵¹ T. K. Adams, “Future Warfare and the Decline of Human Decision-Making,” *Parameters: US Army War College Quarterly* 2, no. 57 (2001), 57–71.

⁵² Kenneth Payne, *Strategy, Evolution, and War: From Apes to Artificial Intelligence* (Georgetown: Georgetown University Press, 2018), 183.

⁵³ Quoted in Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Farnham: Ashgate, 2009), 167.

research associate with ZRS Koper. She holds degrees in Computer Science and Mathematics from The Pennsylvania State University and a PhD in Theology from The Graduate Theological Union, Berkeley. Herzfeld is the author of *In Our Image: Artificial Intelligence and the Human Spirit* (Fortress, 2002), *Technology and Religion: Remaining Human in a Co-Created World* (Templeton, 2009), *The Limits of Perfection in Technology, Religion, and Science* (Pandora, 2010) and editor of *Religion and the New Technologies* (MDPI, 2017).