



Applying AI on the Battlefield: The Ethical Debates

Gregory M. Reichberg and Henrik Syse

Contents

Introduction	147
Background Considerations	149
Principled Arguments for and against Battlefield Use of LAWS	151
Technical and Pragmatic Considerations	154
Virtue Ethics and Human–AI Interaction	156
References	158

Abstract

Because lethal autonomous weapon systems (LAWS) are designed to make targeting decisions without the direct intervention of human agents (who are “out of the killing loop”), considerable debate has arisen on whether this mode of autonomous targeting should be deemed morally permissible. Surveying the contours of this debate, the authors first present a prominent ethical argument that has been advanced in favor of LAWS, namely, that AI-directed robotic combatants have an advantage over their human counterparts, insofar as the former operate solely on the basis of rational assessment, while the latter are often swayed by emotions that conduce to poor judgment. Several counter arguments are then presented, inter alia, (1) that emotions have a positive influence on moral judgment and are indispensable to it; (2) that it is a violation of human dignity to be killed by a machine, as opposed to

being killed by a human being; and (3) that the honor of the military profession hinges on maintaining an equality of risk between combatants, an equality that would be removed if one side delegates its fighting to robots. The chapter concludes with a reflection on the moral challenges posed by human-AI teaming in battlefield settings, and how virtue ethics provides a valuable framework for addressing these challenges.

Keywords

AI · Autonomy · Combatants · Emotions · Ethics · Rationality · Robots · Weapons

Introduction

The battlefield is an especially challenging domain for ethical assessment. It involves the infliction of the worst sorts of harm: killing, maiming, destruction of property, and devastation of the natural environment. Decision-making in war is carried out under conditions of urgency and disorder. Clausewitz famously referred to this as the “fog of war”; indeed, the

G. M. Reichberg (✉)
Peace Research Institute Oslo (PRIO), Research School on Peace and Conflict | Political Science, University of Oslo, Grønland, Norway
e-mail: greg.reichberg@prio.org

H. Syse
Peace Research Institute Oslo (PRIO), Oslo, Norway
Björknes University College, Oslo, Norway

very root of our word “war” is derived from the Germanic root *wirr*, which signified “thrown into confusion.” Showing how ethics is realistically applicable in such a setting has long taxed philosophers, lawyers, military practitioners, and educators. The advent of artificial intelligence (AI) has added a new layer of complexity. Hopes have been kindled for smarter targeting on the battlefield, fewer combatants, and hence less bloodshed; simultaneously, stern warnings have been issued on the grave risks of a new arms race in “killer robots,” the loss of control over powerful machinery, and the risks associated with delegating lethal decisions to increasingly complex and autonomous machines.

While war has remained a constant in human existence, the progressive introduction of new technologies (e.g., gunpowder, mechanized infantry, air power, nuclear munitions) has led to dramatic shifts in battlefield dynamics. Warfare has been extended into new domains—air, undersea, cyber, and now outer space—that in turn interact in novel ways. How transformative AI will ultimately be within this multilayered battlefield has been the subject of much speculation, but already military forces the world over, not least the major powers but also many lesser ones, are investing heavily in AI-based weapons systems and platforms.¹ Ethical reflection on the likely implications is imperative. This chapter aims to outline the main directions of current debate in this field. Our focus is on AI-based weapons technology; we will largely leave to the side how AI more broadly supports military activity: monitoring troop movements and capabilities, administration of aid to wounded personnel or their extraction from the battlefield, diffusing explosive munitions, and so forth.

At the outset, it can be noted that AI is not itself a weapon. Rather, it is a cognitive tool that facilitates the application of weaponry to selected targets. AI does this both through the mediation of robots and by assisting human agents in applying the weaponry themselves. In either case, AI mimics² cognitive abilities—sensation, memory, and inference—that are found in human beings. AI is patterned after these abilities, sometimes falling short of them, and other times surpassing them. At the present stage of scientific advancement, general artificial intelligence has not been achieved (and it remains an open question whether it ever will be); for the foreseeable future at least, machine intelligence will remain highly selective in its operations. For this reason, in what follows, we proceed from the assumption that AI is a tool—albeit the most sophisticated tool yet devised by human beings—and

even when implemented in robots, it does not possess agency in the proper sense of the term (which entails a capacity for self-awareness, ability to set goals, and so forth). This is not to say that AI *qua* tool cannot run in autonomous mode, making its own decisions and learning from previous decisions. On the contrary, this is already possible. However, even when operating in autonomous mode, AI serves in a support capacity; full agency, selfhood, and personhood cannot be attributed to it. Responsibility for any wrongdoing that might result from AI-powered operations must be traced back to the human agents who propose, design, operate, or direct their use.

AI is a tool that extends human cognitive abilities beyond their normal range of functioning. AI can enhance human sensory capacities, as when it is used for purposes of surveillance or detection; AI can increase the speed with which humans process information; and AI can contribute to human decision-making, either by providing input that supports decisions made by humans or, as with autonomously functioning AI, the decision itself is delegated to the machine. Decision is the cognitive act whereby an antecedent phase of deliberation (whether extended or instantaneous) issues into a course of action. A decision is a special form of judgment: “*x* shall be done.” The doing of *x* changes some aspect of the world. This is what philosophers call “practical” (as opposed to “speculative” or “theoretical”) judgment. In what follows, we are concerned with practical judgments within the sphere of military action, particularly those decisions that result in harm done to human beings or the natural environment. The chapter proceeds as follows:

1. To provide context for our discussion, we review (i) the principal reasons that have induced military planners to develop AI-based warfighting capabilities, (ii) how autonomy in AI-based weapons systems is a matter of degree, and (iii) current attempts to code ethical norms into autonomous AI systems for military applications.
2. Thereafter, we review ethical arguments for and against the use of autonomously functioning AI battlefield targeting systems, focusing first on the more principle-based arguments, and thereafter on arguments belonging more in the domain of the technological and pragmatic, although, admittedly, there is not a clear-cut distinction between the two categories, and some arguments will overlap.
3. By way of conclusion, we look at how AI–human collaboration (“the force mix”) on the battlefield can be expected to impact on the practical judgment of military personnel, their ability to engage in ethical (“virtuous”) conduct, and their moral integrity. Using the tradition of virtue ethics as our point of departure we formulate a number of questions for further research.

¹For a survey of the current technology see Swett et al. in this volume.

²“Mimic” is not here taken in a pejorative or dismissive sense, as when we say that a parrot “mimics” human speech. It is taken rather in the sense of Platonic “imitation,” as when it is said that one reality can possess (“participates in”) what another has but in a diminished or analogous manner. See Patterson (1985).

Background Considerations

In promoting the development of AI-based warfighting capabilities, military planners have responded to several different needs and technological developments.

First, there is the robotics revolution, which has led to an increasing deployment of remote-piloted unmanned ground, surface, underwater, and aerial vehicles. Best known of these are the “drones” (unmanned aerial vehicles—UAVs), which have been extensively used to deliver lethal attacks most notably in Afghanistan, but elsewhere as well. The remotely controlled deployment of these vehicles by human pilots (often sitting thousands of kilometers away from the theater of operations) presents as threefold difficulty: such deployment (1) is very labor intensive (one or more operators are needed to control a single vehicle), (2) requires communication links that are subject to adversarial disruption or are inoperative in some locations, and (3) functions relatively slowly given its dependency on human cognitive reflexes and decision-making. AI-directed unmanned vehicles provide a way around these three difficulties, freeing up human operators for other tasks, obviating the need for constant communications links, and allowing for a more rapid response time. The last feature has become especially important in the context of “swarm warfare,” whereby multiple vehicles proceed against a single target (or against another swarm), in a high-speed, tightly coordinated attack. Speed of response is also highly beneficial in related settings, for instance, in cyber confrontations that occur in the milliseconds, or radar-based defensive action to protect against incoming missiles.

It goes without saying that the use of unmanned attack vehicles has the added advantage of protecting military personnel from lethal harm; AI-directed attacks decrease the number of personnel that need be placed on the battlefield, thereby preserving them from injury and death. After World War I force protection has become a paramount concern for conventional armies, and the US experience in Viet Nam showed how soldierly casualties can have a very adverse political impact, even on an otherwise dominant military force.

Replacing human agents in combat settings, in the ways summed up above, is possible only when AI enables weapon systems to operate in autonomous mode. For purposes of this discussion, artificial intelligence may be defined as *intelligent behavior embedded in artificial matter*. “Intelligent” designates an ability to solve complex problems to achieve some goal, while “artificial” excludes biological systems—most importantly: living, breathing, thinking human beings. This definition covers both autonomous and non-autonomous systems. An artificially intelligent system is autonomous if the selection of the means for reaching a preset goal is left to the system itself, as in what has become known as “machine learning.” Here the machine “has flexibility in how

it achieves its goal” (Scharre 2018: 31). By contrast, a system is *non*-autonomous if the means for reaching a preset goal are predetermined by an external agent, as in the case of a cruise missile that follows a given program, however complex the program might be. It goes without saying that autonomy is very much a matter of degree. There is a spectrum of intelligence in lethal machines (Scharre 2018: 31–34), from systems that are automatic (simple, threshold based), for instance an improvised explosive device, to those that are automated (complex, rule-based), for instance a precision-guided munition, and finally those that are autonomous (goal-oriented, self-directed with respect to the selection of means), for instance the Israeli manufactured Harpy that destroys specific radar installations, none of which are determined in advance, within a particular radius.

Judgments made about machine “autonomy” are very much in the mind of the beholder. Machines whose inner workings we do not understand will often seem to be wholly unpredictable and to produce effects that follow from a decision made by the machine itself, thus the ascription of autonomy to the machine. But once we understand the complex logic on which a machine operates, its resulting actions will usually be redescribed as being merely automated (Scharre 2018: 26 ff.). Of course, this begs the question of whether any machine could ever be autonomous in the proper metaphysical sense of possessing free will. Being free in this sense entails the ability to dominate the reasons for one’s action; no one reason requires me to do this or that, i.e., necessitates my action (Simon 1969). Whether freedom from the necessitation of reasons can be achieved by a machine is in all probability impossible. The machine is bound by its underlying architecture and that is why our initial judgment of a machine’s autonomy eventually gives way to more moderate characterization in terms of automaticity. In other words, here as elsewhere we need to be on guard against the anthropomorphic imagination.

With respect to lethal weaponry, autonomous functioning is usually described in terms of a threefold distinction (see Scharre 2018: 29–30) between modes of human presence in the “killing loop.” First, (1) there is semiautonomous machine killing. Such a system can detect the external environment, identify hostile targets, and even propose a course of action, but the kill decision can only happen through the intervention of a human being. Here a human operator remains *in the killing loop*, in the sense that he/she must take positive action if the lethal attack is to be consummated. Then, (2) there is supervised autonomous machine killing. Such a machine can sense, decide, and act on its own, but it remains under the supervision of a human being who can veto the passage from decision to action. Should no veto be issued, the machine is fully capable of running through the combat cycle (observe, orient, decide, act) on its own. Here a human being remains not *in*, but *on the killing loop*. Third, (3)

there is fully autonomous machine killing whereby a human being is needed only to activate the machine, but afterwards it carries out its assigned task without communication back to the human user. Here the human being is *out of the killing loop*. This threefold distinction of *in*, *on*, and *out of* the loop refers to three modes of operation but not necessarily three kinds of machine, as one and the same machine could be set to run on each of these three modes.

Finally, with respect to AI-driven weapon systems that operate in autonomous mode (humans out of the loop), current research has sought to devise algorithms that embed ethical principles into the targeting decisions adopted by these machines. The issue here is to consider whether and how autonomous robots can be programmed to function legally and ethically as battlefield combatants. As already noted, robots can have other tasks on the battlefield, such as retrieving injured soldiers (the Battlefield Extraction Assist Robot), or monitoring human battlefield conduct so that norm violations will be reported back to central command or an international agency, and perhaps even prevented (fear of punishment arising from robotic observations might dissuade soldiers from acting wrongly in the first place). Our interest in this chapter is, however, with weaponized robots, usually termed LAWS (Lethal Autonomous Weapons Systems); AWS (Autonomous Weapons Systems) is used by some as an alternative.

The question here is whether “the rules governing acceptable conduct of personnel might perhaps be adapted for robots” (Lin et al. 2008: 25). Is it possible “to design a robot which has an explicit internal representation of the rules and strictly follows them?” (ibid.) Attempts at answering these questions have focused on the following considerations.

As a point of departure, we have to distinguish between operational and functional morality. An operational morality is one in which all possible options are known in advance and the appropriate responses are preprogrammed. “The actions of such a robot are entirely in the hands of the designers of the systems and those who choose to deploy them” (ibid. 26). Such robots have no ability to evaluate their operations and correct errors. An operational morality has the advantage of being entirely derivatory on the decisions made by the designer/user, so the lines of control, and hence responsibility, are crystal clear. However, apart from very narrow operating environments, it is impossible to preconceive all possible options in advance, because of the complexity of the environments in which the robots will be deployed, because the systems are introduced in settings for which they were not planned, or even because of the complexity of the technology involved such that “the engineers are unable to predict how the robot will behave under a new set of inputs” (ibid.). As the battlefield is a notoriously disorderly environment, for deployment in real-life battlefield settings, LAWS must be

programmed with a *functional morality*, namely a built-in capacity to evaluate and respond to moral/legal considerations.

Following on from this, the design of functional morality in robots—namely a capacity for moral reasoning so that unanticipated situations can be dealt with appropriately—has been approached in two different ways, *top-down* and *bottom-up*:

- (a) In a top-down approach, a particular moral theory is encoded in the software. This will typically involve some version of deontology or consequentialism, which are then detailed into a set of rules that can be turned into an algorithm. There are many challenges here, for instance the possibility of conflict between rules. This could lead to paralysis if the rules are meant to function as hard restraints, or if the rules are designed only as guidelines, this could open the door to robotic behavior that should be prohibited. A further and especially pressing issue concerns what is termed the “frame-problem” (Dennett 1984; Klinecicz 2015), namely to grasp the relevant features of a situation so the correct rules are applied. To borrow an example: How would a robot programmed with Asimov’s “First Law [of robotics, which says that a robot may not injure a human being or, through inaction, allow a human being to come to harm] *know*, for example, that a medic or surgeon welding a knife over a fallen fighter on the battlefield is not about to harm the soldier?” (Lin et al. 2008: 31). In a famous case, a Soviet colonel saw his computer screen flash “launch,” warning him that the US had initiated a nuclear attack. Thinking there was a bug in the system he waited, and it happened again repeatedly; finally, “missile strike” replaced “launch” and the system reported its highest confidence level. Still the colonel paused. Having seconds left to decide the matter, he called the ground-based operators for confirmation, but they had detected nothing. It turns out the system had malfunctioned; it had mistaken light reflecting off a cloud configuration for the trace of an incoming missile. The frame-problem, if unresolved, can have enormous consequences if a powerful weaponized AI system is in play (Scharre 2018: 1–29).
- (b) In a *bottom-up* approach to functional machine morality, systems mimicking evolutionary or developmental processes are implemented within machine learning. The basic idea here is that “normative values are implicit in the activity of agents rather than explicitly articulated . . . in terms of a general theory” (Lin et al. 2008: 35). This has led to the application of virtue ethics to autonomously functioning machines. Just as people are taught to acquire the right set of character traits (virtues) and on that basis come progressively to understand what morality requires (it has been suggested by Kolberg and others that this is how children learn about morality), likewise

neural networks might provide a pathway toward the engineering of robots that “embody the right tendencies in their reactions to the world” (Lin et al. 2008: 40). This “bottom-up development of virtuous patterns of behavior might be combined [the hybrid approach] together with a top-down implementation of the virtues as a way of both evaluating the [resulting] actions and as a vehicle for providing rational explanations of the behavior.” In this way, “a virtuous robot might emulate the kind of character that the armed forces value in their personnel” (ibid.). Even if feasible, development of such technology appears to be still well off in the future.

Principled Arguments for and against Battlefield Use of LAWS

Because LAWS are designed to make lethal targeting decisions without the direct intervention of human agents (who are “out of the killing loop”), considerable debate has arisen on whether this mode of autonomous targeting should be deemed morally permissible. A variety of arguments have been proposed that we classify below into four different main types. Alongside this ethical discussion, calls for the international legal regulation of LAWS, including a ban on their use, have multiplied (see, e.g., the campaigns and web sites of Human Rights Watch and the Campaign to Ban Killer Robots).³ As there is not a perfect overlap between ethics and law—the latter proceeds from principles and a methodology quite different from the former—the legal issues surrounding LAWS fall outside the scope of the present chapter and will be considered only indirectly.

A principal line of argumentation in favor of LAWS has focused on the qualities that combatants should possess in order to make sound targeting decisions in the heat of battle. Proponents of LAWS have maintained that AI-directed robotic combatants would have an advantage over their human counterparts, insofar as the former would operate solely on the basis of rational assessment, while the latter are often swayed by emotions that conduce to poor judgment (Arkin 2010). The negative role of the emotions is amplified in battlefield settings, when fear, rage, hatred, and related passions often deflect human agents from the right course of action. Machines operating under instructions from AI software would not be prone to emotive distortion; thus, if properly programmed, they could be counted on to function in strict conformity with recognized laws of armed conflict (domestic and international) and the appropriate rules of engagement. The occurrence of wartime atrocities would be reduced if combat missions could be undertaken by au-

tonomously functioning robots. Not only would robotic fighters avoid killing civilians (the *sine qua none* of international humanitarian law), in addition, they could be programmed to assume risk upon themselves to protect civilians from side-effect harm, something that human combatants often shy away from. Finally, robots would be capable of precise targeting, thereby enabling them to disable rather than kill the opposing human combatants. Although human soldiers sometimes have the intent to engage in disabling action, the stress and rapidity of the battlefield, as well as the lack of needed weapons, often results in higher kill rates than might otherwise be the case. Human soldiers do not always have the luxury of precise aiming—say, at the feet rather than the torso—and end up striking sensitive organs that seriously wound or kill their adversaries, despite a wish to cause only minor damage. The same could be said of damage to property and sites of cultural or environmental significance, which are often used to provide cover or expose adversaries to live fire. Robots could be equipped with a sophisticated range of weapons (being strong enough to carry them all), enabling them to select for each situation the weapon best suited to the task at hand; much as expert golfers select from among their clubs.⁴

Against this endorsement of LAWS, counterarguments have been advanced, some principled and others more pragmatic and contingent upon the current state of technological development. We will first be treating the principled objections, which are oriented around four considerations.

First, given that practical decisions take place in circumstances that are inherently affected by contingency, no prior process of reasoning is adequate to render the morally right conclusion, namely a sound decision about what should be done here and now, and in what precise way. Beginning with Socrates’s claim that virtue reduces to knowledge, it has long been an aspiration of some philosophers (epitomized by August Comte), and more lately of social scientists, to devise a science of action that is both predictive and unfailingly right. Such knowledge might be deflected in us by wayward passion, but kept on its own trajectory, it will be unfailingly right. The invention of machine learning with algorithms that program ethical behavior can be viewed as the most recent permutation of the philosophical project to reduce virtue to knowledge.

But against this aspiration often associated with Socrates, other philosophers, beginning with Aristotle, have maintained that knowledge alone, no matter how sophisticated and reliable it may be, can ever serve as a proximate guide to morally right human action. This holds doubly true for action undertaken under the chaotic conditions of the battlefield. Even more challenging than predicting the weather (notorious for its difficulty, which even supercomputers cannot

³See <https://www.hrw.org/topic/arms/killer-robots> and <https://www.stopkillerrobots.org/>

⁴See Kalshoven (1984).

fully overcome), on the battlefield soldiers must confront contingencies relating not only to the terrain, local buildings and other installations (which may or may not be damaged), the weather, and their own and their adversaries' weapon systems (which even when known do not always function as planned), but most significantly, they face off against other agents who are possessed of free will, and who can choose among alternative courses of action, both tactically and strategically. How another individual (or group of individuals) will react can never with certitude be known in advance. It is for this reason that Aristotle emphasized how practical reasoning will succeed (i.e., will conduce to morally good choices) only if directed by an upright will. By the affective orientation of an upright will, the agent's attention is directed toward the most morally salient features of the immediate environment and decides which of these, amidst complexity and rapid change, to prioritize in the resulting choice. Thus, the affective disposition of will, oriented to the moral good, substitutes for the lack of perfect knowledge, impossible under these circumstances, thereby enabling right action under conditions of inherent contingency. This ability, at once cognitive and affective, Aristotle termed *phronesis* (*prudentia* in Latin). Through a combination of intellectual skill (enabling correct apprehension of moral principles) and well-ordered desire (that the later Latin tradition would term *voluntas* or "will" in English) the morally prudent person is able to judge well within the singular contingencies of experience. Intellect supplies an abstract grasp of the relevant moral truths, e.g., "noncombatants should never be intentionally harmed," while the will, which desires things in their very concreteness,⁵ directs the intellect toward specific items in the perceptual field ("I do not want to harm this person"). The more the will is well ordered to the moral good, the better it will orient the intellect in its judgment of things in their particularity. Thomas Aquinas later explained how, in dealing with the challenges of warfare, a special mode of *phronesis* is requisite (see Reichberg 2017).⁶ This he termed "military prudence" (*prudentia militaris*). AI-based machines, no matter how sophisticated their cognitive mimicking may be, will never be possessed of the affective disposition we term "will."⁷ LAWS, by consequence, can never attain to *phronesis*, and, for this reason, cannot function as a trustworthy substitute for human combatants (see Lambert 2019).

⁵Thomas Aquinas emphasized how knowledge is oriented to things under the condition of abstractness and universality, while desire is oriented to things in their concrete singularity, and hence their very contingency. The respective faculties, intellect and will, operate in synergy, and thus each provides input to the other in carrying out its respective operation.

⁶On the interrelation of intellect and will in the cognition of things in their particularity and contingency, see Reichberg (2002).

⁷This claim is of course not self-evident; providing the argumentation would take us too far afield. Our aim at present is to describe the contours of one principled argument against LAWS.

A *second* principled argument against LAWS is orientated around the indispensability of emotions for the exercise of right judgment on the battlefield. As noted already, proponents of LAWS have assumed that practical judgment functions best when freed from the distortions arising from emotion. This is a viewpoint that was originally articulated by the ancient Stoics, who held that as emotion invariably leads humans astray, only a life lived under the austere dictates of reason will be a morally successful one. From this perspective, AI, which operates with cognitive capacity only, can be counted on to adhere unfailingly to the ethical guidelines that have been programmed into it, such that all its subsequent learning will develop in conformity with these guidelines.⁸ Since it operates without emotion, AI has the potential to exceed human beings in making reliable ethical decisions on the battlefield.

By contrast, opponents of LAWS refuse to accept the fundamental premise on which this argumentation is built, namely that emotions are a hindrance to sound judgment and the action that follows from it (see Johnson and Axinn 2013; Morkevicius 2014). On the contrary, they maintain that emotions provide indispensable support for moral agency, and that without emotion, judgment about what should (or should not) be done on the battlefield will be greatly impoverished. Reason, without proper emotional support, will lead combatants astray. From this perspective, emotion can provide a corrective to practical judgments that derive from erroneous abstract principles ("ideology" we would say today) or from blind obedience to authority. There are numerous accounts of soldiers who, acting for instance under the impulse of mercy, desist from wrongful commands or provide aid to enemy combatants whose demise would make no contribution to the military effort (Makos 2014). Human beings have a unique ability to perceive when preordained rules or established plans should be set aside and exceptions made, in the interests of humanity. Our emotions ("feelings") are often what prompt us to see how a given situation requires, "calls for," a response out of our ordinary patterns of behavior. An emotionless machine, operating on sequential reasoning alone, would have no basis to depart from its preprogrammed course of action, and thus no ability for making exceptions. While this could be reckoned only a shortcoming under ordinary operating conditions, should a LAWS be reprogrammed for malicious ends (say, by cyber intrusion or other means), and oriented toward the commission of atrocities, there would be no internal mechanism by which it would resist the new operating plan. No emotional response would provide the necessary cognitive dissonance. By contrast, human beings always have the ability, and sometimes even an obligation, to disobey orders, in other words, to act against instructions

⁸This being the case, it is not surprising that stoicism is the philosophy of choice in Silicon valley (see Manthorpe 2017).

received from a commanding officer. But it is hard if not impossible to imagine how a machine could override its software (Johnson and Axinn 2013: 135).

A *third* principled argument against LAWS proceeds from the moral intuition that battlefield killing will be compatible with human dignity only when it is carried out by the direct decision of a human being. To be killed by machine decision would debase warfare into mere slaughter, as though the enemy combatant were on a par with an animal killed on an automated conveyer belt, or, as two authors put the point:

A mouse can be caught in a mousetrap, but a human must be treated with more dignity. . . . A robot is in a way like a high-tech mousetrap; it is not a soldier with concerns about human dignity or military honor. Therefore, a human should not be killed by a machine as it would be a violation of our inherent dignity (Johnson and Axinn 2013: 134).

The operative supposition is that the killing which is done on the battlefield is special (compared to other modes of killing) insofar as it is done according to an established code of honor, a convention, whereby soldiers face off as moral equals. Each is a conscious participant in a profession which requires that a certain deference be shown to the other even in the process of killing him. Shared adherence to the same calling sets military warfare apart from other sorts of confrontations, say, between police and thieves, where there is no moral reciprocity. In a famous essay (Nagel 1979), philosopher Thomas Nagel maintained that the hostility which is characteristic of war is founded, paradoxically, on a mode of interpersonal relations. Respect for the other (hence treating him as an end and not merely as a means) is maintained even in the process of killing him, when he is targeted precisely as a subject, as someone who is aware that he is the target of hostile acts for a specific reason, namely because he too is directing lethal harm against the one who is trying to kill him. What is often called the “war convention” is based on this reciprocity, namely the mutual exposure to intentional harm, and for this reason the personal dignity of soldiers is maintained even in the killing. But this personal dimension, having lethal harm directed against oneself insofar as one is a member of a distinct class, that of arms-bearing soldier, could not be maintained in the event that the opposing soldier were not himself a person, but only a machine.

Against this view, one could say that it proceeds from a conception of warfare—associated with “chivalry”—that is no longer operative today.⁹ Today’s conflicts are often waged by reference to a justice motive wherein the adversary is deemed immoral, a terrorist or a kind of criminal, such that moral equality with him would be unthinkable. Moreover, if one were to exclude (as morally dubious) “impersonal killing” of the sort carried out by LAWS, then by the same

token much of the technology employed in modern warfare would have to be excluded as well: high altitude bombing of enemy positions, roadside bombs, booby traps, and similar devices. But thus far, few if any are militating for a ban on these methods of warfare, except in cases where civilians might indiscriminately be harmed (land mines or biological weapons) or where the harm to combatants (by, e.g., poisonous gas or chemical weapons) results in long-lasting suffering.

In arguing for an essential difference between human and machine killing, namely that even in a situation where a combatant would justifiably be killed by his human counterpart, it would be wrong in the identical situation to have him killed by a LAWS, Johnson and Axinn (2013) nonetheless draw a distinction between offensive and defensive LAWS. The former would include unmanned ground, surface, underwater, or airborne *vehicles* that are able to attack targets—wherever they may be found—based on preset autonomous decision procedures. Insofar as they are directed against *manned* targets, autonomous killing systems of this sort should never be used, for the reason given above, namely that machine killing is incompatible with human dignity. By contrast, defensive LAWS do not fall under the same moral stricture. “Defensive” would include integrated air defense systems that shoot down aircraft or missiles flying within a specific radius, as well as land-based autonomous turrets or perimeter patrol robots that fire on anyone entering a designated perimeter. Autonomously functioning machines would have moral license to kill anyone entering the said zones, provided these no-go areas were well announced in advance. It could then be assumed that trespassers were engaged in hostile activity, and there could be ample justification to program a machine to kill them upon entry. This would be an AI-based extension of the electric fence.

While the distinction here drawn between the two types of LAWS (offensive and defensive) is useful, it does seem to undermine the authors’ argument that it is an inherent violation of human dignity to be killed by a machine. After all, the basic supposition behind the deployment of LAWS is that such machines can effectively be programmed to distinguish combatants from noncombatants, the former being engaged in hostile activity and the latter not. But if advance warning is what enables a differentiation of allowable versus wrongful machine killing, then anytime it is publicly known that a war is underway, there could be moral justification in having a LAWS kill adversarial combatants. After all, by a tacit convention, combatants know that once they step out on the battlefield they are “fair game”; this is the “perimeter” they have entered, and in doing so they assume risk upon themselves. On this reasoning, all may rightly be made lethal targets of a machine.

A *fourth* principled argument against LAWS is focused, as was the previous argument, on the moral equality of

⁹On the applicability of chivalric norms to contemporary conflict, in light of recent technological developments, see Olsthoorn (2019).

combatants as a prerequisite for maintaining a rule-based order on the battlefield. An expression coined by political theorist Michael Walzer, but with antecedents in international law, the “moral equality of combatants” refers to the idea that the moral standing of combatants can be determined without reference to the cause, just or unjust, for which they fight (Walzer 1992). All soldiers, whether fighting in just or unjust wars, are bound by the same set of rules. On this conception, “when conditions of enmity arise between states, this does not automatically amount to interpersonal enmity between the individual members of the opposing states. In war, those who do the fighting may consequently do so without special animosity toward their adversaries on the other side, because, like themselves, they are mere instruments of the state. This positions them to confront each other as peers in a rule-bound, albeit bloody competition” (Reichberg 2017: 231–232). Because the actual fighting is conducted in detachment from substantive justice (the question of which side is in the right with respect to the *casus belli*), combatants deploy force against each other in view, not of hatred or vengeance, or even high-minded goals such as the upholding of justice, but for the preservation of personal security. Moral license to kill derives, in other words, not from the personal moral guilt of the opposing combatant (after all he is acting out of obedience to his leadership as I am to mine), but from a right to self-defense. “Each possesses this license [to kill in war] because each acts in self-defense *vis-à-vis* the other. The reciprocal imposition of risk creates the space that allows injury to the morally innocent [i.e., combatants on the opposing side]” (Kahn 2002: 2). Rule-based warfare, and the moral equality of combatants that it entails, depends on a mutual assumption of risk by combatants. This reciprocal exposure to serious injury and death is what justifies each in directing self-defensive lethal harm against the other. But should one side prosecute war without assuming any risk upon itself, i.e., its own combatants, its moral right to use lethal force is thereby removed. There can be no moral justification in fighting riskless war. This is exactly the situation that would arise by the introduction of LAWS on the battlefield. The side deploying these “killer robots” against human combatants on the other side might prevail militarily, but the resulting victory would be morally pyrrhic and hence wholly without honor. The professional ethos of soldiering, which rests on a voluntary and reciprocal assumption of risk, would be undermined, and with it the expectation, built up over many centuries, that war can be conducted in a rule-based and even virtuous manner, namely in a way that preserves (and enhances) the moral integrity of those who actively take part in it (Riza 2013).

One could of course respond (Arkin 2010) that the ultimate goal behind LAWS is to reconfigure the battlefield so that in the future robots will fight only robots, not men,

thereby removing the asymmetry outlined above. This, however, is unlikely to produce the desired outcome—bloodless war. Unless belligerents agree to a convention whereby defeat of one’s robotic army will entail capitulation to the political demands of the victor, the war will simply shift to another plane, one in which human combatants are again pitted against the robotic combatants of the (tactically but not strategically victorious) other side, with a morally dubious risk asymmetry reintroduced onto the battlefield.

Another line of response would question whether the moral equality of combatants, and the mutual assumption of risk that underlies it (Renic 2018), is indeed a needed precondition for the maintenance of rule-based warfare. A lively debate has been ongoing on this topic over the last decade (Syse 2015; Barry and Christie 2018). This is not the place to elucidate the details. For our present purpose it can be said that the alternative viewpoint—which posits a moral *inequality* in favor of the combatants who prosecute the just cause—will entail that the just side has moral warrant to engage in risk-free warfare. Or put differently, if LAWS effectively enable force protection, while simultaneously aiming their fire only at enemy combatants and not civilians, no sound moral argument stands in the way of their use. This moral justification derives wholly from the *ad bellum* cause and from nothing else.

One may, however, make much the same argument without going as far as nullifying *in bello* considerations in favor of making *ad bellum* concerns alone morally decisive. One may more simply, with James Cook, argue that we should avoid romanticizing risk and death in war when we can clearly, with the aid of unmanned and AI-based technology, protect our own warfighters better by offering them the opportunity of better defenses, lower risk, and more accuracy. The latter argument can be made even if we do not reject the moral equality of combatants (Cook 2014).

Technical and Pragmatic Considerations

We choose in the following to treat separately a group of ethical arguments for and against the battlefield use of AI that can broadly be termed “pragmatic,” centering partly on the current state of technologies and partly on broader considerations of whether AI will do more harm and good. There is indeed overlap between these considerations and the ones we have called “principled” arguments, and some of the arguments below have been foreshadowed above. Nonetheless, the distinction is useful, since the arguments below do center more on the technical aspects and the consequences of the use of AI technology and take a less principled stand for or against the application of such tech-

nologies. These arguments can be classified by reference to what one might term AI battlefield optimists and pessimists (Syse 2016).

By optimists we mean those who see the introduction of autonomous weapons as representing a net gain in terms of much higher precision, less suffering, and fewer fatalities in war (Arkin 2010; Strawser 2010). Beyond the obvious point, outlined above, that robots fighting each other would spare human combatants from having to do the same, thus resulting in less loss of life, the optimists also claim that robots, even when fighting human beings, would show greater battlefield probity, because they would not be misled by emotion. Wartime atrocities would be eliminated if robots were left to do our fighting for us, provided of course that they were programmed properly, to the extent possible, to avoid harm to noncombatants.

Moreover, robotic fighters would have an additional benefit insofar as they could be counted on to assume risk upon themselves to protect noncombatants, something human combatants often avoid. Attacks could, for instance, be carried out at closer range, thus with greater precision, resulting in decreased rates of side-effect damage to civilians and the infrastructure on which they depend. Moreover, given that AI can provide better situational awareness to human soldiers, targeting decisions will prove to be less lethal, as enemy combatants can more readily be incapacitated than killed. In other words, proportionality calculations could be implemented with enhanced accuracy.

Optimists are also quick to acknowledge the economic as well as tactical advantages of autonomous lethal systems (already mentioned above in the section on Background Considerations). For instance, whereas remotely controlled and supervised battlefield robots require much human capital for their operation, these high costs could be bypassed by means of autonomously functioning robots.

There are tactical benefits also, insofar as autonomous robots eliminate the need for electromagnetic communications links, which are hotly contested in wartime, and inoperative in some settings, for instance, deep undersea. Moreover, much of current military planning increasingly makes use of swarm warfare and the resulting maneuvers happen far too rapidly to be directed by human guidance.

Concluding from these lines of reasoning, B. J. Strawser (2010) holds that the use and further development of unmanned (or “uninhabited”) and increasingly autonomous weapons may be a duty, if the likelihood of fewer casualties and less suffering is significant. Opposing or delaying such development and use would be akin to holding back on the use of life-saving tactics and strategies, even when we know that they will be effective.

In short, the “optimist” arguments hold that the likely overall result of AI technology on the battlefield will be one of more accuracy and fewer human beings put in harm’s way.

Pessimists, by contrast, have offered a set of opposing arguments:

- The anticipation of decreased battlefield human casualty rates (through the introduction of robots) would lower the perceived risks of waging war. In anticipation of fewer battlefield casualties to the deploying side, political leaders who possess a strong LAWS capability will increasingly view initiation of war as a viable policy option. The number of wars will grow accordingly (Asaro 2007).
- It is an illusion to think that robotic warfare will render wars entirely bloodless. Ultimately, the fruit of defeat on the battlefield will be the vulnerability of one’s civilian population to lethal robotic attack, which, given the new technologies developed (e.g., swarmed drone attacks), could lead to massive deaths on a par with nuclear detonations. In this connection, Russell (2019: 112) refers to these AI-based technologies as “scalable weapons of mass destruction.” A new arms race will emerge, with the most unscrupulous actors prevailing over those who show restraint. For instance, AI engineers at leading US technology firms are refusing to engage in military design projects with lethal applications. It is said that at Chinese defense contracting firms, where the development of AI systems is a priority, the engineers have not expressed the same reservations (Knight 2019). An additional worry, recently voiced by Paul Scharre, is that an arms race in AI military applications will lead to a widespread neglect of safety considerations. “[T]he perception of a race will prompt everyone to deploy unsafe AI systems. In their desire to win, countries risk endangering themselves just as much as their opponents” (Scharre 2019: 135).
- The differentiation between combatants and noncombatants depends on a complex set of variables that, in today’s asymmetric battlefields, cannot be reduced to a question of the uniform one may or may not be wearing. Irregular combatants often pose as civilians, and subtle judgments of context are needed to ferret them out from their innocent counterparts. For instance, human beings are adept at perceiving whether their fellows are animated by anger or fear, but machine intelligence in its current form is largely unable to detect this crucial difference.
- Similar problems arise from AI black-boxing, namely the difficulty knowing in advance how an algorithm would dictate a response in an unanticipated set of circumstances (see Danzig 2018 for a survey of the relevant risk factors). Given such immense complexities, teams of

programmers need to collaborate on algorithm design for any one project. Consequently, no one programmer has a comprehensive understanding of the millions of lines of code required for each system, with the result that it is difficult if not impossible to predict the effect of a given command with any certainty, “since portions of large programs may interact in unexpected, untested ways” (Lin et al. 2008: 8). Opening lethal decision-making to such uncertainty is to assume an unacceptable level of moral risk. This unpredictability, amplified by machine learning, could result in mistakes of epic proportions, including large-scale friendly-fire incidents, thereby nullifying the benefits that might otherwise accrue from the use of LAWS. “In the wrong situation, AI systems can go from supersmart to superdumb in an instant” (Scharre 2019: 140). Given this unreliability, military personnel would be unwilling to put their trust in AI systems (see Roff and Danks 2018). The use of such systems would accordingly be avoided, thereby nullifying the tactical benefits that might otherwise accrue. Perhaps a way will be found to overcome algorithmic black-boxing, but at the current stage of AI design, a solution is still well off in the future.

- Likewise, no matter how effectively LAWS might be programmed to act in accordance with ethical norms, cyber intrusion cannot definitively be excluded, such that its code would henceforth dictate unethical behavior, including the commission of atrocities. Advances in cryptology and other defenses against cyber intrusion have still not reached the point where malicious interference can be ruled out.
- Moreover, even if the use of autonomous battlefield robots could, if programmed effectively with moral norms, lead to reduced bloodshed in war, there is no guarantee that all relevant militaries would program their robots in this way. The opposite could easily happen under a variety of scenarios, including states that might refuse to sign onto AI-related treaties that may eventually be negotiated, the assumption of control over such systems by rouge actors or third-party hackers, or the theft, reuse, and reprogramming of battlefield robots.

As this brief summary of core technological and pragmatic arguments shows us, whether the use of complex AI capacities in battlefield weaponry will lead to more or less suffering, more or less casualties, is subject to intense debate. Hence, the uncertainties of the accompanying calculus of moral utility are far-reaching. The “optimists” will, however, insist that their arguments are not meant to be *ipso facto* true: rather, they are unequivocally dependent on the development of AI technologies that discriminate clearly and assuredly. Much of the “pessimist” argument, on the other hand, centers on the unlikelihood that we will—at least in the foreseeable

future—be able to trust, or truly and safely harness, the powers of such technologies.

Virtue Ethics and Human–AI Interaction

Thus far we have mainly considered the question of whether autonomous robots should be allowed on the battlefield. The resulting debate should not blind us to a wider set of questions that are important to address as human–AI interactions—semiautonomous and autonomous—become increasingly prevalent in military planning and execution. How these tools impact on the military personnel who make use of them and their ability to undertake responsible action on the battlefield (whether directing the conduct of hostilities or directly engaging in these hostilities) must complement the reflections delineated above, and may play a vital role as we try to draw conclusions to the quandaries with which we are faced.

Human–machine interaction within the conduct of hostilities is referred to in military jargon as the “Force Mix” (Lucas 2016). Ethics research into the human and machine Force Mix, especially the moral implications for the human agents who use AI-based weapons systems, has arguably failed to keep pace with accelerating technological developments. This lacuna urgently needs to be filled. How service within the Force Mix affects the moral character of the human personnel involved is also our central focus in a new research project undertaken by the authors, in collaboration with a team of experts.¹⁰

We propose that a *virtue ethics* perspective is especially useful when investigating the ethical implications of human participation in the “Force Mix.” Virtue ethics, a philosophical approach (see Russell 2009) associated most closely with Aristotelianism (and the Thomistic school within Catholic moral thought), has been adopted within military training programs (see Moelker and Olsthoorn 2007 for a good overview of the interaction between virtue ethics and professional military ethics). Virtue ethics is uniquely flexible: rather than espouse fixed moral principles, it emphasizes acquiring appropriate dispositions for the proper exercise of one’s professional role (Vallor 2016: ch. 1). Paramount is the structural context—including, in the military setting, such important factors as combat unit membership and type of battlefield, as well as the technological setting—within which individuals act, and the ways in which their actions must be adjusted to fit that specific context. The use of AI within combat units

¹⁰“Warring with Machines: Military Applications of Artificial Intelligence and the Relevance of Virtue Ethics”; see <https://www.prio.org/Projects/Project/?x=1859> for a description of this project and the principal collaborators.

will inevitably alter these structural conditions, including the prerequisites for force cohesion, within which virtue is exercised. How should we think about the virtue of military personnel in light of this momentous, ongoing change?

Let us add that while often associated with Greek and Christian thought, virtue ethics also has significant parallels in Asian traditions of thought (as discussed e.g. in Vallor 2016), thus making it eminently suitable for a global conversation about ethics and AI.

Within military ethics, virtue has occupied a central role, pertaining not least to the inculcation of soldier identity, unit cohesion, pride, discipline, and conscience. The virtue-based ideal of the good and reliable soldier can be found across cultures and over time, albeit with different emphases and priorities. In spite of the many differences, the idea of the soldier and the officer as someone who must acquire and develop defined character traits or virtues, courage and prudence foremost among them, is central to most military cultures. In Western philosophy, it is exactly this way of thinking that has gone under the name of virtue ethics or, more specifically for the armed forces, professional military ethics.

It could be argued (Schulzke 2016) that an increased reliance on automated weapons and AI makes virtue ethics *less* central to the military enterprise, and that a more rules-based focus will be needed, since machines *per se* cannot have virtues, while they can indeed be programmed to follow rules. We would rather argue that the question of virtue becomes even more pressing in the face of AI, since the very role and competence of the human soldier is what is being augmented, challenged, and placed under great pressure. How do we ensure that soldiers and officers *maintain* those virtues that make them fit for military service and command, instead of delegating them to AI systems—and in the process, maybe, ignoring or losing them (“de-skilling”).

The debates between optimists and pessimists, delineated above, are also debates about the role of human virtue. As we have seen, while technology optimists will typically claim that lethal autonomous weapons systems (LAWS) will be superior to more human-driven systems because they tend toward an elimination of mistakes based on errors of judgment and the distortions introduced by strong emotion (Arkin 2010), those on the other side of the debate fear that military decision-making will suffer greatly if prudential reasoning, including such typically human phenomena as doubt and conscience, will be weakened, and war will be waged “without soul” (Morkevicius 2014; Riza 2013).

The questions that virtue ethics helps us posit concern what special moral challenges are faced by the human decision-maker and user of AI systems. Closely linked to that is the following question: How should that decision-maker and user be prepared and trained for action in the

battlefield? AI forces us to ask these questions in a new way, since the human–AI encounter is not only a traditional encounter between a user and a tool, in which the tool is essentially an artificial extension of the intentions of the user. Human–AI interaction also represents an encounter between a human agent on the one hand and a nonhuman system which is capable of making seemingly superior and not least self-sufficient, autonomous decisions, based on active learning, on the other. How should the human user of such systems think about his or her role and relationship vis-à-vis them?

In order to answer this question and determine the ethical implications of implementing AI technology into human practices of war, we hold that the following further questions have to be asked, and we conclude the present paper with an attempt at formulating them, indicating thereby the direction of our further research on the AI–human encounter in military settings:

Firstly, which are the specifically human qualities or virtues that remain crucial in guiding decisions about strategy as well as battlefield actions in war? How can we ensure that these qualities are not weakened or ignored as a result of the use of AI-based weapons systems? Or put in other words, how can we ensure that the implementation of AI systems does not lead to a “de-skilling” of the human actor?

Secondly, the Stoic ideal of peace of mind, balance, and moderation is often touted as a military ideal, based on a virtue-ethical tradition (Sherman 2007). But, as also intimated above, we must ask to what extent this ideal denies the importance of *emotions* for proper moral understanding. How does AI play into this debate about the role of emotions, such as fear, anger, distrust, doubt, and remorse—all feelings with significant relevance for decision-making in war? How will AI change the ways in which we understand, appreciate, critique, and develop emotions associated with the use of military force?

Thirdly, in the Socratic tradition, dialogue is considered a crucial prerequisite for the development of virtues and proper decision-making. What kind of a dialogue takes place in the human–AI encounter? Is an AI-based system with significant linguistic and machine-learning capabilities a real, trustworthy dialogue partner? The term “digital twin” is increasingly used in describing the interaction between AI-based and human-operated systems. Does this conceptualization truly capture the nature of the human–AI encounter in the deployment and operation of weapons systems? (Kim 2018).

And finally, and most generally, which of the virtues are most relevant to humans in the human–AI force mix? To what extent do those virtues—such as moderation, prudence, and courage, which we assume are among them—change their character as a result of the use of AI-based systems?

It is worth noting that these are questions with relevance well beyond military ethics. Virtue ethics is not only the dominant ethical framework for moral training within the military but is today also the dominant framework for thinking about professional ethics in a wide array of fields, as for example in the case of nursing ethics, and more widely the ethics of care, fields that are also very much undergoing change due to increased digitalization and widespread use of AI. Raising them within the context of military ethics first, however, does have the benefit of attacking the problem within the context where AI research arguably has come the furthest, and where the stakes are particularly high. Our belief is that a virtue-ethical approach helps us raise awareness and ask questions about how AI can be integrated in military settings in a way that takes seriously the crucial role of the human developer, operator, and user of AI systems. This, in turn, will help us focus on the sort of training and self-awareness that must be developed and fostered in tandem with the development and deployment of AI weapons systems.

To both ask and answer such questions intelligently requires, however, that we have a good overview of the nature of the ethical debates that confront us, and it is to that aim that we hope our reflections have contributed.

Acknowledgements The authors thank Joe Reeder and Sigurd Hovd for valuable comments to earlier drafts of the article.

References

- Arkin, R. (2010). The case for ethical autonomy in unmanned systems. *Journal of Military Ethics*, 9(4), 332–341.
- Asaro, P. M. (2007). Robots and responsibility from a legal perspective. *Proceedings of the IEEE*, 2007, 20–24.
- Barry, C., & Christie, L. (2018). The moral equality of combatants. In S. Lazar & H. Frowe (Eds.), *The Oxford handbook of ethics of war*. New York: Oxford UP.
- Cook, J. (2014). Book review: Killing without heart. *Journal of Military Ethics*, 13(1), 106–111.
- Danzig, R. (2018). *Technology roulette: Managing loss of control as many militaries pursue technological superiority*. Washington, DC: Center for a New American Security.
- Dennett, D. (1984). Cognitive wheels: The frame problem of AI. In C. Hookway (Ed.), *Minds, machines, and evolution* (pp. 129–151). Cambridge: Cambridge UP.
- Johnson, A. M., & Axinn, S. (2013). The morality of autonomous robots. *Journal of Military Ethics*, 12(2), 129–141. <https://doi.org/10.1080/15027570.2013.818399>.
- Kahn, P. W. (2002). The paradox of riskless warfare. *Philosophy and Public Policy Quarterly*, 22(3), 2–8.
- Kalshoven, F. (1984). The soldier and his golf clubs. In C. Swinarski (Ed.), *Studies and essays on international humanitarian law and Red Cross principles: In honor of Jean Pictet* (pp. 369–385). Geneva/The Hague: International Committee of the Red Cross/Martinus Nijhoff.
- Kim, J. (2018). *Digital twins and data analysis*. Available via Signal. Retrieved February 25, 2020, from <https://www.afcea.org/content/digital-twins-and-data-analysis>
- Klincewicz, M. (2015). Autonomous weapons systems, the frame problem and computer security. *Journal of Military Ethics*, 14(2), 162–176. <https://doi.org/10.1080/15027570.2015.1069013>.
- Knight, W. (2019). *China's military is rushing to use artificial intelligence*. Available via MIT Technology Review. Retrieved February 25, 2020, from <https://www.technologyreview.com/f/612915/chinas-military-is-rushing-to-use-artificial-intelligence/>
- Lambert, D. (2019). *La robotique et l'intelligence artificielle*. Paris: Fidélité/Éditions jésuites.
- Lin, P., Bekey, G., & Abney, K. (2008). *Autonomous military robotics: Risk, ethics, and design*. Washington, DC: U.S. Department of the Navy, Office of Naval Research.
- Lucas, G. (2016). *Military ethics: What everyone needs to know*. New York: Oxford UP.
- Makos, A. (2014). *A higher call: An incredible true story of combat and chivalry in the war-torn skies of World War II*. New York: Caliber.
- Manthorpe, R. (2017). *All that's good and bad about Silicon valley's stoicism's fad*. Available via Wired. Retrieved February 25, 2020, from <https://www.wired.co.uk/article/susan-fowler-uber-sexism-stoicism>
- Moelker, R., & Olsthoorn, P. (Eds.). (2007). Special issue: Virtue ethics and military ethics. *Journal of Military Ethics*, 6(4), 257–258.
- Morkevicus, V. (2014). Tin men: Ethics, cybernetics and the importance of soul. *Journal of Military Ethics*, 13(1), 3–19.
- Nagel, T. (1979). War and massacre. In *Mortal questions* (pp. 53–74). Cambridge: Cambridge UP.
- Olsthoorn, P. (2019). Risks, robots, and the honorableness of the military profession. In B. Koch (Ed.), *Chivalrous combatants* (pp. 161–178). Baden-Baden: Nomos.
- Patterson, R. (1985). *Image and reality in Plato's Metaphysics*. Indianapolis: Hackett.
- Reichberg, G. (2002). The intellectual virtues. In S. J. Pope (Ed.), *The ethics of Aquinas*. Washington, DC: Georgetown UP.
- Reichberg, G. (2017). *Thomas Aquinas on war and peace*. Cambridge: Cambridge UP.
- Renic, N. (2018). UAVs and the end of heroism? Historicising the ethical challenge of asymmetric violence. *Journal of Military Ethics*, 17(4), 188–197.
- Riza, M. S. (2013). *Killing without heart: Limits on robotic warfare in an age of persistent conflict*. Washington, DC: Potomac Books.
- Roff, H. M., & Danks, D. (2018). 'Trust but verify': The difficulty of trusting autonomous weapons systems. *Journal of Military Ethics*, 17(1), 2–20.
- Russell, D. C. (2009). *Practical intelligence and the virtues*. Oxford: Oxford UP.
- Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. New York: Viking.
- Scharre, P. (2018). *Army of none: Autonomous weapons and the future of war*. New York: W. W. Norton & Co..
- Scharre, P. (2019). Killer apps, the real dangers of an AI arms race. *Foreign Affairs*, 98(3), 135–144.
- Schulzke, M. (2016). Rethinking military virtue ethics in an age of unmanned weapons. *Journal of Military Ethics*, 15(3), 187–204.
- Sherman, N. (2007). *Stoic warriors: The ancient philosophy behind the military mind*. Oxford: Oxford UP.
- Simon, Y. R. (1969). *Freedom of choice*. New York: Fordham UP.
- Strawser, B. J. (2010). The duty to employ uninhabited aerial vehicles. *Journal of Military Ethics*, 9(4), 342–368.
- Syse, H. (2015). The moral equality of combatants. In J. T. Johnson & E. D. Patterson (Eds.), *The Ashgate research companion to military ethics* (pp. 259–270). Farnham: Ashgate.
- Syse, H. (2016). Dronenes etikk: Noen glimt fra debatten [The ethics of drones: Some snapshots from the debate]. In T. A. Berntsen, G. L. Dyndal, & S. R. Johansen (Eds.), *Når dronene våkner: Autonome våpensystemer og robotisering av krig* (pp. 243–254). Oslo: Cappelen Damm.
- Vallor, S. (2016). *Technology and the virtues*. New York: Oxford UP.
- Walzer, M. (1992). *Just and unjust wars* (2nd ed.). New York: Basic Books.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

