This is the author’s accepted manuscript of a book chapter published in Rocci Luppicini’s *The Changing Scope of Technoethics in Contemporary Society*. Cite as: Lambèr Royakkers and Peter Olsthoorn (2018) ‘Lethal Military Robots: Who is Responsible When Things Go Wrong?,’ in: Rocci Luppicini (ed.) *The Changing Scope of Technoethics in Contemporary Society*, IGI Global.

**Lethal Military Robots: Who is Responsible When Things Go Wrong**

Lambèr Royakkers

*Eindhoven University of Technology, The Netherlands*

Peter Olsthoorn

*Netherlands Defence Academy, The Netherlands*

**ABSTRACT**

*Although most unmanned systems that militaries use today are still unarmed and predominantly used for surveillance, it is especially the proliferation of armed military robots that raises some serious ethical questions. One of the most pressing concerns the moral responsibility in case a military robot uses violence in a way that would normally qualify as a war crime. In this article, we critically assess the chain of responsibility with respect to the deployment of both semi-autonomous and (learning) autonomous lethal military robots. We will start by looking at military commanders, as they are the ones with whom responsibility normally lies. We will argue that this is typically still the case when lethal robots kill wrongly – even if these robots act autonomously. Nonetheless, we will next look into the possible moral responsibility of the actors at the beginning and the end of the causal chain: those who design and manufacture armed military robots, and those who, far from the battlefield, remotely control them.*

Key words: Military ethics, Responsibility, Military robots, Unmanned systems, Accountability

**INTRODUCTION**

Although the use of unmanned systems is still in its infancy in most armed forces, some militaries, especially those of the US and Israel, have developed and deployed highly advanced drones. Even though the majority of these unmanned systems used in operations today are unarmed and mainly used for reconnaissance and mine clearing, the increase of the number of armed military robots, especially airborne ones, is undeniable. Certainly, on the face of it, unmanned systems have some strong benefits that could reduce the number of ‘unfortunate incidents’ on the battlefield. To start with, the main causes of misconduct on the battlefield: frustration, boredom, and anger are diminished.[[1]](#endnote-1) What’s more, these unmanned systems have no instinct of self-preservation, and are able to hold their fire in critical situations. On the other hand, the use of robots raises some serious ethical questions. For instance, under what circumstances, and to what extent, do we allow robots to act autonomously? What precautions should (and can) we take to prevent robots from running amok? Would the use of military robots not be counterproductive to winning the hearts and minds of occupied populations, or result in more desperate terrorist-tactics given an increasing asymmetry in warfare? (See for an overview Lin, Bekey, and Abney, 2008; Lichocki, Kahn & Billard. 2011; Olsthoorn & Royakkers 2011; Schwarz 2017). A particularly pressing question is what to do when things go wrong: who, if anyone, can be held morally accountable *in reason* for an act of violence that a) involves a military robot; and b) would normally be described as a war crime?

The answer to that latter question depends on the answer to a prior one: when is there reasonable ground to hold an agent morally responsible for a certain outcome in the first place? Following Fischer & Ravizza (1998) on moral responsibility, we will assume here that agents can only in reason be held responsible if they are *moral* agents, that is, persons (or organizations) who have *control* over their behavior and the resulting consequences. This means that agents can be held responsible for a certain decision only insofar as they have been able to make it *in freedom* and *knowingly*. The first term means that it is not reasonable to hold agents responsible for actions or their consequences if they were not coerced or under duress. The second term, ‘knowingly,’ has an important normative aspect in that it relates to what people should know, or can with reason be expected to know, with respect to the relevant facts surrounding their decision or action.[[2]](#endnote-2)

According to some authors (Asaro, 2007; Sparrow, 2007; Sharkey, 2008), the use of armed military robots makes the attribution of responsibility problematic, as it is not sufficiently clear who can be held responsible for civilian casualties and other collateral damage that result from the use of military robots, whether by mechanical error or failing judgment. Is it the designer/programmer, the field commander, the robot manufacturer, the robot controller/supervisor, or the nation that commissioned the robot? The answer to that question depends on a number of factors. For instance, was the cause a programming error, a malfunctioning, an accident, or intentional misuse? Or did the procedure include a ‘man-in-the-loop,’ that is, an element of human control, or was the military robot a fully autonomous or even learning machine?

As to that last question, this paper distinguishes between semi-autonomous robots, autonomous (but non-learning) robots, and learning robots. To start with the latter, learning robots are able to develop new behavioral patterns without human intervention; these robots are able to go beyond the parameters they left the factory with, as the robot itself can change them in its interaction with the operating environment (Matthias, 2004). An example is a system that was developed for the automatic diagnosis of lung cancer, and which is able to learn to identify cancer cells on the basis of microscope images of specimens of needle biopsies obtained from the bodies of the persons to be diagnosed (Matthias, 2004). More learning systems are in development, and most of them will pose no moral problems. But this would be different in the case of lethal military learning robots: seeing that it will often be impossible to predict the future behavior of these robots, it is hard to see how one could have sufficient *control* over their actions, which in its turn makes it difficult to determine who can be held responsible with reason.

Autonomous robots are robots that are based on conclusions derived from gathered information and preprogrammed constraints, and are capable of independently selecting and engaging targets (Crootof 2015). This in contrast to semi-autonomous robots where a human must take some affirmative action before a specific target is elected or engaged, that is, they require a ‘man-in-the-loop.’ This means that the decision to open fire, or more in general, the taking of any action that could threaten human life, is to be considered and approved by a human agent. According to Sparrow (2007) and Singer (2009), this precondition is essential to avoid the problems with respect to the attribution of responsibility; it is not without reason that the “International Law of Armed Conflict dictates that unmanned systems cannot fire their weapons without a human operator in the loop” (Isenberg, 2007).

Yet, while it is certainly true that currently human agents are kept ‘in-the-loop,’ it is not certain, or even likely, that this will remain so. The logic that brought unmanned systems more or less leads naturally to the wish to take the human agent out of the system altogether (Sparrow, 2011, 121; see also Sullins, 2010), and it seems almost a given that the future will bring more autonomous, and possibly learning, robots. Illustrative for this direction is that The United States is investing heavily in autonomous robots, and the U.S. Department of Defence has described increasing autonomous capabilities as a high priority. One of the ultimate goals is to take the man out-of-the-loop (US Department of Defense 2013), which raises the essential moral question: should we relinquish the decision to kill a human to a non-human machine? (Johnson & Axinn, 2013). That a lethal decision should be made only by a human and not a machine was one of the main reasons that in 2009 four leading scientists (Noel Sharkey, Peter Asaro, Robert Sparrow and Jürgen Altmann) established the International Committee for Robot Arms Control, ICRAC. The committee wants to restrict the use of armed robots for military purposes as much as possible (Altmann et al., 2013). In 2010 in Berlin, the committee organised the expert workshop Limiting Armed Tele-Operated and Autonomous Systems and invited scientists, politicians and military delegates. Following the workshop, a statement was made – signed by a large majority of those present – which emphasised that a ban ought to be imposed on autonomous armed military robots because a human must always take life-and-death decisions. Part of the statement read, “it is unacceptable for machines to control, determine, or decide upon the application of force or violence in conflict or war. In all cases where such a decision must be made, at least one human being must be held personally responsible and legally accountable for the decision and its foreseeable consequences”.[[3]](#endnote-3) In 2015, the Future of Life Institute gained considerable publicity in 2015 when it distributed an open letter signed by more than twenty thousand artificial intelligence and robotics researchers (and others), which called for a “ban on offensive autonomous weapons beyond meaningful human control”[[4]](#endnote-4). “Meaningful human control” has been identified as key for the responsible design of autonomous systems operating in circumstances where human life is at stake. By preserving meaningful human control human safety can be better protected and “responsibility gaps” can be avoided. However, we still lack a satisfactory definition of what meaningful human control precisely means in relation to killer robots (Ekelhof, forthcoming, see also Crootof, 2016). Sparrow (2007) was one of the first authors to discuss this question, and he claims that since no one can reasonably be held responsible for autonomous robots’ behavior, robots should not be allowed to make life-and-death decisions. In Sparrow’s reasoning, considerations of responsibility occupy centre stage. Mukerji (2016) agrees with Sparrow’s conclusion that robots should not be allowed to make life-and-death decisions, but on a different argument. He argues that only subjects who are capable of responsibility should make morally significant choices. Since autonomous robots seem to lack that capability, it follows that they should not be entrusted to make life-and-death decisions and act on them. However, some researchers dispute these arguments and that the problem of determining responsibility for autonomous military robots can be solved by addressing it within the context of the military chain of commands (e.g., Schulzke, 2013, and Champagne and Tonkens, 2015), or by developing responsibility practices that clearly establish lines of accountability (Noorman & Johnson, 2014, and Noorman, 2014).

The above raises plenty of issues to critically assess the attribution of responsibility with respect to the deployment of both semi-autonomous and (learning) autonomous lethal military robots. In this article, we will argue that, in contrast to what some other authors hold, that we can reasonably assign responsibility for the development and deployment of these robots to their developers and commanders. A neglected problem, however, regards the assignment of responsibility at the end of the causal chain: the human operators who remotely control armed military robots from behind a computer screen. The depersonalization of war, the dehumanization of the enemy, and the moralization of technology strengthened by the speed of decision-making, can disengage the human operator morally and emotionally. We will show that to make deliberate, and thus responsible, decisions it is essential that operators have control over their decisions, and a vivid awareness of what is at stake. We will start by looking at the role of those with whom responsibility normally lies, the commanders. After that, we will turn to those at the beginning and the end of the causal chain, respectively the manufacturers and designers, and the human operators. We end by drawing some conclusions.

**Ultimately responsible: the commanding officer**

Discrimination and proportionality are the two main *jus in bello* principles of the Just War Tradition. The first forbids the intentional killing of ‘the innocent,’ and the underlying idea is that civilians should not be made to suffer in war; overall, this is a rights-based principle. The principle of proportionality is a consequentialist one: enemy combatants should not be subjected to unnecessary suffering and superfluous injury, and a mission is permitted only if the expected military gain outweighs the expected number of unintended (the intended killing of civilians being prohibited anyway) civilian casualties. These principles have materialized in international humanitarian law and, for instance, in international treaties banning, regulating, or limiting the possession and use of certain weaponry.

Sparrow has pointed out that these principles assume that “we can identify the persons responsible for the actions that these principles are intended to govern” (Sparrow, 2007; see also Fieser & Dowden, 2007). Although some authors (Sparrow, 2007; Asaro, 2007; Sharkey, 2008) hold that the use of military robots makes this attribution of responsibility complex, it seems evident that the ultimate responsibility for, for instance, a war crime that crucially involved a military robot lies in the hands of the operators, the commanding officers who ordered the deployment of that robot, the military organization on whose behalf they did so, and the state of which that military is an instrument of. So, at the heart of it, the use of unmanned systems is, as far as responsibility concerns at least, not *very* different from using an aircraft to drop a bomb from a high altitude.

Still, this straightforward approach only works as long as there is ‘a human in the loop’ – from a legal and ethical perspective that human element seems a precondition for the attribution of responsibility (cf. Singer, 2009). Some have therefore argued that the attribution of responsibility in the case of autonomous military robots is so problematic that for that reason alone their use should be forbidden (Asaro, 2007; Sparrow, 2007). One of the most outspoken advocates of this position on the use of autonomous military robots is Noel Sharkey. In his view, the problem is that armed autonomous robots simply cannot be held accountable for their actions, as “[t]here is no way to punish a robot” (2010, 380). According to Sharkey this “leaves the question about who is responsible somewhere along the long causal chain that includes: the manufacturer for the description they gave, the programmer, the designer, the Department of Defense, the generals or admirals in charge of the operation and the operator” (2010, 380-1). If it proves impossible to establish where the responsibility lies, the use of unmanned robots should be, according to Sharkey, illegal. Other experts in the field hold similar views: a majority of the participants of an Expert Workshop on “Arms Control for Robots – Limiting Armed Tele-Operated and Autonomous Systems” organized by the International Committee for Robot Arms Control (ICRAC) in 2010, for example, found it ‘unacceptable for machines to control, determine, or decide upon the application of force or violence in conflict or war. In all cases in which such a decision must be made, at least one human being must be held personally responsible and legally accountable for the decision and its foreseeable consequences’ (Altmann et al., 2013).[[5]](#endnote-5) The general drift of these comments is that responsibility evaporates without a human in the loop.

But on second though it seems clear that, also without a human agent directly controlling the robot, it is still a human agent that has decided on whether or not to deploy this robot: ‘even if a system is fully autonomous, it does not mean that no humans are involved. Someone has to plan the operation, define the parameters, prescribe the rules of engagement, and deploy the system’ (Quintana, 2008). Krishnan therefore thinks that “the legal problems with regard to accountability might be smaller than some critics of military robots believe. (...) If the robot does not operate within the boundaries of its specified parameters, it is the manufacturer’s fault. If the robot is used in circumstances that make its use illegal, then it is the commander’s fault” (2009, 105). The basis of this argument is the Doctrine of Command Responsibility, and although this doctrine, ‘that goes back to the dawn of time’ (Garraway 2009, 704), is interpreted differently by different authors (Garraway, 2009), it will usually cover the deployment of armed military robots. So, if such a robot is used in circumstances that make its use illegal, then it is the commander’s fault. The question then arises: what degree of knowledge must a commander have about the workings of a military robot in order to be reasonably responsible? Commanders are expected to take “all necessary and reasonable measures in their power” to prevent war crimes (Heckaerst & Doswald-Beck, 2006), and that requires according to Dunlap (forthcoming) a reasonable understanding of the foreseeable consequences of the deployment of armed military robots, and that we cannot hold the commanding officer responsible if an armed military robot would perform unforeseeable consequences. The latter is in line with the argument of Sparrow (2007), that it is unfair to assign responsibility to the commanding officer of learning autonomous robots, since “[t]he use of autonomous weapons (…) involves a risk that military personnel will be held responsible for the actions of machines whose decisions they did not control”. This problem. However, can be overcome by using a more explicit social contract or “blank check”. According to Champagne and Tonkens (2015), Sparrow overlooks the possibility of a sufficient high-ranking commanding officer accepting responsibility for the robot’s performance, and thus being responsible for any violation of the rules for the ethical conduct of warfare. So, “the moral blame and accompanying punishment could be placed squarely on a human agent (or agents) who, through her own volition, has traded a part of her freedoms for the prestige of occupying a high-ranking position in a given social hierarchy (…). If no one in willing to accept this responsibility, then they should not deploy autonomous killer robots in the first place” (Champagne & Tonkens, 2015, 136; see also Schulzke 2013).

That the future will hold more autonomous systems seems almost a given, and although renouncing certain types of autonomous robots might be a good idea for many reasons, a lack of clarity as to who is responsible for their use is probably not among them (see also Kershnar, 2013). Whether one would want to have that responsibility is a different question altogether.

**Responsibility of designers**

Although in general commanders can be held responsible for the decisions they make, in particular cases the matter is not that clear cut; in some specific cases it might be unjust to hold a commanding officer responsible for the actions of military robots. For instance, it is dubious whether we can hold a commanding officer responsible if, for example, civilian casualties (or casualties among own military personnel) are the result of a malfunctioning military robot, or of a programming error. There have already been instances of such malfunctioning: in October 2007 in South Africa a malfunctioning robotic gun fired five hundred high explosive rounds, killing nine soldiers and seriously injuring fourteen (Shachtman, 2007), while in April 2008 a SWORDS, an armed unmanned ground vehicle, produced unintended movements (Magnuson, 2008). Although partly a result of their flying low and slow, the relatively high loss rate of military robots in military operations is perhaps also an indication of their unreliability. For example, some years ago the US Air Force reported that it had lost 50 percent of its then 90 Predators (Jordan, 2007). Although it is clear that both military personnel and local civilians should be protected from malfunctioning military robots, the question is whether designers can be reasonably held responsible for casualties caused by a malfunctioning system. This is a difficult question for more than one reason. Especially with regard to learning autonomous robots, since according to Sparrow (2007) the designers then lose their responsibility for how the robot will act in the future: “[t]he connection between programmers/designers and the results of the system, which would ground the attribution of responsibility, is broken by the autonomy of the system” (Sparrow 207, p. 70). Sparrow treats the autonomy of military robots as a ‘black box’, leaving its working opaque, and then drawing out implications as if the autonomy of military robots are the same as human autonomy (Noorman & Johnson, 2014). This is, however, misleading since although autonomous military robots “would be able to go beyond their programming by learning new information and developing novel solutions to problems, the range of possible actions open to them would still constrained by their initial programming” (Schulzke, 2013, 213), and do not exhibit random behavior. These robots are designed to perform particular tasks and the extent to which their actions can vary is constrained by the designers. These constrains for military robots should be based on International Humanitarian Law (including the principle of discrimination, the principle of proportionality, and the principle of superfluous injury or unnecessary suffering) and the rules of engagement.

The possibility that a military robot may engage the wrong targets – because it cannot adequately distinguish citizens from soldiers – could be an acknowledged limitation of the system. If the designers have made this clear to those who have purchased or deployed a system, then, Sparrow (2007) argues, they can no longer be held responsible should this occur; in that case the responsibility should be assumed by the nation or commanding officer who (willfully and knowingly) decided to send the robot into the battlefield despite its known limitations. Conversely, one could argue that designers have an (at least moral) obligation to make products that are ‘safe’ and satisfy the legal and ethical conditions, and that it poses an unacceptable risk to soldiers and civilians alike to introduce a robot with such ‘an acknowledged limitation’ onto the battlefield. Although it might seem somewhat inconsistent to require armed military robots to be safe, as their main purpose is to destroy, in reality there is of course ample reason to be extra cautious with systems that are designed to be lethal. As it stands, there is, regarding the responsibility of designers of military robots (and militaries that deploy them), no lack of relevant legal and ethical concepts we can turn to, with *liability* seeming to be the most appropriate one (see also Lucas, 2011). The advantage of responsibility as liability *in this case*, that is, of malfunctioning robots, is that it does not require the government to foresee the consequences of new technology but rather – under certain conditions – makes the ones developing the technologies legally liable for these consequences. This places the responsibility where it can be met best: in the hands of the ones developing the technology. They have the best knowledge of new innovations and their possible effects and they are in the best position to avert certain disadvantages. Moreover, a scheme of liability would stimulate them to take their (active) moral responsibility seriously (Van de Poel & Royakkers, 2011).

Regrettably, in an international environment liability issues are not always that straightforward; for starters, the conditions that have to be met for someone to be liable depend on national laws that differ from country to country. Generally speaking, however, in the Western world the main condition for liability is negligence. Negligence implies that a manufacturer failed to do something that was morally or legally required, and thus can be held responsible for certain harms produced by his product – in legal terminology this is called *reasonable care*. Reasonable care is based on the sometimes implicit moral responsibilities of programmers and designers – it is not something made explicit in law. In military practice, however, designers are rarely held responsible for accidents caused by poor design (Thompson, 2007). It is therefore important that there will be a regulatory framework that entails ‘reasonable care’ to be met by designers of armed military robots. A good candidate to offer some guidelines for a formal approach of reasonable care would be the *precautionary principle*.[[6]](#endnote-6) This principle was initially proposed to deal with environmental problems, yet it has far wider applications. It is especially suited for situations in which it is not possible to calculate the risks because we have insufficient scientific knowledge. In general, the precautionary principle states that precautionary measures must be taken if there are indications that there are certain hazards despite the fact that the precise nature and extent of these hazards cannot be scientifically established (Raffensperger & Tickner, 1999). Applying the precautionary principle to armed military robots would imply that the potential hazards of military robots should first be properly assessed before they could be introduced into the armed forces on a large scale, and that these robots, for example, should design with some emergency stop mechanism in case of malfunctioning such as happened in South Africa with a SWORDS (see above). The present practice to test if a system is properly functioning is done through simulations and experiments in laboratories or in small-scale field tests. Although such simulations, experiments, and tests are very important with respect to technological innovations, they do not always provide complete and reliable knowledge of the functioning of technological products and the potential hazards and risks involved. For example, laboratory and field tests are not always representative of the dynamic and complex circumstances in which military robots have to function. Evidently, it is difficult to capture the fog of war in a test setting.

Nevertheless, it seems that regulation on risk assessment can provide us with a regulatory framework for the design of military robots by formulating a set of boundary conditions for the design, production, and use of military robot technologies, such as that a military robot follows the rules of engagement and the safety protocols. If such a regulatory framework meets certain criteria, it could be considered an adequate way of dealing with the ethical issues the design of a technology raises with regard to safety and responsibility. Such a regulation for risk management should be laid down in specific norms, such as the ISO 10218 requirements for industrial robots. This norm ‘specifies requirements and guidelines for the inherent safe design, protective measures, and information for use of industrial robots. It describes basic hazards associated with robots, and provides requirements to eliminate or adequately reduce the risks associated with these hazards’ (ISO 2011). Although this ISO norm was developed for industrial robots, it can be used outside that context: “Examples of non-industrial robot applications include, but are not limited to: undersea, military and space robots” (ISO 2011).

Until now, however, such a regulatory framework specifically for military robots has been lacking (see also Lin et al., 2008, and Lucas, 2011). The potential use of military robots in saving ‘friendly’ human lives possibly explains why most of today’s governments refrain from regulation with respect to the design of military robots, and especially from regulation that could lead to a ban on certain designs of military robots. On the other hand, if no-one is legally responsible for the malfunction of a military robot, there is hardly any incentive to learn from mistakes or do better in future.

Driven by this faith in the rationalisation of warfare by constraining autonomous armed robots so that they meet certain criteria, Arkin (2009) investigates the introduction of an adequate artificial ethical conscience into the armed robots by programming international humanitarian law and the rules of engagement into the robot. An advantage of these robots is that they “have the potential capability of independently and objectively monitoring ethical behaviour in the battlefield by all parties and reporting infractions that might be observed. This presence alone might possibly lead to a reduction in human ethical infractions” (Arkin, 2007). The idea of the development of autonomous machines with an additional ethical dimension is quite a new emerging field of machine ethics. These robots, with the help of formal logic, are “able to calculate the best action in ethical dilemmas using ethical principles” (Anderson & Anderson, 2007). It is thus assumed that it is sufficient to represent ethical theory in terms of a logical theory and to deduce the consequences of that theory. Beavers (2012) sees machine ethics as a threat to *ethical nihilism*, “the doctrine that states that morality needs no internal sanctions, that ethics can get by without moral ‘weight,’ that is, without some type of psychological force that restrains the satisfaction of our desire and that makes us care about our moral condition in the first place” (Beavers, 2012, 343). According to Johnson and Axinn (2013), robots with an artificial ethical conscience can only mimic moral actions with no human emotions and no feelings about the seriousness of killing a human. Therefore, they conclude that autonomous robots cannot be moral.

This view that ethics can be made computable also misunderstands the unique – non-reducible – nature of ethics. Arkin (2009) agrees that some ethical theories, such as virtue ethics, do not lend themselves well by definition to a model based on a strict ethical code. But he claims that the solution is simply to eliminate ethical approaches that refuse such reduction. However, the hallmark of ethics is its non-reducibility (Royakkers & Topolski, 2014). While many ethical situations may be reducible, it is the ability to act ethically in situations that call for a judgement that is distinctly human. A consequence of this approach is that ethical principles themselves will be modified to suit the needs of a technological imperative: “Technology perpetually threatens to co-opt ethics. Efficient means tend to become ends in themselves by means of the ‘technological imperative’ in which it becomes perceived as morally permissible to use a tool merely because we have it” (Kaagman & Kaufman, 2009).

Another problem concerning the attempt to program the rules of the law of war and engagement into robots is that these rules are subject to interpretation and they do not provide a ready answer in any given situation. Moreover, the application of these rules is context dependent and requires the ability to understand complex social situations (Asaro, 2009). This ability is necessary in order to make ethical decisions, and as long as robots lack this ability, they will certainly not be able to take better ethical decisions than humans (Sullins, 2010).

In conclusion, although risks are inherent to the use of military robots, there is a moral obligation to limit these risks to an acceptable level by procedures of risk assessment. To ensure the allocation of responsibility for malfunction or program errors of military robots, it should be specified what this acceptable level is. Still, the possibility of an accident remains, and Krishnan suggests that in such cases all the military robots of the type that was involved in the accident should be abandoned: “If the weapon is not withdrawn from service, it can only be interpreted as a failure of politics and maybe as a war crime or crime against humanity committed by the political leadership of a state” (2009, 105).

**Responsibility of human operators**

What is often underplayed in the literature on responsibility and military robots is the role of the human operators; an omission that is all the more regrettable since it is, in fact, especially there where a number of complexities lie. In this section, we will work out the position of the human operator, often working far away from the actual battlefield. Today’s UAVs connect human operators with the war zone; they are the eyes – and sometimes hands – of the tele-soldier. Partly autonomous military robots such as the Predator or the Reaper, which are able to navigate independently to their objective but require a human operator to open fire,send GPS-coordinates and camera images back to an operator who then, based on the information projected on his computer screen, decides whether or not to engage a target.

The point is that the things they see on screen can emotionally and psychologically affect these human operators, influencing their decision-making. Although fighting from behind a computer is not as emotionally potent as being on the battlefield, killing from a distance remains stressful; various studies have reported physical and emotional fatigue and increased tensions in the private lives of military personnel operating the Predatorsin Iraq and Afghanistan (Donnelly, 2005; Kaplan, 2006; Lee, 2012). For example, a drone pilot may witness war crimes which he is unable to prevent, or he may even see how his own actions kill civilians; unfortunately no longer a hypothetical case.

The ensuing ‘residual stress’ of human operators has led to proposals to diminish these tensions. In particular, the visual interface can play an important role in reducing stress; interfaces that only show abstract and indirect images of the battlefield will probably cause less stress than the more advanced real images (Singer, 2009). From a technical perspective this proposal is a feasible one, since it will not be hard to digitally recode the war scene in such a way that it induces less psychological discomfort with the war operator. This cure may have some unwanted side-effects though. Showing abstract images would in fact dehumanize the enemy, and as a result could desensitize military personnel operating unmanned systems.

Today’s operators already find it at times difficult to distinguish between playing a video war game and operating an actual drone (Singer, 2009). From a technological perspective it is only a minor step to let operators think they are playing a computer game, and destroying enemy ‘avatars,’ while they are actually killing real people at the other side of the globe. In that case, it would be no longer the real war that would numb the soldier, but the digital recoding of it. From a moral point of view this would mean that a soldier is likely to become even further physically and emotionally detached from his actions than is already the case at present (cf. Royakkers & Van Est, 2010). Thus, where increased detachment may reduce or even eliminate human operator stress, it could at the same time limit reflection; human operators only focused on the outcome – the targeting of the blips on a screen – might not be fully aware of the consequences of their decisions.

This last observation brings us to the important role dehumanization can play. Social psychologists point out that dehumanization, that is, seeing people for something less than human, can open the door to more serious forms of unethical conduct (see for instance Bandura, 1999, 200; Mastroianni, 2011; Moller & Deci, 2010, 43-4; see also Slim, 2007, 218). It is rather hard to imagine how to have respect for the local population, a vital element of the hearts and minds approach, from, for instance, a control room in Nevada (where the pilots of Predators and Reapers mostly work from). The most effective remedy for preventing unethical conduct is ‘humanization,’ a not so clearly defined concept that, however, includes the affirmation of common humanity, instead of distancing oneself ‘from others or divesting them from human qualities’ (Bandura, 1999, 202-3). Hugo Slim explains in *Killing Civilians* that, to be effective, civilian immunity “requires that armed people find a fundamental *identification* with those called civilians and not an excessive *distinction* from them” (2007, 34). As shown by the famous Milgram experiments on obedience, it is easier to be cruel when the other does not have a face (Milgram, 1974). At a time that unmanned aerial vehicles take out insurgents from afar, and a human operator thinks that his job is “like a video game. It can get a little bloodthirsty. But it’s fucking cool” (Singer, 2009, 332), that ‘face’ is hardly ever there. With such a distance – physical, but also psychological – between a soldier and the horrors of war, it has to be feared that this can limit, or even eliminate, proper reflection of the human operators on their decisions to open fire (see for a different view Lee, 2012; Vincente, 2013).[[7]](#endnote-7)

Unclear responsibility plays a role here: the social-psychologist Albert Bandura counts the displacement and diffusion of responsibility among the “many social and psychological manoeuvres by which moral self-sanctions can be disengaged from inhumane conduct” (1999, 194), which leaves the answer to the question of responsibility of the human operator far from straightforward.

Sullins (2010) believes that in the foreseeable future such inhumane conduct could be a thing of the past, as life-and-death decisions might then be mediated by a computer-aided diagnosis of the war situation, and that military robots may even have ethical constraints built into their design – a so-called ‘ethical governor’ that suppresses unethical lethal behavior. For example, Arkin (2007) has done research (sponsored by the US Army) to create a mathematical decision mechanism consisting of prohibitions and obligations derived directly from the laws of war. The idea is that future military robots might give a warning if orders, according to their ethical governor, are illegal or unethical. For example, a military robot might advise a human operator not to fire because from an evaluation of the camera images the robot informs the operator that he or she is about to attack non-combatants, that is, the software of the military robot diagnosing a war situation provides a human operator with ethical advice.

Achterhuis (1998) has called this development, in which for instance an ethical governor helps to shape moral decision-making, the ‘moralization of technology.’ Instead of only moralizing the *human operator* (‘do not shoot non-combatants’), it is the *material environment* that should be moralized, Achterhuis claims. In our case, the military robot should then see to it that no rules of engagement are violated (cf. Verbeek, 2005). Achterhuis’ plea for a moralization of technology has received severe criticism, with the main criticism being that human freedom is diminished when human actions are explicitly and deliberately steered with the help of technology (cf. Verbeek, 2009). A consequence may be that humans would simply show a type of behavior that was desired by the designers of the technology instead of explicitly choosing to act this way, diminishing the freedom of, and control over, their actions. According to Cummings (2006), this would also be the case with ethical governors, since they may form a ‘moral buffer’ between human operators and their actions, allowing human operators to tell themselves that it was the military robot that took the decision.

The moralization of military robots could blur the line between semi-autonomous and autonomous systems, as the decision of a human operator is not the result of human deliberation, but is mainly determined or even enforced by a military robot. Another possible effect is that human operators may come to over-rely on military robots (Cummings, 2006). This is bound to happen more often in future, due to an ongoing shift from controlling to monitoring.

In brief, the moralizing of a military robot may deprive the human operator from *controlling* the situation; his future role may be restricted to *monitoring*. For our case, this would imply that we could no longer hold a human operator in reason responsible for his decisions, since it would not really be the operator taking the decisions but a military robot. This could have consequences for the question of responsibility in another way too: Detert et al. (2008) have argued that people who believe that they have little personal control in certain situations – such as those who monitor – are more likely to go along with rules, decisions, and situations even if they are unethical or have harmful effects. However, Leveringshaus (2016) argues that this “lack of moral awareness” does not relieve the human operator of taking “adequate steps to attain the morally relevant facts that enabled him to assess the risks arising from automation”, and that he can still be judged negligent for deploying and relying on a technology that put innocent lives at risks.

**Conclusion**

Niccolò Machiavelli held that in war nothing ever really changes, and hence thought that the invention of the firearm amounted to nothing more than just a new variety of the age-old catapult. It is tempting to think likewise about the progress of unmanned systems, that is, as a development that does not really raise issues different from those raised a long time ago by artillery, and more recently by highflying bombers. And in part, there is something to be said for this view. But the fact is that Machiavelli was, of course, wrong; the invention of the firearm proved as crucial for warfare as the spread of the stirrup some thousand years before. Possibly, the use of unmanned systems will prove to be equally significant, especially since the development of these systems has only just begun. That will raise a host of ethical issues that are truly new, and this article addressed one of them.

We have seen that most military robots currently find their applications in surveillance, reconnaissance, and the location and destruction of mines and IEDs. These robots are unarmed, harm no one, and save lives. But not all military robots are unarmed; different types of armed military robots are currently used on the battlefield. This development may cause an important ethical problem in the assignment of responsibility in the long causal chain involved in the design and deployment of armed military robots, stretching from the manufacturer, programmer, and designer, to the departments of defense, commanding commander, and human operator. According to Sparrow (2007) a necessary condition for fighting a just war, under the principle of jus in bello, is that there is someone that can reasonably be held responsible for an atrocity, such as killing non-combatants, in the course of war; this condition also applies when a military robot is employed. Especially with respect to armed learning military robots there are serious legal and ethical questions regarding responsibility due to the problematic attribution of responsibility: since no-one can have sufficient *control* over the actions of these robots, and no one can foresee the consequences of these actions.

The issue of responsibility is the main reason why many lawyers and ethicists argue that a human agent should remain in-the-loop, so that traditional accountability can be ensured, and that this would therefore avoid the problem of the allocation of responsibility. Yet, in section 2 we have argued that even in the case of learning autonomous robots the assignment of responsibility is still very well possible: it normally lies with the commanding officer. However, in the subsequent two sections we have shown that the existence of responsibility gaps cannot always be avoided. Müller (2016) states that possibility of this existence should not morally preclude the use of autonomous weapons, since in civil life we only expect ‘due care’ from those involved. He states that if technology produces rare cases of killings where no person is responsible, this does not by itself compel a ban of autonomous robots, but that we have to regulate autonomous robots is such a way that they reduce the frequency of responsibility gaps. Therefore, we should be focused on developing responsibility practices that work to minimize risk and that clearly establish lines of responsibility. This means, that we have to negotiate about how to best assign responsibility and what that entails. Noorman and Johnson conclude that “[d]elegation of responsibility to human and non-human components is a sociotechnical design choice, not an inevitable outcome of technological development. Robots for which no human actor can be held responsible are poorly designed sociotechnical systems (Noorman & Johnson 2014, 61; see also Noorman 2014). This will be a significant challenge, since this implies that this shared responsibility is distributed broadly and which can lead to ‘the problem of many hands’ (see Van de Poel et al. 2015), meaning that it can be very difficult to determine what share of responsibility each person has in this sociotechnical system when something goes wrong.

For example, developing responsibility practices is necessary for a lethal mishap due to a malfunctioning of a military robot. The application of the legal concept of ‘product liability’ is inadequate to deal with this problem since it misses a clear regulatory framework for the design of military robots about safety and risk assessment, and more specifically, what degrees of reliability should autonomous robots meet in terms of proportionality and discrimination in their decision to attack. Therefore, governments should have a moral obligation to determine the appropriate tolerance level for autonomous robots, and to licence them for use only after a thorough testing and inspection process (Simpson and Müller 2016).

As far as semi-autonomous robots are concerned, we have discussed the largely neglected problem of the responsibility of those at the other end of the causal chain, the human operators. After all, they are the ones who decide to use lethal force, and at first sight it might seem that the attribution of responsibility is clear-cut in their case. We have seen, however, that the criteria for attributing responsibility might not always be met. The depersonalization of war, the dehumanizing the enemy, and the moralization of technology, emotionally and morally disengages the human operator. The aim of this disengagement is to reduce the psychological stress among human operators being simultaneously ‘present’ in and absent from the battlefield. Unfortunately, this can also limit, or even eliminate, proper reflection on their decisions; human operators have partly lost control over their decisions. Hence, if we want to hold human operators in reason responsible for their decisions, it is essential that they have control over their decisions, and a clear awareness of what is at stake.

It is, in all, important to strike a proper balance between emotional and moral attachment and detachment. This requires a lot from the computer systems that human operators use to make their life-and-death decisions. On the one hand, such systems should communicate the moral reality of the consequences of the decisions of human operators, and on the other hand such systems should reduce the strong emotions human operators feel to reduce the number of war crimes (Sparrow 2009). To develop such systems is a real challenge, but necessary to solve the problem of the attribution of responsibility, and a precondition for fighting justly in the robotic era.

**References**

Achterhuis, H. (1998). *De erfenis van de utopie*. Amsterdam: Ambo.

Alley, R. (2013). *The drone debate. Sudden bullet or slow boomerang?* Wellington: The Centre for Strategic Studies, New Zealand, Victoria University of Wellington.

Altmann, J., Asaro, P., Sharkey, N., & Sparrow, R. (2013). Armed military robots: editorial. *Ethics and Information Technology*, *15*(2), 73-76.

Anderson, M. & Anderson, S.L. (2007). Machine ethics: creating an ethical intelligent agent. *AI Magazine*, *28*(4), 15-26.

Arkin, R.C. (2007). *Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture* (Technical report GIT-GVU-07-11). Atlanta: Georgia Institute of Technology.

Arkin, R.C. (2009). Ethical robots in warfare. *IEEE Technology and Society Magazine*, *28*(1), 30-33.

Asaro, P. (2007). Robots and responsibility from a legal perspective. In: *Proceedings of the 8th IEEE 2007 International Conference on Robotics and Automation*. Workshop on RoboEthics, Rome, 14 April 2007.

Asaro, P.M. (2008). How just could a robot war be? In A. Briggle, K. Waelbers, & Ph. Brey (Eds.), *Current issues in computing and philosophy* (pp. 50-64). Amsterdam: IOS Press.

Asaro, P.M. (2009). Modeling the moral user. *IEEE Technology and Society*, *28*(1), 20-24.

Beavers, A.F. (2012). Moral machines and the threat of ethical nihilism. In: P. Lin, K. Abney & G.A. Bekey (Eds.), *Robot ethics. The ethical and social implications of robotics* (pp. 333-344. Cambridge: The MIT Press.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, *3*(3), 193-209.

Champagne, M., & Tonkens, R. (2015). Bridging the responsibility gap in automated warfare. *Philosophy & Technology, 28*(1), 125-137.

Coeckelbergh, M. (2013). Drones, information technology, and distance: mapping the moral epistemology of remote fighting. *Ethics and Information Technology*, *15*(2), 87-98.

Crootof, R. (2015). War, responsibility, and killer robots. *North Carolina Journal of International Law and Commercial Regulation*, *40*(4), 909-932.

Crootof, R. (2016). A meaningful floor for ‘meaningful human control’. *Temple International & Comparative Law Journal*, *30*, 53-62.

Cummings, M.L. (2006). Automation and accountability in decision support system interface design. *Journal of Technology Studies*, *32*(1), 23-31.

Detert, J.R., Treviño, L.K., & Sweitzer, V.L. (2008). Moral disengagement in ethical decision making: A study of antecedents and outcomes. *Journal of Applied Psychology*, *93*(2), 374-391.

Donnelly, S.B. (2005). Long-distance warriors. *Time Magazine*, 4 December.

Dunlap, Ch.J., Jr. (forthcoming). Accountability and autonomous weapons: Much ado about nothing, *Temple International & Comparative Law Journal*.

Ekelhof, M.A.C. (forthcoming). Complications of a common language: Why it is so hard to talk about autonomous weapons. *Journal of Conflict and Security Law*.

Fielding, M. (2006). Robotics in future land warfare. *Australian Army Journal*, *3*(2), 99-108.

Fieser, J., & Dowden, B. (2007). Just war theory. *The internet encyclopedia of philosophy*. Retrieved June 26, 2017, from http//www.iep.utm.edu/j/justwar.htm.

Fischer, J.M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.

Garraway, C. (2009). The doctrine of command responsibility. In J. Doria, H.-P. Gasser, & M. C. Bassiouni (Eds.), *The legal regime of the international criminal court* (pp. 703-725). Leiden: Nijhoff.

Hellström, T. (2013). On the moral responsibility of military robots. *Ethics and Information Technology*, *15*(2), 99-107.

Henckaerts, J.-M., & Doswald-Beck, L. (2009). *Customary international humanitarian law. Volume I: Rules.* Cambridge: ICRC and Cambridge University Press.

Human Right Council (2013). *Report of the special rapporteur on extrajudicial,summary or arbitrary executions, Christof Heyns* (A/HRC/23/47), United Nations.

Isenberg, D. (2007). Robots replace trigger fingers in Iraq. *Asia Times Online*. Retrieved June 26, 2017, from http://www.atimes.com/atimes/Middle\_East/IH29Ak01.html.

International Organization for Standardization (2011). *Robots for industrial environments - Safety requirements*. Retrieved June 26, 2017, from http://www.iso.org/iso/iso\_catalogue/catalogue\_tc/catalogue\_detail.htm?csnumber=36322.

Johnson, A.M., & Axinn, S. (2013). The morality of autonomous robots. *Journal of Military Ethics*, *12*(2), 129-141.

Jordan, B. (2007). Half of predators fielded have been lost. *Air Force Times*. Retrieved June 26, 2017, from http://www.airforcetimes.com/news/2007/02/AFpredatorlosses070223.

Kaag, J., & Kaufman. W. (2009). Military frameworks: technological know-how and the legitimization of warfare. *Cambridge Review of International Affairs*, *22*(4), 585-606.

Kaplan, R.D. (2006). Hunting the taliban in Las Vegas. *Atlantic Monthly*, *September*, 81-84.

Kershnar, S. (2013). No moral problem. In B. J. Strawser (Ed.), *Killing by remote control: The ethics of an unnmanned military*.Oxford: Oxford University Press.

Krishnan, A. (2009). *Killer robots. Legality and ethicality of autonomous weapons*. Farnham: Ashgate Publishing Limited.

Leveringhaus, A. (2016). Drones, automated targeting and moral responsibility. In E.D. Nucci, & F. Santoni de Sio (Eds.), *Drones and responsibility. Legal, philosophical and socio-technical perspectives on remotely controlled weapons* (pp. 169-181). London: Routledge.

Lee, P. (2012). Remoteness, risk and aircrew Ethos. *Air Power Review*, *15*(1), 1-19.

Lichocki, P., Kahn Jr, P., & Billard, A. (2011). The Ethical Landscape of Robotics. *IEEE Robotics and Automation Magazine*, *18*(1), 39-50.

Lin, P., Bekey, G., & Abney, K. (2008). *Autonomous military robotics: Risk, ethics, and design*. San Luis Obispo: California Polytechnic State University.

Lucas, G. (2011). Industrial challenges of military robotics. Paper presented at the conference *The Ethics of Emerging Military Technologies* of International Society for Military Ethics (ISME), San Diego, January 2011. (http://isme.tamu.edu/ISME11/isme11.html)

Magnuson, S. (2008). Armed robots sidelined in Iraq. *National Defense Magazine* (May).

Mastroianni, G. R. (2011). The person–situation debate: Implications for military leadership and civilian–military relations. *Journal of Military Ethics*, *10*(1), 2-16.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, *6*(3), 175-183.

Milgram, S. (1974). *Obedience to authority: An experimental view.* London: Tavistock.

Moller, A. C., & Deci, E.L.(2010). Interpersonal control, dehumanization, and violence: A self-determination theory perspective. *Group Processes and Intergroup Relations*, *13*(1), 41-53.

Mukerji, N. (2016). Autonomous killer drones. In E.D. Nucci, & F. Santoni de Sio (Eds.), *Drones and responsibility. Legal, philosophical and socio-technical perspectives on remotely controlled weapons* (pp. 197-214 ). London: Routledge.

Müller, V.C. (2016). Autonomous killer robots are probably good news. In E.D. Nucci, & F. Santoni de Sio (Eds.), *Drones and responsibility. Legal, philosophical and socio-technical perspectives on remotely controlled weapons* (pp. 67-81 ). London: Routledge.

National Research Council (2005). *Autonomous vehicles in support of naval operations*. Washington: The National Academies Press.

Noorman, M. (2014). Responsibility practices and unmanned military technologies. *Science and Engineering Ethics*, *20*(3), 809-826.

Noorman, M., & Johnson, D.G. (2014). Negotiating autonomy and responsibility in military robots. *Ethics and Information Technology*, *16*(1), 51-62.

Olsthoorn, P., & Royakkers, L.M.M. (2011). Risks and robots – Some ethical issues. Paper presented at the conference *The Ethics of Emerging Military Technologies* of International Society for Military Ethics (ISME), San Diego, January 2011. (http://isme.tamu.edu/ISME11/isme11.html)

Orend, B. (2006). *The morality of war*. Orchard Park, N.Y.: Broadview Press.

Raffensperger, C., & Tickner, J. (Eds.) (1999). *Protecting public health and the environment: Implementing the precautionary principle*. Washington, DC: Island Press.

Royakkers, L.M.M., & Topolski, A.R. (2014). Military robotics & relationality: criteria for ethical decision-making. In J. van den Hoven, N. Doorn, T. Swierstra, B.-J. Koops & H. Romijn (Eds.), *Responsible innovation 1. Innovative solutions for global issues* (pp. 351-367). Dordrecht: Springer.

Royakkers, L.M.M., & Van Est, Q. (2010). The cubicle warrior: The marionette of digitalized warfare, *Ethics and Information Technology* *12*(3), 289-296.

Schulzke, M. (2013). Autonomous weapons and distributed responsibility. *Philosophy and Technology*, *26*(2), 203-219.

Schwarz, E. (2017). *Death machines: The ethics of violent technologies*. Manchester University Press (forthcoming).

Shachtman, N. (2007). Robot cannon kills 9, wounds 14. *Wired.com* (18 October).

Sharkey, N. (2008). Cassandra or false prophet of doom: AI robots and war. *IEEE Intelligent Systems* *23*(4), 14-17.

Sharkey, N. (2010). Saying ‘no!’ to lethal autonomous targeting, *Journal of Military Ethics*, *9*(4), 369-383.

Simpson, T.W., & Müller, V.C. (2016). Just war and robots’ killings. *The Philosophical Quarterly*, *66*(263), 302-322.

Singer, P. W. (2009). *Wired for war: The robotics revolution and conflict in the twenty-first century.* New York: Penguin Books.

Slim, H. (2007). *Killing civilians: Method, madness and morality in war*. London: Hurst & Company.

Sparrow, R. (2007). Killer robots. *Journal of Applied Philosophy*. *24*(1), 62-77.

Sparrow R. (2009). Building a better warbot: ethical issues in the design of unmanned systems for military applications. *Science and Engineering Ethics*, *15*(2), 169-87.

Sparrow R. (2011). Robotic weapons and the future of war. In P. Tripodi, & J. Wolfendale (Eds.), *New wars and new soldiers: Military ethics in the contemporary world* (pp. 117-133). Farnham Surrey, UK: Ashgate.

Sullins, J. (2010). RoboWarfare: can robots be more ethical than humans on the battlefield? *Ethics and Information Technology*, *12*(3), 263-75.

Thompson, M. (2007).V-22 Osprey: A flying shame. *Time*. Retrieved June 26, 2017, from [http://www.time.com/time/politics/article/0,8599,1665835,00.html](http://www.time.com/time/politics/article/0%2C8599%2C1665835%2C00.html).

US Department of Defense (2013). *Unmanned systems integrated roadmap 2013-2038*. Washington: Government Printing Office.

Van de Poel, I.R., & Royakkers, L.M.M. (2011). *Ethics,* e*ngineering and technology*. Oxford: Blackwell.

Van de Poel, I.R., Royakkers, L.M.M, & Zwart, S.D. (2015). *Moral responsibility and the problem of many hands*. New York: Routledge.

Verbeek, P.P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. University Park: Pennsylvania State University Press.

Verbeek, P.P. (2009). Moralizing technology: On the morality of technological artifacts and their design. In D.M. Kaplan (Ed.), *Readings in the philosophy of technology* (pp. 226-242). Plymouth: Rowman & Littlefield Publishers.

Vicente, J.P. (2013). The American way of remote air warfare. *Journal of Military Studies*, *4*(1).

Welsh, A. (2008). *What is Honor? A Question of Moral Imperatives*. New Haven: Yale University Press.

**ENDNOTES**

1. As the authors of a report on autonomous military robots, written for the US Navy, put it: unmanned systems are ‘unaffected by the emotions, adrenaline, and stress that cause soldiers to overreact or deliberately overstep the Rules of Engagement and commit atrocities, that is to say, war crimes’ (Lin, Bekey, and Abney 2008, 1). [↑](#endnote-ref-1)
2. There are more criteria to hold someone reasonably responsible, such as a *causal relationship* between the behavior and the consequences, but in this paper we focus on *control.* [↑](#endnote-ref-2)
3. For the statement see www.icrac.co.uk/Expert%20Workshop%20Statement.pdf. [↑](#endnote-ref-3)
4. https://futureoflife.org/open-letter-autonomous-weapons/ (retrieved July 3, 2017). [↑](#endnote-ref-4)
5. A footnote made clear to what kind of decisions this principle should apply to ‘The decision to kill or use lethal force against a human being; The decision to use injurious or incapacitating force against a human being; The decision to initiate combat or violent engagement between military units; The decision to initiate war or warfare between states or against non-state actors.’ The statement can be found at <http://icrac.net/statements/> (retrieved July 4, 2017) [↑](#endnote-ref-5)
6. This principle originates from the Rio Declaration, the closing statement of the first conference of the United Nations on sustainable development, which was held in Rio de Janeiro in 1992. [↑](#endnote-ref-6)
7. According to Coeckelbergh (2013), the practice of using drones is illustrative of the claim that new technological practices that aim to bridge physical distance create more moral distance and make it difficult for people to exercise moral responsibility. [↑](#endnote-ref-7)