(This is the penultimate draft of the version published in 2021. It was previously listed on Philpapers as “Fire and Forget: A Defense of the Use of Autonomous Weapons in War”)

**Fire and Forget: A Moral Defense of the Use of Autonomous Weapons Systems in War and Peace**[[1]](#footnote-1)

Duncan MacIntosh

**Introduction**

While Autonomous Weapons Systems—AWS—have obvious military advantages, there are prima facie moral objections to using them. I have elsewhere argued (MacIntosh 2016) that there are similarities between the structure of law and morality on the one hand and of automata on the other, and that this plus the fact that automata can be designed to lack the biases and other failings of humans, require us to automate the administration and enforcement of law as much as possible.

But in this paper I want to argue more specifically (and contra Mary Ellen O’Connell 2014, Christof Heyns 2013, Peter Asaro 2016, and others) that there are many conditions where using AWSs would be strategically, morally and rationally appropriate. This will occupy Part I of the paper. In Part II I deal with the objection that use of robots is inherently wrong or violating of human dignity.[[2]](#footnote-2)

**Part I: Occasions of the Ethical Use of Autonomous Fire and Forget Weapons**

An AWS would be a “fire-and-forget” weapon. And some see such weapons as legally and morally problematic. For surely a human and human judgment should figure at every point in a weapon’s operation, especially where it is about to have its lethal effect on a human. After all, as O’Connell (2014) argues, that is the last reconsideration moment, and arguably to fail to have a human doing the deciding at that point is to abdicate moral and legal responsibility for the kill. (Think of the final phone call to the Governor to see if the Governor will stay an execution.) Asaro (2016) argues that it is part of law, including International Humanitarian Law, to respect public morality even if it has not yet been encoded into law, and that part of such morality is the expectation that there be meaningful human control of weapons systems, so that this requirement should be formally encoded into law. In addition to there being a public morality requirement of meaningful human control, Asaro suspects that the dignity of persons liable to being killed likewise requires that their death, if they are to die, be brought about by a human, not a robot.

The positions of O’Connell and Asaro have an initial plausibility, but they have not been argued for in depth, it is unclear what does or could premise them, and it is doubtful, I think, whether they will withstand examination.[[3]](#footnote-3) For example, I think it will prove false that there must always be meaningful human control in the infliction of death. For given a choice between control by a morally bad human who would kill someone underserving of being killed and a morally good robot who would kill only someone deserving of being killed, we would pick the good robot. What matters is not that there be meaningful human control, but that there be meaningful moral control, that is, that what happens be under the control of morality, that it be the right thing to happen. And similar factors complicate the dignity issue—what dignity is, what sort of agent best implements dignity, and when the importance of dignity is an overridden factor, all come into play. So let us investigate more closely.

Clarity requires breaking this issue down into three sub-issues, namely whether, when an autonomous weapon (an AWS) has followed its program and is now poised to kill, i) there should always be reconsideration of its decision at least in the sense of revisiting whether the weapon should be allowed to kill, and then, in any given case, whether ii) there should be reconsideration in the sense of reversing the decision to kill. And if there is to be either or both, iii) what sort of agent should do the reconsidering, the AWS or a human being?

It might be thought that there should always be reconsideration by a human in at least the revisiting sense, if not necessarily the reversing. For what could it cost? And it might save us from making a moral mistake.

But there are several situations where reconsideration would be inappropriate. In what follows, I assume that the agent deciding whether to use a fire and forget weapon is a rational agent with all-things-considered morally approvable goals seeking therefore to maximize moral expected utility. That is, in choosing among actions, she is disposed to do that action which makes as high as possible the sum of the products of the moral desirability of possible outcomes of actions and the probability of those outcomes obtaining given the doing of the various actions available. She will have considered the likelihood of the weapon’s having morally good effects given its design and proposed circumstance of use. If the context is a war context, she would bear in mind whether use of the weapon is likely to respect such things as International Humanitarian Law and the Laws of War. So she would be seeking to respect the principles of distinctness, necessity and proportionality. Distinctness is the principle that combatants should be targeted before civilians; necessity, the principle that violence should be used only to attain important military objectives; and proportionality is the principle that the violence used to attain the objective should not be out of proportion to the value of the objective. More generally, I shall assume that the person considering using an AWS would bear in mind whether the weapon can likely distinguish, or can be deployed in such a way as to respect the distinction between those morally liable to being harmed (that is, those whom it is morally permissible or obligatory to harm) and those who are to be protected from harm. She would proceed on the best information available at the time of considering activation of the weapon.

**Among the situations in which activating a fire and forget weapon by such an agent would be rationally and morally legitimate would be these**:

**a) Planning Scenarios**: one initially best guesses that it is at the moment of firing the weapon (e.g., activating the robot) that one has greatest informational and moral clarity about what needs to be done, estimating that to reconsider would be to open oneself to fog of war confusion, or to temptations one judges at the time of weapon activation that it would be best to resist at the moment of possible recall. So one forms the plan to activate the weapon and let it do its job, then follows through on the plan by activating and then not recalling the weapon, even as one faces temptations to reconsider, reminding one’s self that one was probably earlier better placed to work out how best to proceed back when one formed the plan.[[4]](#footnote-4)

**b) Short Term Versus Long Term Consequences Cases**: one initially best judges that one must not reconsider if one is to attain the desired effect of the weapon. Think of the decision to bomb Nagasaki and Hiroshima in hopes of saving by means of the deterrent effect of the bombing more lives than those lost from the bombing, this in spite of the horror that must be felt at the immediate prospect of the bombing.[[5]](#footnote-5) Here one should not radio the planes and call off the mission.

**c) Resolute Choice Cases**: one expects moral benefit to accrue not from allowing the weapon to finish its task, but from the consequence of committing to its un-reconsidered use should the enemy not meet some demand.[[6]](#footnote-6) The consequence sought will be available only if one can be predicted not to reconsider; and refraining from reconsidering is made rational by the initial expected benefit and so rationality of committing not to reconsider. Here, if the enemy does not oblige, one activates the weapon and lets it finish.

It may be confusing what distinguishes these first three rationales. Here is the distinction: the reason one does not reconsider in a) is because one assumes one knew best what to do when forming the plan that required non-reconsidering; in b) because one sees that the long-term consequences of not reconsidering exceed those of reconsidering; and in c) because non-reconsideration expresses a strategy for making choices whose adoption was expected to have one do better, even if following through on it would not, and morality and rationality require one to make the choices dictated by the best strategy—one decides the appropriateness of actions by the advantages of the strategies that dictate them, not by the advantages of the actions themselves. Otherwise, one could not have the advantages of strategies.

This last rationale is widely contested. After all, since the point of the strategy was, say, deterrence, and deterrence has failed so that one must now fulfill a threat one never really wanted to have to fulfill, why still act from a strategy one now knows was a failure? To preserve one’s credibility in later threat scenarios? But suppose there will be none, as is likely in the case of, for example, the threat of nuclear apocalypse. Then again, why fulfill the threat? By way of addressing this, I have favoured a variant on the foregoing rationale: in adopting a strategy, one changes in what it is that one sees as the desired outcome of actions, then refrains from reconsidering because refraining now best expresses one’s new desires—one has come to care more about implementing the strategy, or about the expected outcome of implementing it, than about what first motivated one to adopt the strategy. So one does not experience acting on the strategy as going against what one cares about.[[7]](#footnote-7)

**d) Unreconsiderable Weapons Cases**: one’s weapon is such that, while deploying it would be expected to maximize moral utility, reconsidering it at its point of lethality would be impossible so that, if a condition on the permissible use of the weapon were to require reconsideration at that point, one could never use the weapon. (E.g., one cannot stop a bullet at the skin and re-think whether to let it penetrate, so one would have to never use a gun.) A variant on this case would be the case of a weapon that could be made able to be monitored and recalled as it engages in its mission, but giving it this feature would put it at risk of being hacked and used for evil. For to recall the device would require that it be in touch by, say, radio, and so liable to being communicated with by the enemy. Again, if the mission has high moral expected utility as it stands, one would not want to lower this by converting the weapon into something recallable and therefore able to be perverted. (This point has been made by many authors.)

By hypothesis, being disposed to reconsider in a)-d) would have lower moral expected utility than not. And so being disposed to reconsider would nullify any advantage the weapon afforded. No, in these situations one should deliberate as long as is needed to make an informed decision given the pressure of time. Then one should activate the weapon.

Of course in all those scenarios one could discover part way through that the facts are not what one first thought, so that the payoffs of activating and not reconsidering are different. This might mean that one would learn it was a mistake to activate the weapon, and should now reconsider and perhaps abort. So of course it can be morally and rationally obligatory to stay sensitive to these possibilities. (Although it is a moot point in case d), since there, recalling the weapon is impossible. If the weapon will take a long time to impact, however, it might become rational and morally obligatory to warn the target if one has a communication signal that can travel faster than the speed of one’s kinetic weapon.)

Let us be clear, however, about which possibilities are morally and rationally relevant to deciding to recall a weapon. Suppose one rationally commits to using a weapon and also to not reconsidering even though one knows at the time of commitment that one’s compassion would tempt one to call it off later. Since this was considered at the outset, it would not be appropriate to reconsider on that ground just before the weapon’s moment of lethality.

Now suppose instead that it was predictable that there would be a certain level of horror from use of the weapon, but one then discovers that the horror will be much worse, e.g., that many more people will die than one had predicted. That, of course, would be a basis for reconsideration.

But several philosophers, including Martha Nussbaum, in effect think as follows (Nussbaum 1993, especially pp. 83-92): every action is both a consequence of a decision taking into account moral factors, and a learning moment where one may get new information about moral factors. Perhaps one forms a plan to kill someone, thinking justice requires this, then finds one cannot face actually doing the deed, and decides that justice requires something different, mercy perhaps, as Nussbaum suggests—one comes to find the originally intended deed more horrible, not because it will involve more deaths than one thought, but because one has come to think that any death is more horrible than one first thought. Surely putting an autonomous robot in the loop here would deprive one of the possibilities of new moral learning?

It is true that some actions can be learning occasions, and either we should not automate those actions so extremely as to make the weapons unrecallable, or we should figure out how to have our automata likewise learn from the experience and adjust their behaviours accordingly, perhaps self-aborting.

But some actions can reasonably be expected not to be moral learning occasions. In these cases we have evidence of there being no need to build in the possibility of moral experiencing and reconsideration. Perhaps one already knows the horror of killing someone, for example. (There is, of course, always the logical possibility that the situation is morally new. But that is different from having actual evidence in advance that the situation is new; and the mere possibility by itself is no reason to forego the benefits of a disposition to non-reconsideration. Indeed, if that were a reason, one could never act, for upon making any decision one would have to reconsider in light of the mere logical possibility that one’s decision was wrong.)

Moreover, there are other ways to get a moral learning experience about a certain kind of action or its consequence than by building a moment of possible experience and reconsideration into the action. For example, one could reflect after the fact, survey the scene, do interviews with witnesses and relatives of those affected, study film of the event, and so on, in this way getting the original expected benefit of the weapon, but also gaining new information for future decisions. This would be appropriate where one calculates that there would be greater over-all moral benefit to using the weapon in this case and then revisiting the ethics of the matter, rather than the other way around, because one calculates that one is at risk of being excessively squeamish until the mission is over, and that this would prevent one from doing a morally required thing.

There is also the possibility that not only will one not expect to get more morally relevant experience from the event, but one may expect to be harmed in one’s moral perspective by it. Thus we have…

**e) Protection of One’s Moral Self Cases**: suppose there simply must be some people killed to save many people—there is no question that his is ethically required. But suppose too that if a human were to do the killing, she would be left traumatized in a way constituting a moral harm to her. For example, she would have crippling PTSD and a tendency towards suicidality. Or perhaps the experience would leave her coarsened in a way making her more likely to do evil in the future. Either way, it would then be harder down the road for her to fulfill her moral duties to others and to herself. Here, it would be morally and rationally better that an AWS do the killing—the morally hard but necessary task gets done, but the agent has her moral agency protected. Indeed, even now there are situations where, while there is a human in the decision loop, the role the human is playing is defined so algorithmically that she has no real decision making power. Her role could be played by a machine. And yet her presence in the role means that she will have the guilt of making hard choices resulting in deaths, something that will be a burden on her conscience even where it is the right choice. So, again, why not just spare her conscience and take her out of the loop?

It is worth noting that there are a numbers of ways of getting her out of the loop, and a number of degrees to which she could be out. She could make the decision that someone will have to die, but a machine might implement the decision for her. This would be her being out of the loop by means of delegating implementation of her decision to an AWS. An even greater degree of removal from the loop might be where a human delegates the very decision of whether someone has to die to a machine, one with a program so sophisticated that it is in effect a morally autonomous agent. Here the hope would be that the machine can make the morally hard choices, and that it will make morally right choices, but that it will not have the pangs of conscience that would be so unbearable for a human being.

There is already a precedent for this in military contexts where a commander delegates decisions about life and death to an autonomous human with his own detailed criteria for when to kill, so that the commander cannot really say in advance who is going to be killed, how, or when. This is routine in military practice and part of the chain of command and the delegation of responsibility to those most appropriately bearing it—detailed decisions implementing larger strategic policy have to be left to those closest to battle.

Some people might see this as a form of immorality. Is it really OK for a commander to have a less troubled conscience by virtue of having delegated morally difficult decisions to a subordinate? But I think this can be defended, not only on grounds of this being militarily necessary—there really is no better way of war-fighting—but on grounds, again, of distributing the costs of conscience: commanders need to make decisions that will result in loss of lives over and over again, and can only escape moral fatigue if they do not have to further make the detailed decisions about whom exactly to kill and when.

And if these decisions are delegated to a morally discerning but morally conscienceless machine, we have the additional virtue that the moral offloading—the offloading of morally difficult decisions—is done onto a device that will not be morally harmed by the decisions it must make.[[8]](#footnote-8), [[9]](#footnote-9)

There are also…

**f) Morally Required Diffusion of Responsibility Cases**, e.g., cases of a firing squad sort where many people are involved in performing the execution so that there is ambiguity about who had the fatal effect in order to spare the conscience of each squad member. But again, this requires that one not avail one’s self of opportunities to recall the weapon. Translated to robotic warfare, imagine the squad is a group of drone operators all of whom launch their individual AWS drones at a target, and who, if given the means to monitor the progress of their drone and the authority to recall it if they judged this for the best, could figure out pre-impact whose drone is most likely to be the fatal one. This might be better not found out, for it may result in a regress of yank-backs, each operator recalling his drone as it is discovered to be the one most likely to be fatal, with the job left undone; or it getting done by the last person who clues in too late, him then facing the guilt alone; or it getting done by one of the operators deliberately continuing even knowing his will be the fatal drone, but who then, again, must face the crisis of conscience alone.

Yet another case where it would be appropriate to delegate a killing to AWS would be…

**g) Morally Better For Being Comparatively Random and Non-Deliberate Killings Cases.** These are cases where the killing would be less morally problematic the more random and free of deliberate intention each aspect of the killing was. What is morally worse, throwing a grenade into a room of a small number of people who must be stopped to save a large number of people, or moving around the room at super speed with a sack full of shrapnel, pushing pieces of shrapnel into people’s bodies—you have to use all the pieces to stop everyone, but the pieces are of different sizes, some so large that using them will kill, others only maim, yet others, only temporarily injure, and you have to decide which piece goes into which person. The effect is the same—it is as if a blast kills some, maims others, and leaves yet others only temporarily harmed. But the second method is morally worse. Better to delegate to an AWS. Sometimes, of course, the circumstance might permit the use of a very stupid machine, e.g., in the case of an enclosed space, literally a hand grenade, which will produce a blast whose effect on a given person is determined by what is in effect a lottery. But perhaps a similar effect needs to be attained over a large and open area, and, given limited information about the targets and the urgency of the task, the effect is best achieved by using an AWS that will attack targets of opportunity with grenade-like weapons. Here it is the delegating to an AWS, plus very randomness of the method of grenade, plus the fact that only one morally possibly questionable decision need be made in using the weapon—the decision to delegate—that makes it a morally less bad event. Robots can randomize and so democratize violence, and so make it less bad, less inhumane, less monstrous, less evil.

Of course other times the reverse judgment would hold. In the preceding examples I in effect assumed everyone in the room, or in the larger field, was morally equal as a target with no one more or less properly morally liable to be killed, so that, if one chose person by person whom to kill, one would choose on morally arbitrary and therefore problematic, morally agonizing grounds. But in a variant case, imagine one knows this man is a father, that man, a psychopath, this other man, unlikely to harm anyone in the future. Here, careful individual targeting decisions are called for—you definitely kill the psychopath, but harm the others in lesser ways just to get them out of the war.

Next,

**h) Doomsday Machine Cases**: sometimes what is called for is precisely a weapon that cannot be recalled—this would be its great virtue. The weapons in mutually assured destruction are like this—they will activate on provocation no matter what, and so are the supreme deterrent. This reduces to the case of someone’s being morally and rationally required to be resolute in fulfilling a morally and rationally recommended threat (item c), above) if we see the resolute agent as a human implementation of a Doomsday Machine. And if we doubted the rationality or morality of a free agent fulfilling a threat morally maximizing to make but not to keep, arguably we could use the automation of the keeping of the threat to ensure its credibility; for arguably it can be rational and moral to arrange the doing of things one could not rationally or morally do one’s self. (This is not case d), where we use an unrecallable weapon because it is the only weapon we have and we must use some weapon or other. In the present case, only an unrecallable weapon can work, because of its effectiveness in threatening.)

**i) Permissible Threats of Impermissible Harms Cases:** these are related to the former cases. Imagine there is a weapon with such horrible and indiscriminate power that it could not be actually used in ways compatible with International Humanitarian Law and the Laws of War, which require that weapons use respect distinctness, necessity and proportionality, and must not render large regions of the planet uninhabitable for long periods. Even given this, arguably the threat of its use would be permissible both morally and by the foregoing measures provided issuing the threat was likely to have very good effects, and provided the very issuing of the threat makes the necessity of following through fantastically unlikely. The weapon’s use would be so horrible that the threat of its use is almost certain to deter the behavior against which it is a threat. But even if this is a good argument for making such a threat, arguably the threat is permissible only if the weapon is extremely unlikely to be accidentally activated, used corruptly, or misused through human error. And it could be that, given the complexity of the information that would need to be processed to decide whether a given situation was the one for which the weapon was designed, given the speed with which the decision would have to be made, and given the potential for the weapon to be abused were it under human control, it ought instead to be put under the control of an enormously sophisticated artificial intelligence.

Obviously the real world case of nuclear weapons is apposite here. Jules Zacher (2016) has suggested that such weapons cannot be used in ways respecting the strictures of international humanitarian law and the law of war, not even if their control is deputized to an AWS. For again, their actual use would be too monstrous. But I suggest it may yet be able to be right to threaten to do something it would be wrong to actually do, a famous paradoxes of deterrence identified by Gregory Kavka (1978). Arguably we’ve been living in this scenario for seventy years: most people think that massive nuclear retaliation against attack would be immoral. But many think the threat of it has saved the world from further world wars, and is therefore morally defensible.

Let us move on. We have been discussing situations where one best guesses in advance that certain kinds of reconsideration would be inappropriate. But now to the question what should do the deciding at the final possible moment of reconsideration when it can be expected that reconsideration in either of our two senses is appropriate. Let us suppose we have a case where there should be continual reconsideration sensitive to certain factors. Surely this should be done by a human? But I suggest it matters less what makes the call, more that it be the right call. And because of all the usual advantages of robots—their speed, inexhaustibility, etc. —we may want the call to be made by a robot, but one able to detect changes in the moral situation and to adjust its behaviours accordingly. This suggests yet another specific case where it would be better to have humans out of the loop…

**j) Robot Training Cases**: we are trying to train a robot to make better moral decisions, and the press of events has forced us to beta test it in live battle. The expected moral utility of letting the robot learn may exceed that of affording an opportunity for a human to acquire or express a scruple by putting the human in a reconsideration loop. For once the robot learns to make good moral decisions, we can replicate its moral circuit in other robots, with the result of having better moral decisions made in many future contexts.

Here are some further cases and rationales for using autonomous weapons systems:

**k) Precision in Killing Cases**: sometimes, due to the situations the device is to be used in, or due to the advanced design of the device, an AWS may provide greater precision in respecting the distinction between those morally liable and not liable to being killed—something that would be put at risk by the reconsideration of a clumsy human operator (Arkin 2013). An example of the former would be a device tasked to kill anything in a region known to contain only enemies who need killing—there are no civilians in the region who stand at risk, and none of the enemies in the region deserve to survive. Here the AWS might be more thorough than a human. Think of an AWS defending an air-craft carrier, tasked with shooting anything out of the sky that shows up on radar, prioritizing things large in size, moving at great speed, that are very close, and that do not self-identify with a civilian transponder response when queried. Nothing needs to be over an aircraft carrier and anything there is an enemy. An example of the second—of an AWS being more precise than a human by virtue of its design—might be where the AWS is better at detecting the enemy than a human, e.g., by means of metal detectors able to tell who is carrying a weapon and is therefore a genuine threat. Again, all and only those needing killing get killed.

**l) Speed and Efficiency Cases**: use of an AWS may be justified by its being vastly more efficient in a way that, again, would be jeopardized by less efficient human intervention (Arkin 2013)—if the weapon had to pause while the operator approved each proposed action, the machine would have to go more slowly, and fewer of the right people would be killed, fewer of the right people, spared.

The foregoing, then, are cases where we would not want a human operator “in the loop”, that is, a human playing the role of giving final approval to each machine decision to kill, so that the machine will not kill unless authorized by a human in each kill. This would merely result in morally inferior outcomes. Neither would we want a human “on the loop”, where the machine will kill unless vetoed, but where the machine’s killing process is slowed down to give a human operator a moment to decide whether to veto. For again, we would have morally inferior outcomes.

Other cases involve factors often used in arguments against AWSs. So on to…

**Part III: Objections From the Supposed Indignity of Robot-Inflicted Death**

Some think death by robot is inherently worse than death by human hand, that it is somehow inherently more bad, wrong, undignified, or fails in a special way to respect the rights of persons—it is wrong in itself, mala in se as the phrase used by Wendell Wallach (2013) in this connection has it.

I doubt this, but even if it were true, that would not decide the matter. For something can be bad in itself without being such that it should never be incurred or inflicted. Pain is always bad in and of itself. But that does not mean you should never incur it—maybe you must grab a hot metal door knob to escape a burning building, and that will hurt, but you should still do it. Maybe you will have to inflict a painful injury on someone to protect yourself in self-defence, but that does not mean you must not do it. Similarly, even if death by robot were an inherent wrong, that does not mean you should never inflict or be subject to it. For sometimes it is the lesser evil, or is the means to a good thing outweighing the inherent badness of the means.

Here are cases which show either that death by robot is not inherently problematic, or that, even if it is, it could still be morally called for. One guide is how people would answer certain questions. So,

**Dignity Case 1; Saving Your Village by Robotically Killing Your Enemy**: your village is about to be over-run by ISIL; your only defense is the auto-sentry. Surely you would want to activate it? And surely this would be right, even if it metes out undignified robot death to your attackers? Or consider…

**Dignity Case 2; Killing Yourself By Robot To Save Yourself From A Worse Death From a Man**: you are about to be captured and killed; you have the choice of quick death by Western robot (a suicide machine available when the battle is lost and you face capture), or slow beheading by a Jihadist. Surely you would prefer death by robot? (It will follow your command to kill you where you could not make yourself kill yourself. Or it might be pre-programmed to be able to consider all factors and enabled to decide to kill you quickly and painlessly should it detect that all hope is lost.) A person might prefer death by the AWS robot for any of several reasons. One is that an AWS may afford a greater dignity to the person to be killed precisely by virtue of its isolation from human control. In some cases, it seems worse to die at human than at robot hands. For if it is a human who is killing you, you might experience not only the horror of your pending death, but also anguish at the fact that, even though they could take pity on you and spare you, they will not—they are immune to your pleading and suffering. I can imagine this being an additional harm. But with a machine, one realizes there is nothing personal about it, there is no point in struggle or pleading, there is no one in whose gaze you are seen with contempt or as being unworthy of mercy. It is more like facing death by geological forces in a natural disaster, and more bearable for that fact. (Other cases might go the other way, of course. I might want to be killed gently, carefully and painlessly by a loving spouse trying to give me a good death, preferring this to death by impersonal euthanasia machine.)

If you have trouble accepting that robot inflicted death can be OK, think about robot conferred benefits and then ask why, if these are OK, their opposite cannot be. Would you insist on benefits being conferred to you by a human rather than a robot? Suppose you can die of thirst or drink from a palette of water bottles parachuted to you by a supply drone programmed to provide drink to those in the hottest part of the desert. You would take the drink, not scrupling about there being any possible indignity in being targeted for help by a machine. Why should it be any different when it comes to being harmed? Perhaps you want the right to try to talk your way out of whatever supposed justice the machine is to impose upon you. Well, a suitably programmed machine might give you a listen, or set you aside for further human consideration; or it might just kill you. And in these respects, matters are no different than if you faced a human killer.

And anyway, the person being killed is not the only person whose value or dignity is in play. There is also what would give dignity to that person’s victims, and to anyone who must be involved in person’s killing. On the former, we then have….

**Dignity Case 3; Robotic Avenging of the Dignity of a Victim**: maybe the dignity of the victim of a killer (or of the victim’s family) requires the killer’s death, and the only way to get the killer is by robot.

**Dignity Case 4; Robotic Killing to Save the Dignity of a Human Executioner**: maybe those who inflict monstrosity forego any rights to dignified human-inflicted death (if that is in fact especially dignified), either because denying them this is a fit penalty, or because of the moral and psychological cost, and perhaps the indignity, that would have to be borne by a decent person in executing an indecent person. Better a robot death so no human executioners have to soil their hands. And note for whom we have of late been reserving robotic death, as in automated drone killing, or death by indiscriminate weapon, e.g., a non-smart bomb, namely, people who would inflict automated or indiscriminate killing on us (e.g., by a bomb in a café), terrorists whose modus operandi is to select us randomly for death, rather than by means of specific proper liability to death.

Moreover, dignity is a luxury. And sometimes luxury must yield to factors of greater exigency.[[10]](#footnote-10)

Some of this, of course, is separate from what people perceive as being required by dignity, and from how important they think dignity is; and if we are trying to win not just the war but also the peace, maybe we will do better if we respect a culture’s conception of dignity in how we fight them; and this may, as a purely practical matter, require us not to inflict death robotically.

This might even rise to the level of principle if there is a moral imperative to respect the spiritual opinions even of wrong-headed adversaries, an imperative not to unnecessarily trample on those opinions. Maybe we even have a moral duty to take some personal risks in this regard, and so to eschew the personal safety that use of robots would afford.[[11]](#footnote-11)

**Conclusion**

Summing up my argument, it appears that it is false that it is always best for a human decision to be proximal to the application of lethal force. Instead, sometimes remoteness in distance and time, remoteness from information, and remoteness from the factors that would result in specious reconsideration, should rule the day.

It is not true that fire and forget weapons are evil for not having a human at the final point of infliction of harm. They are problematic only if they inflict a harm that proper reconsideration would have demanded not be inflicted. But one can guestimate at the start whether a reconsideration would be appropriate. And if one’s best guess is that it would not be appropriate, then one’s best guess can rightly be that one should activate the fire and forget weapon. At that point, the difference between a weapon that impacts seconds after the initial decision to use it, and a weapon that impacts hours, days, or years after, is merely one of degree. In fact, this suggests yet another pretext for the use of AWS, namely, its being the only way to cover off the requirements of infrastructure protection. Here is a case, which I present as a kind of…

**Coda**

**m) We are low on manpower and deputizing to an AWS is the only way of protecting a remote power installation; (Justifiable Landmines Cases)**:

Landmines are often cited as a counter-example to this way of thinking. But the problem with landmines is not that they do not have a human running the final part of their action, but that they are precisely devices reconsideration of whose use becomes appropriate at the very least at the cessation of hostilities, and perhaps before. The mistake is deploying them without a deactivation point or plan even though it is predictable that this will be morally required. But there is no mistake in having them be fire and forget before then. Especially not if they are either well-designed only to harm the enemy, or their situation makes it a virtual certitude that the only people whom they could ever harm is the enemy (e.g., because only the enemy would have occasion to approach the mine field without the disarm code during a given period). Landmines would be morally acceptable weapons if they biodegraded into something harmless, for example, or if it was pre-arranged for them to be able to be de-activated and harvested at the end of conflict.

**Reference list**

Arkin, Ronald. 2013. ‘Lethal Autonomous Systems and the Plight of the Non-Combatant’. *AISB Quarterly* 137: pp. 1-9.

Asaro, Peter. 2016. ‘Jus nascendi, Robotic Weapons and the Martens Clause’. In *Robot Law*, edited by Ryan Calo, Michael Froomkin and Ian Kerr, pp. 367-386. Cheltenham, United Kingdom: Edward Elgar Publishing.

Bratman, Michael. 1987. *Intention, Plans and Practical Reason*. Cambridge, Massachusetts: Harvard University Press.

Crootof, Rebecca. 2016. ‘A Meaningful Floor For “Meaningful Human Control”’. Temple International and Comparative Law Journal 30 (1): pp. 53-62.

Gauthier, David. 1984. ‘Deterrence, Maximization, and Rationality’. *Ethics* 94: pp. 474-495.

Gauthier, David. 1986. *Morals By Agreement*. Oxford: Clarendon Press.

Heyns, Christof. 2013. ‘Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Human Rights Council, Twenty-third session, Agenda item 3, Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development.’

Kavka, Gregory. 1978. ‘Some Paradoxes of Deterrence’. *The Journal of Philosophy*: pp. 285-302.

Lin, Patrick. ‘The right to life and the Martens Clause’. Presented at Convention on Certain Conventional Weapons (CCW) meeting of experts on lethal autonomous weapons systems (LAWS), at United Nations in Geneva, Switzerland on 13-17 April 2015.

MacIntosh, Duncan. 2013. ‘Assuring, Threatening, a Fully Maximizing Theory of Practical Rationality, and the Practical Duties of Agents’. *Ethics* 123 (4): pp. 625-656.

MacIntosh, Duncan. 2016. ‘Autonomous Weapons and the Nature of Law and Morality: How Rule-of-Law-Values Require Automation of the Rule of Law’. In the symposium ‘Autonomous Legal Reasoning? Legal and Ethical Issues in the Technologies of Conflict’, *Temple International and Comparative Law Journal* 30 (1): pp. 99-117.

MacIntosh, Duncan. Unpublished (b). ‘Autonomous Weapons and the Proper Character of War and Conflict (Or: Three Objections to Autonomous Weapons Mooted—They’ll Destabilize Democracy, They’ll Make Killing Too Easy, They’ll Make War Fighting Unfair)’ (ms. Dalhousie University, 2017).

MacIntosh, Duncan. Unpublished (a). ‘PTSD Weaponized: A Theory of Moral Injury’. Mooted at the conference, ‘Preventing and Treating the Invisible Wounds of War: Combat Trauma and Psychological Injury’, Clair Finkelstein, organizer, held by the Center for Ethics and the Rule of Law at the University of Pennsylvania Law School.

Nussbaum, Martha. 1993. ‘Equity and Mercy’. *Philosophy and Public Affairs* 83. pp. 83-125.

O’Connell, Mary Ellen. 2014. ‘Banning Autonomous Killing—The Legal and Ethical Requirement That Humans Make Near-Time Lethal Decisions’. In *The American Way of Bombing: Changing Ethical and Legal Norms From Flying Fortresses to Drones*, edited by Matthew Evangelista and Henry Shue, pp. 224-235, 293-298. Ithaca: Cornell University Press.

Pop, Adriadna. 2018. ‘Autonomous Weapon Systems: A Threat To Human Dignity?’, Humanitarian Law and Policy. (last accessed April 19, 2018). http://blogs.icrc.org/law-and-policy/2018/04/10/autonomous-weapon-systems-a-threat-to-human-dignity/

Wallach, Wendell. ‘Terminating the Terminator: What to do About Autonomous Weapons’. *Science Progress: Where Science, Technology and Policy Meet*. Tuesday, January 29th, 2013, http://scienceprogress.org/2013/01/terminating-the-terminator-what-to-do-about-autonomous-weapons/

Zacher, Jules. 2016. ‘Automated Weapons Systems and the Launch of the US Nuclear Arsenal: Can the Arsenal Be Made Legitimate?’. (Manuscript)

1. For helpful discussion, my thanks to a philosophy colloquium audience at Dalhousie University, and to the students in my classes at Dalhousie University and at guest lectures I gave at St. Mary’s University. For useful conversation thanks to Sheldon Wein, Greg Scherkoske, Darren Abramson, Jai Galliott, L.W. and M.D. Thanks also to Claire Finkelstein and other participants at the conference on The Ethics of Autonomous Weapons Systems, sponsored by the Center for Ethics and the Rule of Law at the University of Pennsylvania Law School in November 2014. This chapter is part of a longer paper originally prepared for that event. [↑](#footnote-ref-1)
2. In a companion paper (MacIntosh Unpublished (b)) I moot the additional objections that AWS will destabilize democracy, make killing too easy, and make war fighting unfair. [↑](#footnote-ref-2)
3. Thanks to Robert Ramey for conversation on the points in this sentence. [↑](#footnote-ref-3)
4. On this explanation of the rationality of forming and keeping to plans, see Bratman 1987. [↑](#footnote-ref-4)
5. I do not mean to take a stand on what was the actual rationale for using The Bomb in those cases. I have stated what was for a long time the received rationale, but it has since been contested, many arguing that its real purpose was to intimidate the Russians in The Cold War that was to follow. Of course, this might still mean there were consequentialist arguments in its favor, just not the consequences of inducing the Japanese to surrender. [↑](#footnote-ref-5)
6. The classic treatment of this rationale is given by David Gauthier in his defense of the rationality of so-called constrained maximization, and of forming and fulfilling threats it maximizes to form but not to fulfill. See his1984; and his 1986, Chs. I, V, and VI. [↑](#footnote-ref-6)
7. For details on this proposal and its difference from Gauthier’s, see MacIntosh 2013. [↑](#footnote-ref-7)
8. It is, of course, logically possible for a commander to abuse such chains of command. E.g., arguably commanders do not escape moral blame if they deliberately delegate authority to someone whom they know is likely to abuse that authority and commit an atrocity, even if the committing of an atrocity at this point in an armed conflict might be militarily convenient (if not fully justifiable by the criterion of proportionality). Likewise for the delegating of decisions to machines that are, say, highly unpredictable due to their state of design, for example. See Crootof 2016, especially pp. 58-62. But commanders might yet perfectly well delegate the doing of great violence, provided it is militarily necessary and proportionate; and they might be morally permitted to delegate this to a person who might lose their mind and do something too extreme, or to a machine whose design or design flaw might have a similar consequence, provided the commander thinks the odds of these very bad things happening are very small relative to the moral gain to be had should things go as planned. The expected moral utility of engaging in risky delegation might morally justify the delegating. [↑](#footnote-ref-8)
9. On the use of delegation to a machine in order to save a person’s conscience, especially as this might be useful as a way of preventing in the armed forces those forms of post-traumatic stress injuries that are really moral injuries or injuries to the spirit, see MacIntosh Unpublished (a). [↑](#footnote-ref-9)
10. For some further, somewhat different replies to the dignity objection to the use of AWSs, see Patrick Lin 2015; Pop 2018. [↑](#footnote-ref-10)
11. For more on these last two points, see MacIntosh (Unpublished (b)). [↑](#footnote-ref-11)