"Firing, Forgetting and How Rule-of-Law-Values Require Automation of the Rule of Law; A Defense of the Use of Autonomous Weapons Systems in War and Peace"

Duncan MacIntosh,

Dalhousie University

November 17, 2014

Introduction

This paper defends not only of the use of Autonomous Weapons Systems (AWSs), but the automation and robotization of war and even of the formulation, administration and enforcement of law itself.

While AWSs have obvious military advantages, there are prima facie moral objections to using them. In Part I of this essay, however, I point out similarities between the structure of law and morality on the one hand and of automata on the other, arguing that these 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. Ethically speaking, deploying a robot isn't much different from deploying a more or less well-armed, vulnerable, obedient, or morally discerning soldier into battle, a police officer on to patrol, or a lawyer into a trial. All feature automaticity, in the sense of deputation to an agent we do not then directly control. Such relations are well-understood and well-regulated in morality and law; so there isn't much challenging philosophically in having robots be some of these agents (excepting the implications of the limits of the technology at a given time for responsible deputation).

Having sketched that program, in Part II I argue more specifically (and contra Mary Ellen O'Connell, Christof Heyns, et al) that there are many conditions where using

AWSs would be strategically, morally and rationally appropriate. In Part III I deal with objections -- that use of robots is inherently wrong or violating of human dignity, or will make violence too easy, or that proliferation of AWSs will be dangerous politically. I conclude, in Part IV, with second thoughts based on the differences between two conceptions of law, these distinguished by whether they see laws as unambiguous rules inherently uncontroversial in each application; and I consider the prospects for robotizing law on each. I also identify certain elements of law and morality that robots can participate in only upon their attaining of personhood, others, before.

Part I: Similarities Between Morality/Law/Moral Agents and Automata; How Rule-of-Law-Values Require Automation of the Rule of Law

The debate around autonomous robot weapons is an occasion to reflect on the nature of human-discerned morality and human-created and -administered law. And an awareness of the vast similarity between human and robot in this respect argues against objecting to robotic weapons for their somehow lacking a crucial feature of human morality. Not only would transferring adjudication and enforcement of law and morality to robots be consistent with human law and morality, but the highest ideals of both require this.

Foundational to my case is that, as legal and moral beings, much of the most important part of our lives is the attempt to discipline ourselves to rules, things that could easily be the programming foundation for automated expert systems in law, policing and morality. Here I go beyond those who think that human rights laws of engagement can be programmed into robots, or that robots can be programed to respect distinctness, necessity and proportionality (i.e., protect civilians over soldiers, only inflict harms when necessary to important military objectives, and don't let the weight of harms exceed the

importance of the objectives). I say robots represent the possibility of the de-virtualization of the rule of law, the logical extreme of such rule. For robots are algorithm-driven; and so, in its fullest fruition, is law.

In fact, ironically the hardest things to automate are the low-level perceptual skills we share with animals, things like situational awareness, identifying items in our environments. The "higher" functions sustaining morality and law are easiest. But of course, even robotizing those is relatively hard.

Fortunately there are other features of morality that make it amenable to robotizing. One is that moral labour is divided, its parts apportioned to those most competent at each. Ordinary citizens are entitled and obliged to make decisions affecting their welfare in ways not affecting each other, police, decisions about whom to arrest and how for prima facie and well-evidenced violations of the law, judges decide guilt and penalty, legislators decide laws, Commanders-In-Chief, when and how to make war, etc. And there are similar stratifications in armed forces roles – Privates have fewer responsibilities than Corporals, and so on. No one is obliged to master all of morality or law or the ambit of combat, and so no robot either, this leaving the easier task of robotizing any given part.

Another feature of morality making it amenable to robotizing is that it is not required of any agent within a given part of our system of morality that it in effect pass a moral Turing test (i.e., reliably be able to do things indistinguishable from what a human paradigmatically wise and compliant moral agent could do), only that the system of which the agent is a part pass. Put differently, what it is for a given agent to pass is situation relative and demarcated.

Yet another feature of morality and law suiting them to robotizing is that their institutions are so stratified that those higher in the strata make decisions that those lower implement as if automatically or robotically relative to the higher.

Finally, the point of rule of law is to have people live by principles impartially to the benefit of all. But the problem with having law human administered and enforced is that people are inevitably partial, biased, and weak willed, exhaustible, unable fully to work out the consequences of the principles they follow, and so on. We have to take people as we find them, biases and all. Robots, however, we could make into perfect administrators and enforcers of law, unbiased engines of legal purpose. This is why so deploying them is the perfection of the rule of law, and something requiring by rule of law values.

More on these themes in Part IV. But now, on to…

Part II: 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. (C.f. Mary Ellen O'Connell.) After all, 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 he will stay an execution.)

Clarity requires breaking this issue down into several sub-issues, namely, whether there should always be reconsideration at least in the sense of revisiting whether the weapon should be allowed to kill, and then, in any given case, whether there should be

reconsideration in the sense of reversing the decision to kill. And if there is to be either or both, 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. 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 to maximize the probability and magnitude of morally good outcomes. She seeks therefore to maximize moral expected utility, and has considered the likelihood of the weapon's having morally good effects given its design and proposed circumstance of use. Thus she would bear in mind whether use of the weapon is likely to respect distinctness, necessity and proportionality in war-like situations, and whether it can likely distinguish, or can be deployed in such a way as to respect the distinction between those morally liable to being harmed and those to be protected. 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 now judges it best to resist. 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. (C.f. Michael Bratman on the rationality of forming and keeping to plans.)

b) Short Term vs. 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.

c) Resolute Choice Cases: one expects to benefit not from allowing the weapon to finish its task, but from the consequence of committing to its unreconsidered use should the enemy not meet some demand (c.f. David Gauthier on the rationality of constrained maximization, of forming and fulfilling threats it maximizes to form but not to fulfill). The consequence sought will be a 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 it is: one doesn't reconsider in a) 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 couldn't have the advantages of strategies. (This last rationale is widely contested; after all, why still act from a strategy one now knows will fail? So I favour a variant on that 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.)

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 can't 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 of a weapon that could be made monitorable and recallable as it engages in its mission, but giving it this ability 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 (except d), where reconsideration is impossible), 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.

Let us be clear, however, about which possibilities are morally and rationally relevant. 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. Since this was considered at the outset, it would not be appropriate to reconsider on that ground just before the weapon's 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 many think as follows (e.g. Nussbaum in the her essay on mercy): 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 can't face actually doing the deed, and decides that justice requires something different – one comes to find the idea 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. Won't putting an automated robot in the loop here deprive one of the possibilities of new moral learning?

It's 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's different from having actual evidence in advance that the situation is new; and the mere possibility by itself it is no reason to forego the benefits of a disposition to non-reconsideration.)

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. 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 of the ethical requiredness of this -- but a human's doing the killing would leave her traumatized in a way constituting a moral harm to her (e.g., she would have crippling PTSD and a tendency towards suicidality), or 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 automated system do the killing – the morally hard but necessary task gets done, but the agent has her moral self protected.

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 automated 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 automate a killing 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's 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, deciding into whose body to push so many pieces that he will die, into whose body to push only enough pieces to maim, and so on? The effect is the same – it's as if a blast kills some, maims others, leaves some mostly unharmed. But the second method is morally worse. Better to automate. And better to use a stupid automaton, a blast whose effect on a given person is determined by what is in effect a lottery. Here it is the very automatedness and randomness of the grenade, plus the fact that only one morally possibly questionable decision need be made in using it, 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. I assumed everyone in the room is 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, or harm them not at all.

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 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 couldn't rationally or morally do one's self. (This is not case d), where we use an unrecallable weapon because it's 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.)

We've 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. Shouldn't this 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…

i) 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.

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

j) 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 (c.f. Ronald Arkin, et al). 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 there 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 don't self-identify with a civilian transponder. Nothing needs to be over an aircraft carrier and anything there is an enemy. An example of the second might be where the AWS is better at detecting the enemy than a human, e.g., by 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.

k) 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 (again, Arkin, et al) -- 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.

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

Part III: Objections Considered

III.1 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 persons' rights.

I doubt this, but even if it were true, that wouldn't 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 doesn't mean you shouldn't ever incur it – maybe you must grab a hot metal door knob to escape a burning building, and that'll hurt, but you should still do it. Maybe you'll have to inflict a painful injury on someone to protect yourself in self-defence, but that doesn't mean you must not do it. Similarly, even if death by robot were an inherent wrong, that doesn't 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. Wouldn't you want to activate it? And wouldn't this 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. Wouldn't you prefer death by robot? This could be 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 he could take pity on you and spare you, he won't --

he is 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's 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's 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'd take the drink, not scrupling about the indignity of 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 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 his victims, and to anyone who must be involved in his 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 him 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 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 executioner has to soil his 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 M.O. is to select us randomly for death, rather than by means of specific proper liability to death.

Moreover, dignity is a luxury and sometimes must yield to factors of greater exigency.

Some of this, of course, is separate from what people perceive as being required by dignity, and from how important they think it; and if we are trying to win not just the war but also the peace, maybe we'll 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 and not unnecessarily

trample on those opinions. Maybe we even have a moral duty to take some personal risks in this regard.

I move on now to…

III.2 Objections From AWSs Being a Political Threat to Us All

Some worry that if we permit development of AWSs, soon all states will have them. But fortunately only non-failed states have AWSs now or in the foreseeable future, mostly states western or western trending, so AWSs will begin in the hands of the most responsible users, who can use them to stop others from getting them. Moreover, as Blum points out, defending ourselves from abuses of such weapons will require large state involvement in monitoring and policing our activities, and fortunately it is, again, the non-failed, western or western-trending states that most have these capacities. So there is the hope that AWSs will get initiated into our civilization in way that makes them a continuing force more for good than for evil. The most melodramatic possibilities are least likely; and we can do a lot to prevent them.

But won't non-state actors then acquire them, e.g., mini-drones that can fly a grenade into anyone's bedroom? Maybe; but they won't acquire things demanding a massive industrial culture to produce, e.g., drones as big as passenger liners. Their production requires a massive society, one which can only come to exist by the attractions of being approximately just. So any society that manages to produce them is likely to be just, and therefore justly overwhelming of any society that would oppose it.

More generally, if these things become widely available, much as guns are, the result won't be mass violence. Instead, everyone will have to be nicer to everyone else. And in any case, there are defenses against the apocalypse envisioned by people like

Blum of an AWS-armed war of all against all, namely, education into law-abiding values, the Hobbesian argument for letter the state have superiority in such weapons, and the strategy of making the world a good enough place for everyone so that no one has a grievance worth using violence to settle.

But now to a further case favoring AWSs, namely: l) Special Vulnerability of Unjust Regimes to AWSs Cases: Contra Gabriella Blum, development and use of such weapons could be required by virtue of their inherent comparative effectiveness against non-democratic regimes and harmlessness to democracies. The former can be ended by killing the dictator who rules them, and so are particularly vulnerable to tech that can kill rulers, e.g., a hunter-killer drone programed to seek a leader's cell-phone and explode. But democracies produce leaders by forming a national will the leaders then express, and emplacing them in structures in which leadership is diffusely executed, has clear rules of leadership succession, and so can endlessly resupply leaders if any given leader is killed, so that there is little point in killing a democracy's leader. This argues for encouraging the general proliferation of automata, not just in the military world, but also civilian – let even civilians have AWSs so that they can take down their own dictators. For the proliferation of weapons is more harmful to less democratic rule. In the U.S. there are as many guns as people – hundreds of millions. Yet there are five orders of magnitude fewer gun-deaths than guns in a given year. I expect the reverse holds in less democratic countries where the population at large is not armed, only its radical minorities and military elites. AWSs may be the great equalizer in developing nations.

III.3 Objections From Robots Making Killing Too Easy

Many worry that robots will make killing too easy, this resulting in "unfair fights", or making us careless about using violence.

On the first point, it is sometimes said a fight is only just if fair, in the sense that both parties have a reasonable chance of winning. But in fact this is not a general condition on just fighting. If one side is righteous, the other, not, it is absurd to require that the unrighteous side have a chance. For is it seriously being proposed that no one may go to war unless they stand a reasonable chance of losing? Would the U.N., say, actually enforce this, e.g., go to war on anyone who is prepared to go to war without a reasonable chance of losing? Would the U.N. then govern itself by that rule -- limit itself to technologies and strategies that gave the law-breaking nation a reasonable chance at victory? No and no.

On the other hand, we can imagine situations which rightly evoke concerns similar to the above, although not to be confused with it, situations where: i) there is no point to a "fight" because it's a fait accompli due to overwhelming force on one side, so steps should be taken to avoid a massacre; ii) the weaker side can do no damage, and so might as well be allowed to flail away without response, or with the most non-lethal response possible with minimal risk to the side required to be merciful; iii) it is an open question who should prevail but, for non-morally-relevant reasons, one side has the advantage, so outside interests level the playing field; iv) one side is using means that violate human rights, so outside interests rightly intervene – as cases iii) and iv) illustrate, part of successful warring is not attracting others to fight against you, so you must fight by rules that won't be found outrageous; v) the two sides will one day make peace, and

this will go better if both sides had pain in the conflict so that each can have the self-respect of thinking they fought a good fight; vi) outside interests who will have relations with the parties will have better relations with them if both sides experienced pain, neither seen as having won too easy a victory and so perhaps one undeserved. In cases v) and vi), dignity and so post-conflict peace may require mutual pain. Some of these considerations are, of course, more rooted in the sociology of conflict than in justice; and phenomenon may feature outdated conceptions of chivalry in battle. Similarly, consider case vii): you face an enemy who will resent you more and fight harder, e.g., with suicide attacks, the stronger and more invulnerable you seem; so better to handicap yourself and present as more human and vulnerable -- forego robots. Now case viii): It has been argued that if we use robots to do our fighting and so don't allow enemy combatants our soldiers to shoot at, they'll target our civilians instead. I say this is little matter if our robots make us invincible even in defending our citizens. But in any case, as I've elsewhere argued, our civilians should be the ones accepting the risk of terrorist attack instead of the risk being put on our own soldiers, or on enemy but innocent civilians. The risk is better born by our civilians than our soldiers because it will be more highly attenuated for each of us than for each of them given our vastly larger numbers and greater geographic dispersion; and better born by our civilians than by the innocents in the countries from which our enemies emanate because this is necessary to our not attracting more attacks by acting in callous hegemony in causing collateral damage to foreign innocents.

In any case, none of these considerations is a principled point against robots, only practical. Indeed, for some of the above situations, robots might be just the thing to even the odds.

There are, however, special issues about robots in combat. First, a robot is not a moral patient, a stake-holder, a being with interests. (Or at least let's say we're only talking about robots whose design has not advanced to them being moral patients.) So a robot's actions cannot be justified as self-defence. Second, if the robot is remotely activated by someone out of harm's way, she can't claim self-defence either, or at least not immediate self-defence. Third, since the robot's actions can't count as self-defence, deploying it in such a way that it can be expected to draw fire and then authorizing it to fire back in "self-defence" might be morally problematic entrapment.

But there is a general reply to these complaints, namely, that it is just false that the only legitimate use of violence is for self-defence. Violence can be permissible to defend other persons, or to impose moral restitution, as where one must use violence to recover stolen items or land; or in moral prophylaxis to prevent thefts of items or land, say; or, more generally, to keep states of affairs within the order of the deal that most benefits all. And just as humans can justly use violence other than for self-defence, so they can justly use robotic violence. Even if we thought self-defence was the pre-eminent moral pretext for violence, we might get around a robot's not being able to plead self-defence by putting a man in the machine and making him just vulnerable enough that, to ensure his own safety, he'd have to activate the machine's powers of autonomous combat in fire-fights. But isn't this just a cheat? No. The idea is that the man/machine are there on a righteous pretext as boots on the ground in some operation, and the measure of the

operation's going well is that you don't get fired upon, while if you do, the operation hasn't yet succeeded and you're entitled to fire back. The man's vulnerability has become a feature of the definition of the operation, namely, to make it possible for agents like him to walk safely in the region.

Another dimension of the moral deployment of these machines comes from the idea that since they aren't moral stake-holders, moral patients, they needn't protect themselves, and so can be asked to make more compassionate and considered responses to threats than could reasonably be asked of a vulnerable human. There is some truth in this. And yet these will be very expensive machines, the loss of one probably will mean loss of lives down the road, including non-combatant lives, so arguably there is as great a stake in a given machine's survival as in a soldier 's. Or at any rate, the difference wouldn't be orders of magnitude. (Especially if we quantify the value of a soldier's life, something sadly we must do in making the cost/benefit analyses to which we are driven in war, and so assuming that a soldier is worth, say, on the order of a million dollars.) Thus there might be good reason for the machine to defend itself: we have a right to make it do this as part of defending property needed for morally good effects.

As always, of course, an overarching consideration is to minimize total harms; but using robots might well do that.

The truth is that many who oppose AWSs do so because they oppose the nations with the wealth and knowledge to produce them, e.g., the U.S. Yet ironically, the U.S. is already able to wage war with very little risk to Americans. In some conflicts apparently the ratio of non-American to American lives lost, e.g., in Kosovo, is 12,000 to 1. The U.S. can kill 12,000 people before having even one death. So it's hard to see how robots

are going to make this much worse -- the numbers are already at the level of industrial accidents, not war.

Nevertheless, it might be argued that "gentlemanly" rules are sometimes appropriate, e.g., when the world is still sorting out justice in morality and politics, when there is still disagreement among reasonable people about what arrangements should prevail. For that time we recognize that debates can get heated and people can resort to violence to defend what they take to be their entitlements, but we know that it's just tantrums, metaphorically speaking, and we want to ensure that the two sides are balanced in their ability to act out, while being prevented from doing too much damage to each other; in that case, equality of fighting power is desirable, and maybe that means neither side resorting to robots. But it could mean both sides resorting to them, if that were thought the way to minimize human casualties. And even if we thought there was a problem with robots fighting humans, it might be OK for robots to fight each other – perhaps here two wrongs really do make a right.

But suppose we are beyond the sorting out of morality and politics and are in a trans-political estate where reasonable people no longer disagree, where there are merely the outliers of insanity and chaos. Perhaps gentlemanliness is not what is called for in dealing with the morally verminous who remain, but instead the efficiency and easiness of robotized conflict. Violence easy to inflict with low cost to conscience may be good in situations typical of the war on terrorism, namely, situations that aren't really wars but more like police actions, except so risky and difficult that it would be evil to ask police to go, and against people who are enemies to all of civilization. Their mental instability has been attracted to a convenient cause affording them a pretext to express the violence

within them. They aren't by that point like you and me. Of course if they can be de-radicalized they should be. But what if they can't? Thus AWSs may be appropriate where there is no real probability of moral error, or against a virulent enemy implacable to all other forces.

And we may well be in a historical period marking the end of war. So perhaps inhuman techniques prosecuted by non-human agents are appropriate for the remaining required mopping up action. Their very targetedness means that their use isn't a war use. (For in war collateral damage is expected, not so in police action. Therefore, perhaps it is criterial of whether something is a war whether collateral damage occurs, is expected or is permitted.)

Imagine the perfect social deal for all. Imagine there is no longer any reasonable dissent about it. But imagine there is still stupidity, ignorance, greed, weakness of will, insanity, all things that are the human part of the externalities plaguing the deal. (An externality is anything that affects an outcome of trying to implement a deal that isn't itself part of the deal.) So isolated, the people who are like this are by nature a very small problem, one at present requiring very risky activity to deal with. (It is estimated that there are only somewhere between 14,000 and 40,000 people in the world who are really terrorists –whatever that means, exactly – as opposed to nut-cases, people who can't get jobs to feed their families, people in effect extorted into terrorist acts by fear for their own safety, and so on. There are very few people of settled character who have alternative destinies available to them and yet who set themselves against us.) But this is precisely the part that perhaps could be automated -- rounding up the crazies, or failing that, killing them. Perhaps this can be so automated that only a small number of people,

with a minor effort, need concern themselves with the matter – e.g., the President in the Monday morning drawing up of the drone kill list. The rest is merely a matter of oversight to make sure the process doesn't become corrupted, with periodic review to make sure the algorithm is correct.

For suppose there is an ideal deal. Plausibly, it will maximize commerce and interaction generally, trending therefore to maximal inclusiveness. It will then be the arrangement leaving out as few people as possible. Thus anyone not able to accept a place in it does not have a sane view. This means that, so far as they are a threat, they aren't worth our time, or at least not our lives. They should therefore be dealt with as safely to us as possible, viz., as robotically as possible. This needn't mean killing them, but if that is the only way to protect the deal then it is both permitted and required, likewise, then, robotic killing.

Our political world is verging on such a deal. E.g., it is virtually inconceivable that any of the western nations will ever war with each other again. And nearly every nation is verging on west-affiliating. The only exceptions are China, Russia, North Korea and the proto- and failing states of the mid-East and Africa. But China and Russia have begun west-affiliating, on the attraction of commerce. The rest will either western-affiliate or be problematic outliers of the sort discussed above.

Only a society governed by the ideal deal, and so only a just society, could conceive and afford robotic innovation. Therefore any society with the power to use it tends to have the right to use it. In many senses, then, robotizing marks the end of war days: there will only ever more be police actions, not wars, and the more automatic those actions, the better for all. For the robotizing of police action means fewer and fewer

people who shouldn't be harmed by it will be harmed by it. It will almost be as if a society is policing itself by the very mechanism of conscience and at the speed of thought, robotically facilitated.

President Obama and his drone method prefigures all of this. And the fact that no nation is plausibly telling him to stop, and no plausible nation is telling him to stop, is proof of the correctness of what he's doing. He's killing people who should not be taken seriously, and who can't as efficiently be stopped in any other way. And a great part of morality is figuring out what views do not need to be taken seriously, and what persons are so beyond the rational and moral pale as that we have less duty to put ourselves at risk in dealing with them. That sounds harsh, but it is in fact the functional role played by the deliberations, policies, and policy implementations of the rule of law. But part of his approach being right is that it increasingly eliminate collateral damage, and that it increasingly trend towards non-lethality where possible.

It is worried that if it's too easy or costless to kill or war or transgress national boundaries to attack people who are harming you, you'll do so too readily and without being morally mindful. But this would owe not to robots being available, but to not visualizing the consequences of actions, lacking policy constraints, or not having a free press ensuring the presentation of vivid images of the consequences of the act, and so on.

To those who worry it will be too easy to take violent solutions, I reply that it is presently too hard. There are people causing a lot of trouble and we can't do anything about it because we'd first have to move a whole government, change the culture of a whole people, or mobilize our entire nation's attention and overcome our people's reluctance to risk themselves and their loved ones. But AWSs let us by-pass all of this,

freeing us to help even where our own interest is not much involved – Africa, say, instead of the Middle East.

Thus, consider **case m): It is sometimes objected (e.g., by Gabriella Blum) that AWSs will make it too easy to go to war.** For they are cheaper than human forces, easier to move to places of conflict, they result in fewer human casualties on the friendly side so that the nation using them is less likely to suffer weakness of will in the war, they will probably inflict less collateral damage against enemy non-combatants, so that, again, less guilt will be involved in ordering their use, and they can be used in relative secrecy, thus escaping the public monitoring that would provide pressure against their morally outrageous use.

I reply first that, if you have AWSs, and you need to go to war, you can do so more successfully than with conventional tools and man-power, meanwhile controlling the inappropriate use of AWSs by education, propaganda, law-enforcement, and so on; but if you don't have AWSs, you can't use the foregoing techniques to win your war. In other words, it's easier to compensate for having AWSs than for lacking them. Second, adopting AWSs may be justified by virtue of them making it easier to do the right thing. For since they are vastly cheaper, more efficient, less casualty-involving of friendly troops and more easily re-located, they can be deployed in situations where we have a duty to intervene, but less of a national interest in intervening, e.g., in Africa rather than the oil-rich Middle East. AWSs would empower us to do our moral duty in a way sustainable pro bono, without expectation of immediate business investment return. AWSs can make us more morally responsive to need.

……

Summing up the first three sections of this paper, it appears, then, that it is false that it is always best for a human decision to be proximal to the execution 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 isn't 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's a case:

n) We're 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 don't 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 OK 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.

Part IV: Second Thoughts, Two Conceptions of Law, and the Role of Robots in The Kantian Kingdom of Ends

I began by suggesting that rule-of-law values require that we automate the rule of law. Here's a case to show the advantage of doing so: o) Rule of Law Values Cases: in some situations, the use of an AWS may be more in the spirit of the rule of law rather than the rule of a man, or of the rule of humans generally, given the susceptibility of the latter to moral failings of specious partiality, impulse, exhaustion, weakness of will, and emotionality; and so their use may be required by the very commitment to the rule of law. E.g., suppose we are tempted to arm one side in a civil war, but worry that, due to past religious conflicts, they would use our weapons not just to win the war, but also to take revenge against and oppress the other side. Then we might instead think that using AWSs programmed by us would be more just and temperate, would spare more life, and would have a greater chance of producing a lasting peace, since it interrupts the violence cycle (of endless revenge).

But the overall program I proposed may presuppose a false conception of what law is. There are at least two conceptions of law. One sees it as the unbiased and unexceptioned literal application of unambiguous rules. The other sees it as the

determining of what in each situation that people face should be done, whether what was done in a given situation was right, and, if not, what should be done in consequence. And it sees all this as occurring by argumentation based on interpreting inherently ambiguous rules and negotiating their inherently contestable applicability in given cases. The first conception deploys the metaphor of law as rules of a machine, the second conceives law as ongoing debate about behaviour-regulating norms that is conducted in terms of discussion about human-chosen principles (rather than character traits, best outcomes, divine commands, or the impulses of a powerful person), where the behavior to be regulated could probably instead have been improvised without need of prior meditation upon principles – invoking laws is just a handy conversational trope. Obviously the first conception seems more fully automatable. But even on the second some aspects of law might be robotized, e.g., parking law enforcement; and even on this conception, a robot advanced enough to deliberate about and justify its actions just as a person can might be able to contribute. Although at that point the robot/person distinction becomes merely one of physical composition, not conceptual difference.

How difficult would it be to produce such a robot? The difficulty may be vastly over-estimated. For suppose we're trying to produce a judge machine, and what we have can only do things like co-vary mercy with degree of remorse if the latter is discovered by interview and given to the machine on a rating scale of one to ten. We might think, "that's too crude". But more positively, we might think, "there's a lot of latitude about what morality requires, and this is plenty good enough, good enough that we could operate by this standard, too." In other words, maybe we learn what the level of detail of our morality is, and what a permissible version of it, is from what we could make a

machine follow. Perhaps our conception of morality should evolve with our attempt to reproduce it in machines, much as one might find this in trying to educate people from different moral cultures, and having this change one's expectations of what it's reasonable to require of people. We already see this in our own lives: as one acquires more life experience one realizes just how messy and varied living a moral life is, how much tolerance and forgiveness it requires.

But what of the prospects for machines being able to disambiguate inherently ambiguous rules, or negotiate their proper application in a given case? Lawyers will know about this more than me, but I gather that on this conception, the law is all about frames, imposing frames of interpretation on the facts given book law and case law. A reading of a given case will strike a judge as right, and then she will justify the reading by citing elements of book and case law. For each legal question, there are an enormous number of frames one could impose, and book and case law can be used to support any of them. Is there a right one for each question? Arguably not. What there are is many plausible ones. And if we didn't know that a given verdict had occurred by a machine "flipping a coin" amongst the possibilities and then compiling an argument from book and case law for it, we might have found the result perfectly legally plausible, and an example of an entirely livable and prima facie just system of law. So it's easier than one might have thought to have machine-run law, because the level of resolution in relation to some imagined truth that the law operates at is coarser than one might have thought. Machines can produce good enough law, in part because good enough is all that law ever is.

But what of moral theorizing? Can that be automated? It may seem not. However I think one reason people doubt that we could make fully moral robots is that, even if we feel we have a kind of intuitive moral competence ourselves, we don't feel we could describe it, put it into a followable code; and on the assumption that a robot will acquire morality only if we program it to do so, since we don't know how to do that, a robot can't acquire it. But I believe we know better than we think how to describe it – think of the totality of the body of law in North American countries as a good first approximation, for example. Second, even if we couldn't directly program a robot into morality, we might be able to teach it, exposing it to our unsystematised moral reflexes and letting it learn to copy them, even if none of us can adduce an algorithm for what we're doing. Third, we don't have to give a robot all of morality to responsibly use one in some morally important context, only enough that it can function like a moral Turing Machine for that context, that task. Meanwhile, it may be that we have moral knowledge we don't know how to express in rules yet. Or it may be that the knowledge doesn't exist only because morality is in part made up as we go along. And perhaps one day that process too could be automated. In any case, certainly a large part of instruction in moral judgment is instruction in rules of thumb whose following will get you a long way – think of the law, of bioethical codes, of laws of war, of international human rights, rules of engagement, and so on.

It will be objected creating an expert system out of our extant laws would not be to robotize the highest function of moral expertise, vis., the figuring out of moral obligations , but only some arbitrary conception of them, morality as conceived in the U.S., for example, rather than, say, Iraq. Robots can only follow rules, not create them,

only implement some person's or nation's conception of morality, not investigate moral issues to find the moral truth.

But this mistakenly assumes that, while rules of conduct can be programmed into robots, we cannot program rules for discovering the right rules. In fact, in law there are rules for determining whether a policy has been legitimately enacted. And in morality there are a very small number of principles vying for determining morally correct behaviour – e.g., principles saying that we should do whatever respects natural rights, or maximizes happiness, or fulfills mutually advantageous contracts – and these mostly agree about right conduct. And there is every hope of a "Theory of Moral Everything" that will unify all plausible moral intuitions and provide arguments successfully dismissing unassimilable intuitions. There is already increasingly less debate about vast parts of morality, as seen in the increasing balance between free market and welfarist features in modern states. Here we don't just happen to have less disagreement; rather, we are discovering truths that are compelling consensus.

What then couldn't machines do in law and morality? Well, it is one thing to apply rules that serve goals, another to choose goals qua ends, things to be conceived as valuable in themselves and as objects to be pursued by moral agents. An open question on all counts is what machines can contribute to law if they don't themselves have interests situating them in what Immanuel Kant called the Kingdom of Ends. To have this role they must be interest holders, goal choosers, and agents negotiating what laws morally must be followed in the course of pursuing these goals or ends. If robots don't have and can't adopt ends, they can only ever be human-created compilers of records of human norms, checkers of the consistency of human-created laws, and administrators and

enforcers of these laws. The relevance of that is this: our moral culture, and probably morality itself, has the following structure: for each level of decision and authority, the levels below it are in effect the automated implementers of its decisions. So no one can object to a robot being the implementer of a human-taken moral decision. Even the highest level moral decisions we attempt to bring under the rule of law. But interestingly, the highest levels are subjected to the majority will of all moral patients and agents. So that level is informed by us, but it holds us individually to our collective will, and our collective will holds it accountable. The only place in the loop where there couldn't be automated rule-following is in the fact that the whole thing is operated by beings with interests, and is operated in the service of those interests. And this is something in which a robot can't participate unless and until it has an interest. But arguably that is no objection to its playing instrumental roles anywhere else in the system.

So what would it be for robots to have an interest? Immanuel Kant thought we ought to treat people as ends in themselves (i.e., as beings with goals whose attainment matters to them and ought to matter to us); we shouldn't just use them as means to our own ends or goals. To treat people as ends means either not obstructing them or even positively helping them in their attaining their ends, the things they deem of value for their own sakes. This makes us all co-legislators of the rules we should follow in the Kingdom of ends, rules about when we must help each other, and when we need only help ourselves. And Kant thought that we should do only what we could will without contradiction be done by all people in similar circumstances, for what makes us free agents is us writing laws for our conduct, and that's the test of something's being able to be a law, a principle able to govern everyone always. We can move from the latter point

to the others: imagine what you could will without contradiction be done by all people in similar situations. This is to imagine a law of conduct. Now imagine what states of affairs would result of all people following only permissible laws of conduct. Then these are the states people are permitted to desire, the ends they are permitted to choose. And since we all must use the same test in writing the laws, we are all co-legislators. It is also explained what the relation is between being a moral agent and being able to set ends: being a moral agent consists in being able to write coherent laws, which is the same as requiring and permitting oneself to aim only at those ends that could result of everyone's following such laws. And what is it to have ends? It is the same as being disposed to behave in any of the ways permitted by the law, for all such behaviours will tend to have certain outcomes, which can then be taken for the agent's goals or ends. But what is to determine which subset of possible ends a given person will set for herself? It doesn't matter, morally speaking, so long as they are compatible with the ends of others in the foregoing sense. We teach some ends to our children, and leave others to accidents of their genetics and life-experience; and we could do something similar for robots. (This raises large issues I can't get into here. E.g., if we get to the point where we can make robots like this, would it be politically permissible to do so? This is complicated, for they could be manufactured in large number, they'd have the right to vote, and they could overwhelm our own voting rights, for example. They would also consume resources, and this would have to be balanced against their capacity to make economic contributions benefitting us all. So there would be population control issues. And on and on.)

Now to some general claims. First, under what conditions should an autonomous device whose behaviour could affect the welfare of moral patients be allowed to be

"released into the wild"? Just when it is able to do the right thing for the right reasons. Second, under what conditions should such a device, rather than its manufacturers or deployers, be held accountable for what it does? When it is genuinely able to select ends, in the Kantian sense. Third, what would a device have to be like in order to satisfy the first two conditions? It must be a full moral agent, that is, able to make choices because they are arguably the right choices to make, able to offer a justification of them and to respond to counter-vailing justifications (because moral and legal agents are accountable, in the sense that they must be able to give an account of themselves, of their decisions, if challenged), and to morally learn from other moral agents and from its own experience. And to be like this, it would have to be itself a moral patient, that is, able to know at first-hand what it's like to be mistreated, to suffer, so as to be able to decide to do or not do something because of what it would feel like to have this thing done to it.

Epilogue

Arguably the focus on whether to have machines that kill is problematic in the same way as focusing on the morality and rationality of forming and fulfilling apocalyptic threats is problematic in decision theory, ethics and in the philosophy of war: it focuses attention on extremely hypothetical issues, this sloppily implying permission to think in terms of doing truly horrible things in more ordinary contexts. It also over-focuses us on violent rather than peaceful resolutions of conflict, when the latter should be our emphasis and is where we really need creativity. War, policing and lethality should be only last resorts. But this is no argument against using robots, only for keeping their use in proper perspective.

For such perspective I offer a thought experiment. Imagine a machine with the following properties: you turn it on and it kills all those whose unjust aggression can be stopped only by them being killed; it captures and incarcerates all those whose unjust aggression can be stopped by nothing else short of killing; and as for those whose violent expression of discontent could be stopped by them being brought into a new social contract, employed, given apology and redress for past grievance, given psychotherapy, education and loving kindness, it contracts, employs, apologies, redresses, psychotherapizes, educates, and loves kindly. It pays for this by taxing those unjustly wealthy, or those who wouldn't miss the taxed amount, and who would otherwise eventually have to face the enmity of the aggrieved in a world of excessive inequality. (Better still – if this is not already implied -- would be a machine which first figured out the moral truth about how things should be distributed, and about who should live and who die, and then embarked upon the courses I've described in ways most just, moving us justly from a non-ideal world to an ideal world.)

There is no such machine, of course. So we have to be that machine. But now consider its first two elements – those that kill or incarcerate -- the war machine part. How should we feel about developing and deploying an automatized version of the war machine? A lot better, I think, if we know it will occur in the context of the rest of the more ideal machine. And my remarks in this essay must be seen in that context. I have defended using AWSs from several objections that have been raised, and made some new arguments in their favour, too. But I would take it all back if I didn't think AWSs would mostly be deployed by people in a system that emulates the rest of the ideal machine. For I see no other way for use of AWSs to be just.

It is widely thought that the problem of robotized war is the difficulty of designing new weapons. But ironically a much more difficult problem is trying to get the politically potent forces in the world to collectively emulate the justice machine. The core problem is that each polity is partial to its own people, and so must struggle to act with the impartiality of an engine of pure justice. It is the latter that I see as the great hope of the automation of the rule of law. For just as we can arrange for horrible things to be done that we couldn't make ourselves do, we can arrange for wonderful things of impartial benefit for the good of all to be done that we'd be tempted away from in the moment by our own partiality. Robots offer us a way to credibly issue and fulfill utopian assurances, and to forfend against dystopian tragedies of the commons.