Just say "no!" to lethal autonomous robotic weapons

William M. Fleischman
Department of Computing Sciences, Villanova University, Villanova, Pennsylvania, USA

Received 24 December 2014
Revised 9 March 2015
Accepted 9 June 2015

Abstract

Purpose – The purpose of this paper is to consider the question of equipping fully autonomous robotic weapons with the capacity to kill. Current ideas concerning the feasibility and advisability of developing and deploying such weapons, including the proposal that they be equipped with a so-called "ethical governor", are reviewed and critiqued. The perspective adopted for this study includes software engineering practice as well as ethical and legal aspects of the use of lethal autonomous robotic weapons.

Design/methodology/approach – In the paper, the author survey and critique the applicable literature.

Findings – In the current paper, the author argue that fully autonomous robotic weapons with the capacity to kill should neither be developed nor deployed, that research directed toward equipping such weapons with a so-called "ethical governor" is immoral and serves as an "ethical smoke-screen" to legitimize research and development of these weapons and that, as an ethical duty, engineers and scientists should condemn and refuse to participate in their development.

Originality/value – This is a new approach to the argument for banning autonomous lethal robotic weapons based on classical work of Joseph Weizenbaum, Helen Nissenbaum and others.

Keywords Autonomous lethal robotic weapons, Computational ethics, Software engineering complexity

Paper type Research paper

1. Introduction

In *Wired for War: The Robotic Revolution and Conflict in the 21st Century*, Peter W. Singer comments that when it comes to giving robotic weapons lethal capabilities, official military policy seems to be clear and emphatic: "Humans must be kept in the loop" (Singer, 2009). One imagines this to mean that before any robotic weapon can fire on a human target, a human must give authorization. However, Singer makes the quizzical observation that whenever this matter is raised in serious discussion, the result is averted eyes and a change in topic. The subtitle of this paper makes reference to a kindred and puzzling phenomenon the author has observed, most recently at ETHICOMP 2013, among ethicists whose area of concern is problematic applications of computing technology.

In fact, there has been an active discussion among philosophers, legal theorists and engineers, of the question of creating and deploying autonomous robotic weapons with lethal capability. New contributions to this discussion appear with regularity. It has stimulated an initiative to ban weapons of this type and, in turn, has spawned critical commentary on the supposed premature or wrongheaded nature of this effort.

Journal of Information, Communication and Ethics in Society
Vol. 13 No. 3/4, 2015
pp. 299-313

1477-996X
DOI 10.1108/JICES-12-2014-0065

In the current paper, I argue that fully autonomous robotic weapons with the capacity to kill should neither be developed nor deployed, that research directed toward equipping such weapons with a so-called "ethical governor" is immoral and serves as an "ethical smoke-screen" to legitimize research and development of these weapons and that engineers and scientists should condemn and refuse to participate in their development.

One of the founders of the International Committee for Robot Arms Control, Asaro (2012) has proposed that there is a theoretical foundation for banning autonomous robotic weapons "based on human rights and humanitarian principles that are not only moral, but also legal ones". Asaro argues that:

> In order for the taking of a human life in armed combat to be considered legal it must conform to the requirements of IHL [International Humanitarian Law]. In particular, parties to an armed conflict have a duty to apply the principles of distinction and proportionality.

At root, his argument rests on the finding that human judgment is essential in determining that these conditions hold, and therefore:

> […] there is a duty upon individuals and states in armed conflict situations, not to delegate to a machine or automated process the authority or capability to initiate the use of lethal force independently of human determination of its moral and legal legitimacy *in each and every case* [emphasis added].

Asaro's careful analysis and rebuttal of arguments claiming that such a ban might be considered ethically questionable commands my respect and concurrence. Nonetheless, in the current paper, I want to argue that the conclusion may be reached by appeal to a prior consideration based first of all on the respect owed to an adversary in combat considering the moral symmetry of mortal hazard, but even more fundamentally on the respect each of us owes to our own humanity. My argument against the development and deployment of autonomous robotic weapons with lethal capability is grounded in two different philosophical traditions, those of the eminent philosopher of war, General Robert E. Lee in the first instance and of the profound philosopher of technology, Joseph Weizenbaum in the second. In addition, I will adopt a slightly different rhetorical strategy in dealing with the arguments of those who see subtle philosophical difficulties that stand in the way of condemning this obscene development.

We summarize Singer's findings concerning robotic devices currently deployed or under development by the military and discuss the advantages these weapons confer in combat. In a more general context, we consider the effect the existence of these weapons has on the decision to wage war and on attitudes toward the conduct of war. We argue that these effects have the potential to convert the advantages of robotic weapons into dangerous disadvantages. Although it is unrealistic to expect a rollback of weapons already widely deployed, we consider that their effect in lowering the perceived costs of declaring war or acting pre-emptively should give rise to special consideration concerning limits on their capabilities.

Stimulated by the pressures of asymmetric warfare that followed the US invasion of Iraq, advances in robotic weapons technology have induced a mentality of "Gee, wow! We can do this!" ("Look at this new feature we can introduce"), often finessing the fundamental question, "Should we do this?" The rapid development of "capability creep", as it applies to robotic weapons currently in use generates difficulties for human controllers of robotic weapons systems. Considerations of efficiency and the logic of

"force multiplication" dictate a "many-to-one" relationship between individual devices and human controllers. But humans are not very good at the sort of multi-tasking required to oversee multiple independent devices and missions. The limited effectiveness of humans in this capacity leads ineluctably to the thorny problem of keeping humans "in the loop".

If the superior capabilities of robotic weapons and the limitations of humans acting as controllers so far compromise the military principle of always keeping the human in the loop, then perhaps, we can substitute an "ethical governor" implemented in software for the absent human controller. Properly programmed, weapons acting in autonomous mode could perhaps be constrained to "act ethically in war", observing all the articles of the Geneva Conventions, the Laws of War and, in the local context of the combat in which they are deployed, the Rules of Engagement. As robots are not subject to the psychological and emotional stresses that affect human combatants, we might even expect that they would act more morally than the human soldiers whose combat roles they assume.

This idea is at the heart of the research project of Ronald Arkin of the Georgia Institute of Technology. His book, *Governing Lethal Behavior in Autonomous Robots* (Arkin, 2009), describes formalisms for ethical control, representational choices, design options and a prototype implementation of such a controller. In the current paper, we critique this project from the perspective of a project in software engineering, making reference to the recent work of Gerdes and Øhrstrom (2013) and Englert *et al.* (2014), as well as earlier work of Nissenbaum (1994) on the difficult questions of responsibility and accountability. We present an instructive example (Glover, 2012) from the history of the Cold War that underscores the importance of human deliberation in situations of belligerent confrontation.

In the final section of this paper, we discuss the more fundamental question of the philosophical justification for the development and deployment of lethal autonomous robotic weapons. We reiterate the critical analysis of Norbert Wiener and Joseph Weizenbaum (Weizenbaum, 1976), which reveals the fundamental weakness in Arkin's reasoning. We consider the worrisome significance that resides in Arkin's assertion that the system he envisions will permit us to wage war in a manner better than humans do.

2. Robotic devices in use and under development

The accelerated development of robotic weapons – unmanned ground and aerial vehicles (UGVs and UAVs) – was spurred by the wars in Iraq and Afghanistan. Numbers tell part of the story. There were few of either type of system deployed in the 2003 invasion of Iraq. By 2011, there were an estimated 12,000 UGVs and 7,000 UAVs in the inventory of the US military forces. Significantly, the US Air Force currently trains more UAV operators than fighter and bomber pilots combined (Singer, 2011).

Enemy deployment of improvised explosive devices (IEDs) in Iraq created an instant demand for Packbots – a ground-based, essentially defensive device. The Packbot was used to detect and, if necessary, disarm IEDs without the risk of loss of human life. Initially, it was simply thought of as a "mobile pair of binoculars". With the addition of simple effector arms and grippers, the Packbot acquired the capability to disarm and destroy IEDs concealed by the enemy (Singer, 2009).

Initially, Packbots were used to locate and identify non-human threats. A related task was to locate and identify enemy snipers. Clearly, a mobile device that carries a weapon

in addition to its cameras provides the possibility of eliminating the threat by aiming and firing remotely under control of a soldier who does not have to appear in the sight of the sniper's weapon. Once again, the desire to shield one's own combatants from situations in which their lives are at risk provided the incentive for development of a robotic device with additional capabilities. Singer (2009) notes that the desire to increase the killing effectiveness of one's soldiers while increasing the distance between them and the enemy is a constant in the history of modern warfare. Therefore, this was a natural application for Packbot, Warrior, its more heavily armed successor, and congeners such as the Talon and SWORDS.

Not all UGVs have direct combat roles. The special dangers of the role of human medics serving in battlefield situations led to the development of a version of the Packbot that can search for wounded soldiers and provide a video feed that allows a distant human controller to deploy medical equipment on the so-called "med-bot" to evaluate and treat the wounded individual (Singer, 2009).

UAVs have undergone a similar rapid transformation. Armed UAVs carry out offensive missions under the direction of a human pilot or operator located thousands of miles away. Perhaps the best-known UAV is the 27-foot-long Predator, capable of 24-hour reconnaissance and surveillance missions, returning high-quality images day and night by means of normal and infrared cameras. Its synthetic-aperture radar can provide valuable information even when clouds, smoke or dust obscure the terrain. Along the size and mission-length continua, UAVs range from the Wasp (15 inches in length, 45 minutes of endurance) to the Global Hawk (nearly 48 feet long with an endurance of 35 hours), which provides wide-area search, high-resolution single-target identification and can operate autonomously between the signals to taxi, take off and land from its human operator (Singer, 2009).

In addition to deploying its own force of UAVs, the US Navy is also developing several unmanned surface and underwater vessels (USVs and UUVs) (Singer, 2009).

3. Advantages and disadvantages of robotic weaponry

It is not hard to see (and it is very hard to resist) the advantages of robotic weaponry. The first, and most compelling for an armed force possessing these weapons, is that they replace humans on the battlefield and, therefore, reduce the number of casualties this force will sustain. Beyond this, they are markedly superior to humans in what military strategists describe as the "three D's" – situations that are dangerous, dirty and dull.

Dirty environments include not only those, like desert battlefields affected by smog, smoke, sand and dust, but also those which have been contaminated by biological, chemical or radioactive agents. Robots have a very clear advantage in these environments where humans would be encumbered by bulky protective suits and related gear.

Many military missions require concentration over long periods of time. In addition to physical stress, there is the psychological stress of paying steady attention in otherwise boring circumstances. Humans can do this for limited periods of time and then need downtime to recover the necessary level of acuity. In contrast, robots do not need sleep, food or a break for "rest and recreation".

The human body is limited in speed and range of reaction to threats and forces that occur in combat situations. From g-forces acting on human pilots of advanced aircraft to speed of recognition and reaction to battlefield dangers, robotic systems appear to have

a clear advantage. The primary advantage of a robot in a dangerous environment is that its destruction involves the loss of a machine (although this may be more consequential if it falls into the hands of an enemy who can study and copy it) and not the loss of a human life.

Related to these factors is the calculation regarding risk. Singer (2009) notes that, "The unmanning of [an] operation also means that the robot can take risks that a human wouldn't otherwise, risks that might mean fewer mistakes". He cites friendly fire incidents during the Kosovo campaign in 1999 in which the imperative to avoid loss of NATO pilots resulted in orders that planes not be flown at altitudes below 15,000 feet. One of the most grievous errors occurred when NATO planes flying at these altitudes bombed a convoy of buses carrying Kosovar refugees mistakenly identifying them as a convoy of Serbian tanks. Singer notes that the "removal of risk allows decisions to be made in a more deliberate manner". For soldiers fighting in cities, one of the toughest problems is to burst into a building and, in a matter of milliseconds, figure out who is an enemy and who is a civilian. In this situation, a robot that can enter a room and shoot only at someone who shoots first has a distinct advantage over the human who must take fire and somehow instantly manage to determine the source, return fire and avoid hitting any civilians.

Another advantage that robots have in situations of combat is that they do not suffer from human emotions of rage against adversaries who have caused harm to or death of a comrade. We have learned of many episodes where otherwise good individuals have given way to extreme emotion and committed atrocities after experiencing the loss of or grievous harm to someone with whom they have bonded and upon whom they have depended in situations of peril. Surely, eliminating the danger of such episodes is an important advantage favoring robotic agents over humans.

What could possibly be the downside of the use of robotic weapons? These may be more subtle and harder to see, but in a certain sense, the disadvantages of these weapons are identical with their advantages. One disadvantage, clearly recognized by those in command positions in the military, is that over a long time and haltingly, we have negotiated barriers against barbaric behavior in war. The Geneva Conventions and treaties barring the use of chemical and biological weapons are among these barriers. When one side in a conflict has overwhelming technological superiority, when there is marked asymmetry in the resources each brings to battle, there is an inescapable lessening of the respect that each side owes the other out of recognition of the parity of the risks the combatants share. The sense that the weaker forces can be eradicated like insects by the "magic" of advanced technology acts, in a mutually reinforcing manner, on both sides to undercut the restraints erected against barbarity (Singer, 2010).

Perhaps the most serious disadvantage of robotic weapons has to do with another set of barriers. General Robert E. Lee, commander of the Confederate forces in the American Civil War of the nineteenth century – the war which still holds pride of place as the bloodiest in US history – wrote, "It is good that we find war so horrible, or else we would become fond of it" (Singer, 2010). The act of declaring war is or should be a grave existential decision for any country. But we have seen, perhaps most notably in the case of the ill-considered invasion of Iraq by the USA, how consciousness of technological superiority lowers the barrier against waging war. Paradoxically, to the extent that

atrocities committed by otherwise decent soldiers remind us of the horror of war, they serve as a warning to anyone contemplating "loosing the dogs of war".

4. Keeping humans "in the loop"

This section is the easiest to write and the most frightening. When it comes to giving robotic weapons the capacity to kill, official military policy seems to be very clear and emphatic: "Humans must be kept in the loop". The meaning of this is, or should be, that a human must give authorization before any robotic weapon can fire on a human target. In fact, however, whenever this matter is raised in serious discussion, the result, according to Singer (2009), is averted eyes and a change in topic. The reasons for this are also clear. Although the ideal is to keep humans in the [command] loop, there are so many factors militating against this that in practice, it seems unworkable. Why, if there is risk of loss of life on your side in the interval between identification of a lethal threat and authorization to fire issued by a human controller, should the robotic weapon not be given the capability to fire immediately upon locating the threat? As the authorization requires communication between controller and weapon, and this communication can be cut or disrupted by the enemy, why should there not be an emergency back-up capability for the weapon to operate autonomously in this situation?

Singer points out that the logic of human control of robotic weapons seems to demand a many-one correspondence between weapons and controllers. But humans are notoriously ill equipped and unreliable for the task of controlling multiple units at one time, even under relatively calm conditions. A Pentagon-funded report notes that, "Even if the tactical commander is aware of the location of all his units, the combat is so fluid and fast-paced that it is very difficult to control them" (Singer, 2009).

Human control of automated weapons systems has already been seriously compromised by the human tendency to "believe what the computer says". The paradigmatic example of this is the case of the downing of Iran Air flight 655 over the Persian Gulf in July 1988 by an American naval vessel patrolling the gulf during the Iran-Iraq war (Wikipedia article, 2014). In an earlier paper (Fleischman, 2013), I pointed to another aspect of this story that merits attention. An error that should not have evaded the eye of even an undergraduate software engineering student was one of the primary factors that led to the mistaken characterization of flight 655. In the process of coordinating data on the radars of the three ships in the patrol, a tag used within the previous hour to label a (friendly) fighter jet making a landing (thus descending) was reassigned as the label for Iran Air Flight 655 (allowing for the mistaken reading of an attack profile) on the radars of the Vincennes (Iran Air Flight 655, 2014). So while it is not entirely inaccurate to think of this as an illustration of the way in which humans defer to the "judgment" of computer-controlled systems, it is just as relevant to see this as a warning against placing too much trust in the reliability of even "state-of-the-art" software engineering.

5. Compensating for the human "out of the loop"

If the superior capabilities of robotic weapons and the limitations of humans acting as controllers so far compromise the military principle of always keeping the human in the loop, then perhaps, we can substitute an "ethical governor" implemented in software for the absent human controller. Properly programmed, weapons acting autonomously could perhaps be constrained to "act ethically in war", observing all the articles of the

Geneva Conventions, the laws of war, and the locally relevant rules of engagement. In addition, as they are not subject to the psychological and emotional stresses that affect human combatants, we might even expect that they would act more morally than the human soldiers whose combat roles they assume.

The National Science Foundation (NSF) and US Government agencies associated with the Department of Defense have funded a project with precisely this aim. Ronald Arkin, director of the project, claims that:

> Ultimately these systems could have more information to make wiser decisions than a human could make. Some robots are already stronger, faster, and smarter than humans. We want to do better than people, to ultimately save more lives (Singer, 2010).

We should take note of the logical trap here: To "save more lives", it is necessary to develop weapons that make it easier to take the decision to go to war. That is to say, reduce killing in retail by making more likely killing in wholesale.

In this light, I think it is important to ask, "What is the purpose of the NSF in funding this 'research?'" Why should anyone want to do this? One possible motivation is as a salve to the consciences of those who are participating in and drawing public funds from the Department of Defense and the NSF in research that they know to be, in the last analysis, destructive and anti-human. We are building these lethal autonomous robotic weapons, but they are going to be "stronger, faster and smarter than humans". We are going to do better than mere humans and we will save many lives. We believe (or convince ourselves that), we can achieve this chimera, and therefore, we *must* try (and, of course, inure ourselves to the burden of accepting the public's money in furtherance of this grotesque delusion.)

In an earlier paper (Fleischman, 2013), I argued that the software engineering project to develop an "ethical governor" as envisioned by Arkin is a fantasy. The context of actual combat is one of such fluidity and rapid change as to defy the simple description of "a specific [i.e. closed] context" in which one might attempt to construct a formal system that correctly incorporates all aspects of moral reasoning. Gerdes and Øhrstrom (2013) conclude that such a system would:

> […] in principle require a complete description not only of all relevant moral rules and laws but also of all relevant aspects of the situation in question. However, having such descriptions is tantamount to having a God's eye view of all relevant aspects of reality.

In another recent paper, Englert *et al.* (2014) present a series of *Gedankenexperiments* that further indicate the intractability of the algorithmic approach to "moral behavior of" machines on a complex battlefield. They:

> […] construct an (admittedly artificial but) fully deterministic situation where a robot is presented with two choices: one morally clearly preferable over the other – yet based on the undecidability of the Halting Problem, it provably cannot decide algorithmically which one.

In addition, Asaro (2009) asserts that the Laws of Armed Conflict, Just War Theory and Rules of Engagement taken together constitute the following:

> […] a hodge-podge of laws, rules, heuristics, and principles, all subject to interpretation and value judgments. These rules do have an important role in regulating the conduct of individuals and institutions, mostly because they require people to think about the ethical

implications of their actions in certain ways, rather than dictating to them a specific action in a specific situation.

All such arguments, however carefully reasoned, miss the main point. If we succeed in reducing warfare to an exchange of "ethically justifiable" automatized gestures, we will have "domesticated" warfare as a chronic condition of human life. To anyone who objects that this is the current condition of political life on the international stage, let us go one step further: it will *justify* war as a chronic condition of human life.

The similarity of this case with that of the Strategic Defense Initiative (the so-called "Star Wars" project), from which David L. Parnas withdrew in a well-known letter and series of critical papers (Parnas, 1985), suggests that the appropriate response of computer scientists of good conscience toward Arkin's project or any other claiming to have the purpose of devising an "ethical governor" for autonomous robotic weapons should be to condemn and refuse to participate in such an undertaking.

6. The academy of moral sciences of the island kingdom of Laputa discusses robotic weapons

The reference, of course, is to the Third Voyage of Jonathan Swift's superb satire, *Gulliver's Travels*. During the course of this voyage, having been attacked by pirates and marooned near a small chain of islands "in the latitude of 46 N. and longitude of 183", Gulliver is rescued by and makes the acquaintance of the inhabitants of the flying island of Laputa (Swift, 1727). These are a race of people so given over to intense scientific and philosophic speculation that they live – out of indifference – in houses without right angles, and are dressed – out of indifference – in ill-fitting clothes. They walk in the manner of somnambulists, accompanied by servants who deliver, from time to time, gentle blows upon the mouth and ears of their masters with a sort of bladder filled with dried peas to awaken them from their all-absorbing speculations about abstract and unworldly matters. Reading the passages that follow, I sometimes felt myself Gulliver's companion on this voyage, privy to the inner thoughts of the deep-thinking Laputans:

> The moral support for a ban on the deployment of any autonomous robotic weapons depends entirely on whether it is decided that there is a human right not to be the target of a robotic weapon […] We were unable to come to a full conclusion on that concept. The precautionary principle would suggest that until we do, a ban is justified. But if a supra human robotic moral agent or a good moral reasoning augmentation system of the sort that Arkin proposes with his ethical governor is indeed developed, then it would actually be immoral not to deploy robotic weapons so constructed (Sullins, 2013).

> Autonomous weapons systems are entering the battlefields of the future, but they are doing so one small automated step at a time. The steady march of automation is frankly inevitable, in part because it is not merely a feature of weapons technology, but of technology generally […] Highlighting that the incremental automation of some weapon systems in some form is inevitable is not a veiled threat in the guise of a prediction. Nor is it meant to suggest that the path of these technologies is beyond rationally ethical human control (Anderson and Waxman, 2013).

> Where in this long history of new weapons and attempts to regulate them ethically and legally will autonomous weapons fit? What are the features of autonomous robotic weapons that raise ethical and legal concerns? […] One answer to these questions is to wait and see: it is too early

to know where the technology will go, so the debate over ethical and legal principles for robotic autonomous weapons should be deferred until a system is at hand (Anderson and Waxman, 2013).

If one is a lawyer in a ministry of defense somewhere in the world, whose job is to evaluate the lawfulness of such weapon systems, including where and under what operational conditions they can lawfully be used, it will be indispensable to be able to test each system to know what it can and cannot do and under what circumstances (Anderson and Waxman, 2013).

Programming the laws of war at their conceptually most difficult (sophisticated proportionality, for example) is a vital research project over the long run, in order to find the greatest gains that can be had from machine decision-making within the law (Anderson and Waxman, 2013).

Excessive devotion to individual criminal liability as the presumptive mechanism of accountability risks blocking development of machine systems that might, if successful, reduce actual harms to soldiers as well as to civilians on or near the battlefield. […] It would be unfortunate to sacrifice real-world gains consisting of reduced battlefield harm through machine systems (assuming there are any such gains) simply in order to satisfy an a priori principle that there must always be a human to hold accountable (Anderson and Waxman, 2013).

We must strike a delicate balance between the ability to effectively execute mission objectives and the absolute compliance that the Laws of War will be observed. […] To address these problems for a robotic implementation, normally we would turn to neuroscience and psychology to assist in the determination of an architecture capable of ethical reasoning. This paradigm has worked well in the past […] Relatively little is known, however, about the specific processing of morality by the brain from an architectural perspective […] (Arkin, 2009).

Gazzaniga postulates that moral ideas are generated by an interpreter located in the left hemisphere of our brain that creates and supports beliefs. Although this may be useful for providing an understanding for the basis of human moral decisions, it provides little insight into the question that we are most interested in, i.e. how, once a moral stance is taken, is that enforced upon an underlying architecture or control system. The robot need not derive the underlying moral precepts; it needs solely to apply them. Especially in the case of a battlefield robot (but also for a human soldier), we do not want the agent to be able to derive its own beliefs regarding the moral implications of the use of lethal force, but rather to be able to apply those that have been previously derived by humanity as prescribed in the LOW and ROE (Arkin, 2009).

The focus for the reactive ethical architectural component for ethical behavioral control will not involve emotion directly, however, as that has been shown to impede the ethical judgment of humans in wartime (Arkin, 2009).

[…] it is a thesis of my ongoing research for the USA Army that robots not only can be better than soldiers in conducting warfare in certain circumstances, but they also can be more humane in the battlefield than humans (Arkin, 2009).

But only with the caveat that we take seriously the claim that these weapons could also be significant tools for complying with *jus in bello* and that it would be immoral to limit them if that were the case. An accurate answer to that last question requires much more research. As we have seen in this paper there are many arguments both pro and con on this issue, but we also have the potential of settling this case with information gathered from an analysis of the last two decades of the use of telerobotic and semi-autonomous drones on the battlefield and in covert actions. This kind of research will have to wait for all of these reports to become

declassified but over time they will and we will be able to say with much more certainty whether or not this technology has contributed to a more just and moral world (Sullins, 2013).

7. A brief gloss on the preceding section

In presenting the foregoing quotations, the intention is simply to allow the authors' own words to reveal the absurdity (Arkin), dishonest sophistry (Anderson and Waxman) and sweet confusion (Sullins) of their arguments and, thereby, permit them to sink of their own weight. (At the same time, I would like to record my dismay at the complaisance of the reviewers and editors who permitted the elements of this *catalogue irraisonné* to pass unchallenged into the scholarly literature.)

Anderson and Waxman, for example, counsel patience and advance the proposal that each autonomous lethal system undergo testing "to know what it can and cannot do and under what circumstances". However, they conveniently omit any suggestion as to what might constitute an appropriate platform to test these systems under conditions of war. The coy disclaimer of any concealed threat in their prediction of the inevitability of the introduction of autonomous killing weapons in the battlefield "one small automated step at a time", should be seen for what it patently is: an ill-disguised attempt to pre-empt the legitimacy and good sense of an international effort to ban such weapons. A vigorous rebuttal to this inelegant *petitio principii* is presented in Noel Sharkey's paper, "The Evitability of Autonomous Robot Warfare" (Sharkey, 2012).

Finally, and most destructively, Anderson and Waxman breathtakingly banish accountability from the field in their plea that we not:

> […] sacrifice real-world gains consisting of reduced battlefield harm through machine systems (assuming there are any such gains) simply in order to satisfy an a priori principle that there must always be a human to hold accountable (Anderson and Waxman, 2013).

Sullins, like Anderson and Waxman, counsels patience as we wait, perhaps decades, for the declassification of reports on research into the effects of robotic weapons systems. He is to be commended for the care with which he hedges all of his judgments. But to the extent that they offer comfort to those who wish to proceed in haste with the development of autonomous killing weapons, one might still wish that he had adjoined to his exquisitely qualified and conditioned conclusions the possibility that the development of these weapons, as in instances we know too well (e.g. the atomic and hydrogen bombs), might eventually lead to a severe case of buyer's remorse.

For Arkin, apparently, for the purposes of conduct on the battlefield, moral ideas are no longer "generated by an interpreter located in the left hemisphere of our brain", but rather issue from the barrel of a gun fired by means of an algorithm devised in conformance with a predetermined moral stance formulated, one supposes, by a caring, Christian computer scientist. And, of course, we no longer depend on human agents, as autonomous killing weapons "not only can be better than soldiers in conducting warfare in certain circumstances, but they also can be more *humane* [emphasis added] in the battlefield than humans" (Arkin, 2009). Alas, I was under the false impression that the definition in my old dictionary in two volumes (A – Pocket Veto, Pockmark – Zymurgy), namely:

> Humane: Having or showing the feelings befitting a man, esp. with respect to other human beings or to the lower animals; characterized by tenderness and compassion for the suffering or distressed – *New Century Dictionary* (Emery and Brewster, 1927).

Was still valid. In fact, this is a symptom of one of the worst effects of the entire matter. Those who speak of "ethical governors" realized in software, of "supra human robotic moral agent[s]", are engaging in a form of semantic sleight of hand the ultimate consequence of which is to debase the deep meaning of words for the purpose of reducing human feeling, compassion and judgment to nothing more than the result of a computation.

In addition, one final comment on Arkin's unctuous hypocrisy. In a recent paper, "Lethal Autonomous Systems and the Plight of the Non-Combatant", he remarks:

> It must be noted that past and present trends in human behavior in the battlefield regarding adhering to legal and ethical requirements are questionable at best. Unfortunately, humanity has a rather dismal record in ethical behavior in the battlefield. […] How can we meaningfully reduce human atrocities on the modern battlefield? […] Can technology help solve this problem? I believe that simply being human is the weakest point in the kill chain […] (Arkin, 2013).

In the light of recent history in which hundreds of thousands of non-combatants were grievously affected by the unanticipated consequences of the disastrous USA's adventure in Iraq, one might ask why someone with the very noble aspirations of Ronald Arkin could fail to direct his attention to the recent US President and his advisors who, through arrogant overconfidence in the technological superiority they commanded, wrote a "rather dismal record in ethical behavior", unleashing this bloodshed and the resulting "human atrocities" with which we are rather too well-acquainted.

8. An instructive story

In mid-October of 1962, U2 surveillance photographs revealed the presence of missile sites and Soviet missile components on the island of Cuba, exposing as deception the assurances given by Andrei Gromyko, the Soviet Foreign Minister and Nikita Khrushchev, the leader of the Soviet Union, that no Soviet missiles would be installed in Cuba. This discovery precipitated the most dangerous episode of the Cold War, 15 days in which the two superpowers were on a path to war that would have involved attacks using nuclear weapons. The consequences of this were and are unimaginable.

The resolution of this crisis, brilliantly narrated by Jonathan Glover in *Humanity: A Moral History of the 20th Century*, relates the means by which Khrushchev and Kennedy managed to step back from the brink, in spite of strong forces on both sides that tended toward war and nuclear disaster (Glover, 2012). This riveting and illuminating story should command the attention of anyone considering the role of autonomous weapons in war.

The conditions surrounding the Cuban Missile Crisis recapitulate, in an eerie correspondence, the set of misjudgments, miscalculations and reckless actions that, in 1914, led the European powers into a war that can only be considered a disaster for those who fought and for the generation that survived. How, then, did the leaders of the two superpowers in 1962 avoid the trap? Among the factors, there were two worthy of reflection in the context of this paper.

The first is the publication earlier that year of historian Barbara Tuchman's *The Guns of August*, which carefully dissected European internal political pressures, misunderstandings, ambiguous signals, poor communication among allies and between potential belligerents and the military preparations, once begun, that seemed impossible to roll back. Both President Kennedy and his closest advisors had read the book and

referred to it during the meetings at which the possible responses to the Soviet threat were discussed. According to the memoirs of Robert Kennedy, quoted in Glover (2012), JFK spoke with his brother about the European leaders in 1914 saying, "they seemed to tumble into war through stupidity, individual idiosyncrasies, misunderstandings, and personal complexes of inferiority and grandeur". He said:

> I am not going to follow a course which will allow anyone to write a comparable book about this time, *The Missiles of October*. If anybody is around to write after this, they are going to understand that we made every effort to give our adversary room to move. I am not going to push the Russians an inch beyond what is necessary.

Of equal weight, on the Russian side, Khrushchev, early in the crisis, sent a letter to President Kennedy in which he wrote:

> Should war indeed break out, it would not be in our power to contain or stop it, for such is the logic of war. I have taken part in two wars, and I know that war ends only when it has rolled through cities and villages, sowing death and destruction everywhere […] If people do not display wisdom, they will eventually reach the point where they will clash like blind moles, and then mutual annihilation will commence […] You and I should not now pull on the ends of the rope in which you have tied a knot of war, because the harder you and I pull, the tighter this knot will become. And a time may come when the knot is tied so tight that the person who tied it is no longer capable of untying it, and then the knot will have to be cut (Glover, 2012).

Both the words of Nikita Khrushchev and the import of Barbara Tuchman's analysis that was present in the minds of John Fitzgerald Kennedy and his advisors resonated with the warning articulated by Robert E. Lee: "It is good that we find war so horrible, or else we would become fond of it". This was a crisis that, however it unfolded, both leaders understood would forever indelibly bear their signatures. Personal sense of responsibility and consciousness of the horrors of war were the factors that made it possible to pull back. Let us imagine the computer system, designed and implemented by individuals without names and without the wisdom of those who read and reflect and are conscious of the horror, let us indeed pause and imagine the system capable of the saving wisdom of Khrushchev and Kennedy.

9. The question of accountability and responsibility

The "Problem of Many Hands", articulated by Nissenbaum (1994), has become a commonplace. We routinely cite this problem by name, nod knowingly in acquiescence of the certainty that any large software engineering project is bound to have some unanticipated failure modes with serious negative consequences. If, as is customary, the project is developed over a significant period of time by a team the membership of which is not fixed, it will be difficult, perhaps impossible, to determine who is responsible for the failure, to say who should be held accountable for harms ultimately engendered in the use of the system. This is a fact of life in our technologically sophisticated world.

Perhaps we can agree that there may be circumstances where the expectation of future benefits from the development of a new technology justifies accepting the risks of negative consequences – without, however, relinquishing the understanding that, while we are waiting for the realization of such benefits, someone, some organization must be held accountable and accept responsibility for the harms. There are some areas of

application where we can agree to take these risks. But there are assuredly areas where this attitude is unjustifiable. The development of autonomous robotic killing weapons is one of them.

It is appropriate to recall Norbert Wiener's warning, cited in Weizenbaum (1976):

> An intelligent understanding of [a machine's] mode of performance may be delayed until long after the task which [it has] been set is completed. […] This means that, though machines are theoretically subject to human criticism, such criticism may be ineffective until long after it is relevant.

Whose name will be on the disaster precipitated by the malfunction of one of these weapons? Whose name will be attached, as Khrushchev and Kennedy were aware theirs would be to the nuclear disaster precipitated by a reckless gesture in the course of the Cuban Missile Crisis? Who will own the damage to what little of civilized culture we still imagine we possess? Certainly not computer scientists like Arkin, whose names will have long been forgotten. In a sense, this is appropriate. However, much their work contributes to this damage, the disaster will be ours as a society.

10. An alternative argument and concluding observations

Long ago, Joseph Weizenbaum cautioned against the seduction of technique applied to problems for which its application is utterly inappropriate:

> There are two kinds of computer applications that either ought not be undertaken at all, or if they are contemplated, should be approached with utmost caution. […] The first kind I would call simply obscene. These are ones whose very contemplation ought to give rise to feelings of disgust in every civilized person. […] I would put all projects that propose to substitute a computer system for human understanding, for a human function that involves interpersonal respect, understanding, and love in [this] category (Weizenbaum, 1976).

Here, I believe, is the foundation of the proper argument against the development and use of autonomous, lethal robotic weapons. Military officers are taught that there is a bond of mutual respect that unites opposing combatants in moral symmetry of mortal hazard. Certainly, in the decision to exercise lethal force against an enemy combatant, that respect proscribes the possibility of relegating to a machine the decision to take the life of another human. Even beyond the weight of this consideration is the respect one owes to one's own humanity. To renounce this responsibility, even in time of war, is to convert oneself into nothing more than a machine.

Finally, it is important to recognize that, although many profound thinkers have contributed to our understanding of what it means to act ethically, our ideas about ethical behavior are as much a product of our experience and our emotional wisdom as of our analytical intelligence. Beware the scientist or engineer who claims that technique alone will substitute for human instinct and wisdom and enable us to program a machine to behave ethically. Even to approximate this would require the solution of a software engineering problem of forbidding complexity. A moment's reflection on our discouraging experience with such systems should give us pause.

Again, I want to insist on the question, "Why should anyone *want* to do this?" In the words of Joseph Weizenbaum:

> Technological inevitability can thus be seen to be a mere element of a much larger syndrome. Science promised man power. But, as so often happens when people are seduced by promises of power, the price exacted in advance and all along the path, and the price actually paid, is

> servitude and impotence. Power is nothing if it is not the power to choose. Instrumental reason can make decisions, *but there is all the difference between deciding and choosing* (Weizenbaum, 1976, [emphasis added]).

The proposal to develop and deploy autonomous robotic weapons presents us with a moment in which our humanity is in play. What is it that we are choosing when we choose to develop the ability to make war "in a way that is better than the way humans wage war?"

In addition, it is worth repeating that, over a long time and haltingly, we have negotiated barriers against barbaric and inadmissible behavior in waging war. The Geneva Conventions and treaties barring the use of chemical and biological weapons are among these barriers. More recently, the Ottawa Treaty banning the use of anti-personnel land mines (devices that can be thought of as passive autonomous lethal weapons) has been ratified by 162 countries (List of Parties to the Ottawa Treaty, 2014). Although the refusal of various rogue states, including North Korea, China and the USA, to ratify the Ottawa Treaty forearms us with a sense of the difficulty of the undertaking, it is, at the same time, a proof of concept and a spur to the formulation and adoption of an international treaty banning the development and deployment of lethal autonomous robotic weapons.

Beyond this, in choosing to invest in the chimerical pursuit of the ability to build machines that can "do better than people" at waging war, we are distorting the priorities on which a civilized society should rest. In the USA, we seem unable to make a commitment to educating or providing adequate health care for all the children who live among us, but we find it easy to lavish great sums in the pursuit of an obscenity, oblivious to the warning, "It is good that we find war so horrible, or else we would become fond of it".

References

Anderson, K. and Waxman, M.C. (2013), "Law and ethics for autonomous weapon systems: why a ban won't work and how the laws of war can", *Columbia Public Law Research Paper*, American University Washington College of Law Research Paper No. 2013-2111, Washington, DC.

Arkin, R. (2009), *Governing Lethal Behavior in Autonomous Robots*, CRC Press, Taylor and Francis Group, Boca Raton.

Arkin, R. (2013), "Lethal weapons and the plight of the non-combatant", *AISB Quarterly*, Vol. 12 No. 137, pp. 1-9.

Asaro, P. (2009), "Modeling the moral user", *IEEE Technology and Society Magazine*, Vol. 28 No. 1, pp. 20-24.

Asaro, P. (2012), "On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making", *International Review of the Red Cross*, Vol. 94 No. 886, pp. 687-709.

Emery, H.G. and Brewster, H.K. (Eds) (1927), *The New Century Dictionary of the English Language*, D. Appleton-Century Company, New York, NY.

Englert, M., Sandra, S. and Martin, Z. (2014), *Logical Limitations to Machine Ethics with Consequences to Lethal Autonomous Weapons*, Cornell University Library, Computers and Society, at arXiv:1411.2842v1 [cs.CY].

Fleischman, W. (2013), "Why we should not build lethal autonomous robotic weapons", Proceedings of CACIC 2013, XVII Congreso Argentino de las Ciencias de Computación, Universidad Caece, Mar del Plata.

Gerdes, A. and Øhrstrom, P. (2013), "Preliminary reflections on a moral turing test", *Proceedings of ETHICOMP 2013, The Possibilities of Ethical ICT*, University of Southern Denmark, Kolding, pp. 167-174.

Glover, J. (2012), *Humanity: A Moral History of the 20th Century*, 2nd ed., Yale University Press, New Haven.

Iran Air Flight 655 (2014), in *Wikipedia*, available at: http://en.wikipedia.org/wiki/Iran_Air_Flight_655 (accessed 24 March 2014).

List of Parties to the Ottawa Treaty (2014), in *Wikipedia*, available at: http://en.wikipedia.org/wiki/List_of_parties_to_the_Ottawa_Treaty#Non-signatory_states (accessed 24 December 2014).

Nissenbaum, H. (1994), "Computing and accountability", *Communications of the ACM*, Vol. 37 No. 1, pp. 73-80.

Parnas, D.L. (1985), "Letter to James H. Offutt", *Introduction to Computer Ethics, Parts 1 and 2*, available at: http://web.archive.org/web/20130626112228 www.stanford.edu/class/cs181/materials/CS181-Parts1and2.pdf (accessed 30 March 2014).

Sharkey, N. (2012), "The evitability of autonomous robot warfare", *International Review of the Red Cross*, Vol. 94 No. 886.

Singer, P.W. (2009), *Wired for War: The Robotics Revolution and Conflict in the 21st Century*, Penguin Press, New York, NY.

Singer, P.W. (2010), "The ethics of killer applications: why is it so hard to talk about morality when it comes to new military technology", *Journal of Military Ethics*, Vol. 9 No. 4, pp. 299-312.

Singer, P.W. (2011), "Military robotics and ethics: a world of killer apps", *Nature*, Vol. 477 No. 7365, pp. 399-401.

Sullins, J.P. (2013), "An ethical analysis of the case for robotic weapons arms control", in Podins, K., Stinissen, J. and Maybaum, M. (Eds), *Fifth International Conference on Cyber Conflict*, *KNATO CCD COE Publications*, Tallinn.

Swift, J. (1727), "Travels into several remote nations of the world", in Lemuel, G. (Ed.), *Faithfully Abridged*, printed for Stone, J. and King, J., London.

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

Corresponding author
William M. Fleischman can be contacted at: william.fleischman@villanova.edu