Lethal Autonomous Systems and the Plight of the Non-combatant

Ronald Arkin *

It seems a safe assumption, unfortunately, that humanity will persist in conducting warfare, as evidenced over all recorded history. New technology has historically made killing more efficient, for example with the invention of the longbow, artillery, armored vehicles, aircraft carriers, or nuclear weapons. Many view that each of these new technologies has produced a Revolution in Military Affairs (RMA), as they have fundamentally changed the ways in which war is waged. Many now consider robotics technology a potentially new RMA, especially as we move towards more and more autonomous[1] systems in the battlefield.

Robotic systems are now widely present in the modern battlefield, providing intelligence gathering, surveillance, reconnaissance, and target acquisition, designation and engagement capabilities. Limited autonomy is also present or under development in many systems as well, ranging from the Phalanx system capable of "capable of autonomously performing its own search, detect, evaluation, track, engage and kill assessment functions"[2], fire-and-forget munitions, loitering torpedoes, and intelligent antisubmarine or anti-tank mines among numerous other examples. Continued advances in autonomy will result in changes involving tactics, precision, and just perhaps, if done correctly, a reduction in atrocities as outlined in research conducted at the Georgia Tech Mobile Robot Laboratory (GT-MRL)[3]. This paper asserts that it may be possible to ultimately create intelligent autonomous robotic military systems that are capable of reducing civilian casualties and property damage when compared to the performance of human warfighters. Thus, it is a contention that calling for an outright

* Ronald Arkin is Regents' Professor, Director of the Mobile Robot Laboratory, and Associate Dean for Research in the College of Computing at the Georgia Institute of Technology. This work was supported in part by the U.S. Army Research Office under Contract #W911NF-06-1-0252. Small portions of this essay appeared earlier in a Viewpoint article by the author appearing in the *Journal of Industrial Robots* 38:5, 2011, and from a more comprehensive treatment of the subject in the author's book *Governing Lethal Behavior in Autonomous Systems*, Taylor-Francis, 2009, and are included with permission.

[1] We do not use autonomy in the sense that a philosopher does, i.e., possessing free will and moral agency. Rather we use in this context a roboticists' definition – the ability to designate and engage a target without additional human intervention after having been tasked to do so.

[2] U.S. Navy, "Phalanx Close-in Weapons Systems", United States Navy Factfile, http://www.navy.mil/navydata/fact_display.asp?cid=2100&tid=800&ct=2, accessed 8/2008.

[3] R.C. Arkin, *Governing Lethal Behavior in Autonomous Robots*, Chapman-Hall, 2009.

ban on this technology is premature, as some groups already are doing[4]. Nonetheless, if this technology is to be deployed, then restricted, careful and graded introduction into the battlefield of lethal autonomous systems must be standard policy as opposed to haphazard deployments, which I believe is consistent with existing International Humanitarian Law (IHL).

Multiple potential benefits of intelligent war machines have already been declared by the military, including: a reduction in friendly casualties; force multiplication; expanding the battlespace; extending the warfighter's reach; the ability to respond faster given the pressure of an ever increasing battlefield tempo; and greater precision due to persistent stare [constant video surveillance that enables more time for decision making and more eyes on target]. This argues for the inevitability of development and deployment of lethal autonomous systems from a military efficiency and economic standpoint, unless limited by IHL.

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. Potential explanations for the persistence of war crimes include[5]: high friendly losses leading to a tendency to seek revenge; high turnover in the chain of command leading to weakened leadership; dehumanization of the enemy through the use of derogatory names and epithets; poorly trained or inexperienced troops; no clearly defined enemy; unclear orders where intent of the order may be interpreted incorrectly as unlawful; youth and immaturity of troops; external pressure, e.g., for a need to produce a high body count of the enemy; and pleasure from power of killing or an overwhelming sense of frustration. There is clear room for improvement and autonomous systems may help address some of these problems.

Robotics technology, suitably deployed may assist with the plight of the innocent noncombatant caught in the battlefield. If used without suitable precautions, however, it could potentially exacerbate the already existing violations by human soldiers. While I have the utmost respect for our young men and women warfighters, they are placed into conditions in

[4] Notably Human Rights Watch, International Committee on Robot Arms Control (ICRAC) and Article 36.

[5] Bill, B. (Ed.), *Law of War Workshop Deskbook*, International and Operational Law Department, Judge Advocate General's School, June 2000; Danyluk, S., "Preventing Atrocities", *Marine Corps Gazette*, Vol. 8, No. 4, pp. 36-38, Jun 2000; Parks, W.H., "Crimes in Hostilities. Part I", *Marine Corps Gazette*, August 1976; Parks, W.H., "Crimes in Hostilities. Conclusion", Marine Corps Gazette, September 1976; Slim, H., *Killing Civilians: Method, Madness, and Morality in War*, Columbia University Press, New York, 2008.

modern warfare under which no human being was ever designed to function. In such a context, expecting a strict adherence to the Laws of War (LOW) seems unreasonable and unattainable by a significant number of soldiers[6]. Battlefield atrocities have been present since the beginnings of warfare, and despite the introduction of International Humanitarian Law (IHL) over the last 150 years or so, these tendencies persist and are well documented,[7] even more so in the days of CNN and the Internet. 'Armies, armed groups, political and religious movements have been killing civilians since time immemorial.'[8] 'Atrocity... is the most repulsive aspect of war, and that which resides within man and permits him to perform these acts is the most repulsive aspect of mankind'.[9] The dangers of abuse of unmanned robotic systems in war, such as the Predator and Reaper drones, are well documented; they occur even when a human operator is directly in charge.[10]

Given this, questions then arise regarding if and how these new robotic systems can conform as well as, or better than, our soldiers with respect to adherence to the existing IHL. If achievable, this would result in a reduction in collateral damage, i.e., noncombatant casualties and damage to civilian property, which translates into saving innocent lives. If achievable this could result in a moral requirement necessitating the use of these systems. Research conducted in our laboratory[11] focuses on this issue directly from a design perspective. No claim is made our research provides a fieldable solution to the problem, far from it. Rather these are baby-steps towards achieving such a goal, including the development of a prototype proof-of-concept system tested in simulation. Indeed, there may be far better approaches that the one we currently employ, if the research community can focus on the plight of the noncombatant and how technology may possibly ameliorate the situation.

---

[6] Surgeon General's Office, Mental Health Advisory Team (MHAT) IV Operation Iraqi Freedom 05-07, Final Report, Nov. 17, 2006.

[7] For a more detailed description of these abhorrent tendencies of humanity discussed in this context, see Arkin, R.C., "The Case for Ethical Autonomy in Unmanned Systems", *Journal of Military Ethics*, 9:4, pp. 332-341, 2010.

[8] Slim, H., *Killing Civilians: Method, Madness, and Morality in War*, Columbia University Press, New York, 2008, p. 3.

[9] Grossman, D., *On Killing: The Psychological Cost of Learning to Kill in War and Society*, Little, Brown and Company, Boston, 1995, p.229.

[10] Adams. J., "US defends unmanned drone attacks after harsh UN Report", *Christian Science Monitor*, June 5, 2010; Filkins, D., "Operators of Drones are Faulted in Afghan Deaths", *New York Times*, May 29, 2010; Sullivan, R., "Drone Crew Blamed in Afghan Civilian Deaths", *Associated Press*, May 5, 2010.

[11] For more information see Arkin, R.C., *Governing Lethal Behavior in Autonomous Systems,* Taylor and Francis, 2009.

As robots are already faster, stronger, and in certain cases (e.g., Deep Blue, Watson[12]) smarter than humans, is it really that difficult to believe they will be able to ultimately treat us more humanely in the battlefield than we do each other, given the persistent existence of atrocious behaviors by a significant subset of human warfighters?

Why technology can lead to a reduction in casualties on the battlefield

Is there any cause for optimism that this form of technology can lead to a reduction in non-combatant deaths and casualties? I believe so, for the following reasons.

- The ability to act conservatively: i.e., they do not need to protect themselves in cases of low certainty of target identification. Autonomous armed robotic vehicles do not need to have self-preservation as a foremost drive, if at all. They can be used in a self-sacrificing manner if needed and appropriate without reservation by a commanding officer. There is no need for a 'shoot first, ask-questions later' approach, but rather a 'first-do-no-harm' strategy can be utilized instead. They can truly assume risk on behalf of the noncombatant, something that soldiers are schooled in, but which some have difficulty achieving in practice.
- The eventual development and use of a broad range of robotic sensors better equipped for battlefield observations than humans currently possess. This includes ongoing technological advances in electro-optics, synthetic aperture or wall penetrating radars, acoustics, and seismic sensing, to name but a few. There is reason to believe in the future that robotic systems will be able to pierce the fog of war more effectively than humans ever could.
- Unmanned robotic systems can be designed without emotions that cloud their judgment or result in anger and frustration with ongoing battlefield events. In addition, 'Fear and hysteria are always latent in combat, often real, and they press us toward fearful measures and criminal behavior'[13]. Autonomous agents need not suffer similarly.
- Avoidance of the human psychological problem of 'scenario fulfillment' is possible. This phenomenon leads to distortion or neglect of contradictory information in

[12] http://en.wikipedia.org/wiki/Deep_Blue_(chess_computer),http://en.wikipedia.org/wiki/Watson_(computer)
[13] Walzer, M., *Just and Unjust Wars,* 4th ed., Basic Books, 1977.

stressful situations, where humans use new incoming information in ways that only fit their pre-existing belief patterns. Robots need not be vulnerable to such patterns of premature cognitive closure. Such failings are believed to have led to the downing of an Iranian airliner by the USS Vincennes in 1988.[14]

- Intelligent electronic systems can integrate more information from more sources far faster before responding with lethal force than a human possibly could in real-time. These data can arise from multiple remote sensors and intelligence (including human) sources, as part of the Army's network-centric warfare concept and the concurrent development of the Global Information Grid. 'Military systems (including weapons) now on the horizon will be too fast, too small, too numerous and will create an environment too complex for humans to direct'[15].
- When working in a team of combined human soldiers and autonomous systems as an organic asset, they have the potential capability of independently and objectively monitoring ethical behavior in the battlefield by all parties, providing evidence and reporting infractions that might be observed. This presence alone might possibly lead to a reduction in human ethical infractions.

Addressing some of the counter-arguments

But there are many counterarguments as well. These include the challenge of establishing responsibility for war crimes involving autonomous weaponry, the potential lowering of the threshold for entry into war, the military's possible reluctance of giving robots the right to refuse an order, proliferation, effects on squad cohesion, the winning of hearts and minds, cybersecurity, proliferation, and mission creep.

There are good answers to these concerns I believe, and are discussed elsewhere in my writings[16]. If the baseline criteria becomes outperforming humans in the battlefield with respect to adherence to IHL (without mission performance erosion), I consider this to be ultimately attainable, especially under situational conditions where bounded morality [narrow,

[14] Sagan, S., "Rules of Engagement", in *Avoiding War: Problems of Crisis Management* (Ed. A. George), Westview Press, 1991.

[15] Adams, T., "Future Warfare and the Decline of Human Decisionmaking", in *Parameters*, U.S. Army War College Quarterly, Winter 2001-02, pp. 57-71.

[16] E.g., Arkin, R.C, *op. cit.,* 2009.

highly situation-specific conditions] applies[17], but not soon and not easily. The full moral faculties of humans need not be reproduced to attain to this standard. There are profound technological challenges to be resolved, such as effective *in situ* target discrimination and recognition of the status of those otherwise *hors de combat*, among many others. But if a warfighting robot can eventually exceed human performance with respect to IHL adherence, that then equates to a saving of noncombatant lives, and thus is a humanitarian effort. Indeed if this is achievable, there may even exist a moral imperative for its use, due to a resulting reduction in collateral damage, similar to the moral imperative Human Rights Watch has stated with respect to precision guided munitions when used in urban settings[18]. This seems contradictory to their call for an outright ban on lethal autonomous robots[19] before determining via research if indeed better protection for non-combatants could be afforded.

Let us not stifle research in the area or accede to the fears that Hollywood and science fiction in general foist upon us. By merely stating these systems cannot be created to perform properly and ethically does not make it true. If that were so, we would not have supersonic aircraft, space stations, submarines, self-driving cars and the like. I see no fundamental scientific barriers to the creation of intelligent robotic systems that can outperform humans with respect to moral behavior. The use and deployment of ethical autonomous robotic systems is not a short-term goal for use in current conflict, typically counterinsurgency operations, but rather will take considerable time and effort to realize in the context of interstate warfare and situational context involving bounded morality.

A plea for the noncombatant

How can we meaningfully reduce human atrocities on the modern battlefield? Why is there persistent failure and perennial commission of war crimes despite efforts to eliminate them through legislation and advances in training? Can technology help solve this problem? I believe that simply being human is the weakest point in the kill chain, i.e., our biology works against us in complying with IHL. Also the oft-repeated statement that "war is an inherently human endeavor" misses the point, as then atrocities are also an inherently human endeavor,

[17] Wallach, W. and Allen, C., *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, 2010.

[18] Human Rights Watch, "International Humanitarian Law Issues in the Possible U.S. Invasion of Iraq", *Lancet*, Feb. 20, 2003.

[19] Human Rights Watch, "Losing Humanity: The Case Against Killer Robots", Nov. 19, 2012.

and to eliminate them we need to perhaps look to other forms of intelligent autonomous decision-making in the conduct of war. Battlefield tempo is now outpacing the warfighter's ability to be able to make sound rational decisions in the heat of combat. Nonetheless, I must make clear the obvious statement that peace is unequivocally preferable to warfare in all cases, so this argument only applies when human restraint fails once again, leading us back to the battlefield.

While we must not let fear and ignorance rule our decisions regarding policy towards these new weapons systems, we nonetheless must proceed cautiously and judiciously. It is true that this emerging technology can lead us into many different futures, some dystopian. It is crucially important that we not rush headlong into the design, development, and deployment of these systems without thoroughly examining their consequences on all parties: friendly forces, enemy combatants, civilians, and society in general. This can only be done through reasoned discussion of the issues associated with this new technology. Toward that end, I support the call for a moratorium to ensure that such technology meets international standards before being considered for deployment as exemplified by the recent report from the United Nations Special Rapporteur on Extrajudicial, Summary, or Arbitrary Executions.[20] In addition, the United States Department of Defense has recently issued a directive[21] restricting the development and deployment of certain classes of lethal robots, which appears tantamount to a quasi-moratorium.

Is it not our responsibility as scientists and citizens to look for effective ways to reduce man's inhumanity to man through technology? Where is this more evident than in the battlefield? Research in ethical military robotics can and should be applied toward achieving this end. The advent of these systems, if done properly, could possibly yield a greater adherence to the laws of war by robotic systems than from using soldiers of flesh and blood alone. While I am not averse to the outright banning of lethal autonomous systems in the battlefield, If these systems were properly inculcated with a moral ability to adhere to the laws of war and rules of engagement, while ensuring that they are used in narrow bounded military situations as adjuncts to human warfighters, I believe they could outperform human soldiers with respect to conformance to IHL. The end product then could be, despite the fact that these systems could

[20] Christof Heyns, *Report of the Special Rapporteur on Extrajudicial, Summary, and Arbitrary Execution*, United Nations Human Rights Council, 23rd Session, April 9, 2013.

[21] United States Department of Defense Directive Number 3000.09, Subject: Autonomy in Weapons Systems, November 21, 2012.

not ever be expected to be perfectly ethical, a saving of noncombatant lives and property when compared to human war fighters' behavior.

This is obviously a controversial assertion, and I have often stated that the discussion my research engenders on this subject is as important as the research itself. We must continue to examine the development and deployment of lethal autonomous systems in forums such as the United Nations and the International Committee of the Red Cross to ensure that the internationally agreed upon standards regarding the way in which war is waged are adhered to as this technology proceeds forward. If we ignore this, we do so at our own peril.

The Way Forward?

It clearly appears that the use of lethality by autonomous systems is inevitable, perhaps unless outlawed by international law – but even then enforcement seems challenging. But as stated earlier, these systems already exist: the Patriot missile system, the Phalanx system on Aegis class cruisers, anti-tank mines, and fire-and-forget loitering munitions all serve as examples. A call for a ban on these autonomous systems may have as much success as trying to ban artillery, cruise missiles, or aircraft bombing and other forms of standoff weaponry (even the crossbow was banned by Pope Innocent II in 1139[22]). A better strategy perhaps is to try and control its uses and deployments, which existing IHL appears at least at first glance to adequately cover, rather than a call for an outright ban, which seems unenforceable even if enacted.

The horse is out of the barn. Under current IHL, these systems cannot be developed or used until they can demonstrate the capability of adequate distinction, proportionality, and shown that they do not produce unnecessary suffering, and must only be used given military necessity. Outside those bounds any individuals responsible should be held accountable for violations of International Humanitarian Law, whether they are scientists, industrialists, policymakers, commanders, or soldiers. As these systems do not possess moral agency, the question of responsibility becomes equated to other classes of weapon systems, and a human must always ultimately bear responsible for their use[23]. Until it can be shown that the existing

[22] Royal United Services Institute for Defence and Security Studies, "The Ethics & Legal Implications of Unmanned Vehicles for Defence and Security Purposes", Workshop webpage, held Feb. 27, 2008, http://www.rusi.org/events/ref:E47385996DA7D3, (accessed 5/12/2013).

[23] Cf. Arkin, R.C., "The Robot Didn't Do it.", Position Paper for the Workshop on Anticipatory Ethics, Responsibility, and Atificial Agents, Charlottesville, VA., January 2013.

IHL is inadequate to cover this RMA, only then should such action be taken to restructure or expand the law. This may be the case, but unfounded pathos-driven arguments based on horror and Hollywood in the face of potential reductions of civilian casualties seems at best counterproductive. These systems counterintuitively could make warfare safer in the long run to the innocents in the battlespace, if coupled with the use of bounded morality, narrow situational use, and careful graded introduction.

Let it be restated that I am not opposed to the removal of lethal autonomous systems from the battlefield, if international society so deems it fit, but I think that this technology can actually foster humanitarian treatment of noncombatants if done correctly. I have argued to those that call for a ban, they would be better served by a call for a moratorium, but that is even hard to envision occurring, unless these systems can be shown to be in clear violation of the LOW. It's not clear how one can bring the necessary people to the table for discussion starting from a position for a ban derived from pure fear and pathos.

For those familiar with the Martens clause[24] in IHL, a case could be made that these robotic systems potentially "violate the dictates of the public conscience". But until IHL lawyers agree on what that means, this seems a difficult course. I do believe, however, that we can aid the plight of non-combatants through the judicious deployment of these robotic systems, if done carefully and thoughtfully, particularly in those combat situations where warfighters have a greater tendency or opportunity to stray outside International Humanitarian Law. But what must be stated is that a careful examination of the use of these systems must be undertaken now to guide their development and deployment, which many of us believe is inevitable given the ever increasing tempo of the battlefield as a result of ongoing technological advances. It is unacceptable to be "one war behind" in the formulation of law and policy regarding this revolution in military affairs that is already well underway. The status quo with respect to human battlefield atrocities is unacceptable and emerging technology in its manifold forms must be used to ameliorate the plight of the noncombatant.

[24] The clause reads "Until a more complete code of the laws of war is issued, the High Contracting Parties think it right to declare that in cases not included in the Regulations adopted by them, populations and belligerents remain under the protection and empire of the principles of international law, as they result from the usages established between civilized nations, from the laws of humanity and **the requirements of the public conscience**." (Available at the ICRC website, http://www.icrc.org/eng/resources/documents/misc/57jnhy.htm last visited on 30 April 2013)) [Boldface mine].