

What are autonomous weapon systems and what ethical issues do they rise?

Marek Foss, 30/03/2008

1. Introduction

Autonomous weapon (AW) systems are a new and rapidly developing branch of warfare industry. However, autonomous weapons are not devices that belong strictly to the XIX century, in fact some authors date the birth of first autonomous weapons back to 1920's[1]. But the delicate matter of the definition of such weapons brings questions whether it is not the case that today's machines are really autonomous while yesterday's were just enhanced weaponry, preset to react to a certain, small number of input conditions. Also, should the definition state what humans would like the AW systems to **be**, or what they really **are**[2]? Moreover, does the enhanced capabilities of such systems change the way humans should treat the actions of such machines? Do they pose a threat to humans like a kitchen knife, that has to be misused by a person to cause harm — or like an enemy soldier, that nevertheless has to take responsibility for his actions?

This paper tries to establish a definition of AW systems and compares that definition to those presented by others. After establishing what these weapons are, the paper analyzes the ethical issues connected to the application of these devices. An argument is derived basing on case studies and author's thoughts, and is confronted with several points of view and common beliefs.

2. Autonomous Weapons Systems

By dictionary, being 'autonomous' is the ability to act independently or have a freedom to do so. Thus, I define AW systems as those that operate without human intervention and as such are able to complete their tasks by processing, responding to and acting on the environment they operate in. The key feature of an autonomous weapon is the ability to 'pull the trigger' — attack a selected target without human initiation nor confirmation, both in case of target choice or attack command.

As such, this definition differs from others because it eliminates human factor — both Sparrow[3], Arkin[4] and Guetlein[5] approve some 'man-in-the-loop', and therefore "include cruise missiles, torpedoes, submersibles, robots for urban reconnaissance"[3]. Some researchers[1] went on to include the first guided missiles and German V-1 rocket from 1945. And although notable in terms of AW history, they do not comply with the ethic issues that arise with current, sophisticated devices. My definition eliminates the distinction between semi- and fully autonomous systems, putting the former in the box of conventional weapons. Reason behind this is purely in the context of ethical issues.

3. Ethical Issues

The key ethical issue concerning AW systems is responsibility, which spans three subproblems — control, i.e. who is controlling the system, man or machine; consciousness, i.e. does the system interfere with man orders or is self-aware; and crimes, i.e. can a machine commit it and who takes account for it. Responsibility is the reason why I want to distinguish semi- and fully autonomous weapon systems.

Imagine a boomerang thrown into air — its trajectory is not exactly predictable (except that it should fly back to the thrower), thus its flight is sort of autonomous. Now this

boomerang hits and injures another person — who is responsible for this accident? Of course, the person 'pulling the boomerang trigger', the thrower — although in modern western societies the manufacturer could be brought to court for not explaining in the manual that the boomerang can act in this way.

Analogically, when the guided cruise missile is fired from i.e. gunship by army personnel, that personnel is responsible for missile action, even if they set it to attack any hostile targets it can locate. Similarly with Unmanned Aerial Vehicle (UAV) like US Air Force (USAF) RQ-1 Predator[6], which located and killed possible terrorists in Yemen in 2002[7] — and although it was able of unmanned take-off, flight, routing and landing, its Hellfire missiles launch was initiated by the human Command&Control (C2) unit that had live feed of the UAV's mission theatre, not by the Predator itself.

However, the future is the complete elimination of human control. Take for example Low Cost Autonomous Attack System (LOCAAS)[8], which can autonomously hover over a desired mission theatre "at an altitude of 750 feet over the battlefield, flying at a speed of 200 knots for about 30 minutes, covering a footprint of 25 square nautical miles, and take out high-priority targets such as mobile air defenses, mobile surface/surface missile launchers and long range rocket systems"[8]. US Navy is developing a similar weapon for underwater operation[9], Israel deploys autonomous snipers in Giza[10], and South Korea constructs autonomous guards for its Demilitarized Zone on North Korean border[11].

Who is in control of these systems, who is responsible for them? Is it the company who programmed the machine? Or is it the command-in-chief who accepted the military implementation of the machine? Or is it the commander closest to machine who sets it up with parameters before the mission? If an accident is caused by the control program, which fails regardless of set parameters, then the responsibility lies on the producer. If the command-in-chief was aware of these flaws, but still implemented the systems, it is responsible as well. Finally, if it was the commander who incorrectly set up the system, he should be held responsible on the same basis as a mortar operator is held responsible for miscalculating the grenade trajectory and hitting a civilian village instead of a terrorist hideout.

Notice the top-down approach — it ensures that a responsible person can be found even if the system is fully autonomous, non-configurable, 'fire-and-forget' type - until machines will autonomously build and configure machines on a large scale without human intervention, that is. Responsibility is essential, otherwise "AW [system] will violate this important condition of *jus in bello*", quoting Sparrow([3], p. 68), who doubts that the responsibility will not be blurred by application of AW. But the army has a different point of view, believing that "AW can preserve the legitimacy of the cause because the use of force is constrained by a rigid set of heuristics preprogrammed to comply with the [Rules of Engagement] and the The [Law of Armed Conflict]"([5], p. 11). The claim suggests the machine basis on probability, and even when it is very low, it "can be preprogrammed to ask 'mother-may-I' prior to engaging the objective"([5], p. 11). Sparrow[3] asks whether it is the case that the 'man-in-the-middle' of the decision chain is enough to ensure correct behavior of machine and sufficient responsibility execution, because the final goal is to exclude man and rely on the AI as the decision maker. In fact, I agree with Sparrow that human interaction may disrupt the operation outcome of the system — for example in 1988, Iran Air Flight 655 (IR655) civilian airliner was shoot down by US missile killing 290 people[12]. The command to attack was given basing on the incorrectly processed output of the semi-autonomous AEGIS US Navy combat system[13] — the crew of missile cruiser USS Vincennes was said to be in a "simultaneous psychological condition"[12], disobeying the correct readings(see [13]) from AEGIS saying that IR655 is a civilian plane and deliberately assigning an enemy F-14 markup onto it, therefore proceeding to the attack of the plane.

Although this does not mean that the elimination of human control is essential nor desired. In October 2007, South African National Defence Force performed an upgrade on an automated anti-aircraft gun that is able of fully autonomous target location, elimination and

munition reloading. During the upgrade, the gun went berserk, killing 9 soldiers and leaving 11 wounded[14]. Who may be held responsible for this act? In the press coverage, nobody blames the machine, rather the manufacturer and the military upgrade personnel. It's still a common belief that regardless of autonomy, machines do not possess any soul or self-awareness, thus humans don't put the blame directly on them (and even if they do, they look at the label of the manufacturer). But in the distant future humans will develop machine autonomy even further, creating conscious or semi-conscious machines with reasoning abilities similar to animals or even humans — what then?

A situation may happen that the machine will override commander orders and for example eliminate surrendering enemy soldiers on the basis of calculating the mission costs (i.e. Sparrow[3]). The question of who is 'inside' the conscious machine, and what value its existence possesses is a difficult one — because in this case how can we punish the machine, should we lock it in a hangar, or disassemble into pieces? Would it teach that machine (or other machines) a lesson? Would it make any difference to them?

I agree with Sparrow([3], p. 72) that punishment cannot occur without suffering, and that 'suffering' as such is far ahead of the technology that AW systems are capable of today, if ever will be. Therefore the responsibility for machine crimes should be held by the owners. Consider these two sets of examples — firstly, a dog bites a person to death - it is probably silent, and the owner is prosecuted; or a child beats his schoolmate - it is punished and the parent is responsible; or a mechanical machine causes a fatal factory accident - it is examined and the manufacturer is sued. Imagine that AW would be a dog of a field commander, a child of a system programmer and a machine of the command-in-chief and the manufacturer. I suggest here an equally distributed responsibility, not divided responsibility. I suggest that if by law the account for AW systems was taken by each of these entities, it is very likely each of them would maintain caution in terms of AW operation.

If not, if the responsibility was blurred, and held on an individual that cannot fully impose his will onto the AW (like the commander cannot reprogram it, or the programmer cannot control its operation) we may end up with machines that are left without orders, without responsible personnel, but still operational and ready to kill. Like land mines — these devices work autonomously, but for one purpose, long after the conflicts end, and bring havoc to the young population living in the area where land mines were deployed[16]. It is very likely that abandoned AW machines may turn into land mines of XIX century, but even more deadly.

But there is another side of the ethical issue, the machine side — in other words, are we ethically suitable to pose such a claim against machines? "Machines ... don't have hate, vindictiveness, cruelty, or psychosis. Machines don't rape innocent women. Machines don't abuse prisoners. Machines don't massacre civilians." [17] Maybe an automated, algorithm-precise war, with machines doing cold calculations would be more humane than human wars? What of an irony that would be, but how possible it is to see machines that do what they were programmed to, without acts of vandalism and war crimes. And the concern that Sharkey[17] displays over the machines not having the Geneva Pact, is really a false pride of a human being, that does not want to live in a world where machines act more righteous, more correct? Because even though humans have various pacts, laws etc. we still have Holocaust, Vietnam and recently — Abu Ghraib[18] in our history.

3. Conclusion

Defining Autonomous Weapon Systems as the fully autonomous machines brought the ethical issues to the key point of responsibility for the actions of such machines. It is the responsibility of the owners to maintain the correct execution of orders and lawful application of AW. The owners in this case are the cumulative group of people involved in the creation and control of the AW. Humans already are developing systems that can malfunction, and it is their, not the machines, responsibility to correct these errors. Also, it is humans

responsibility to control what machines are doing. However, despite the nightmare visions of several authors, it is equally likely that the war using AW systems will be less humane than the conventional human conflicts.

References:

- [1] Stanford University CSE, "Autonomous Weapons", S. Chen, T. Hsieh, J. Kung, V. Beffa (<http://cse.stanford.edu/classes/cs201/Projects/autonomous-weapons/>)
- [2] The Risks Digest Volume 3: Issue 64, "Towards an effective definition of 'autonomous' weapons" (<http://catless.ncl.ac.uk/risks/3.64.html#subj4>)
- [3] Journal of Applied Philosophy, Vol. 24, No. 1, 2007, "Killer Robots", Robert Sparrow
- [4] Technical Report GIT-GVU-07-11, "Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture", Ronald C. Arkin
- [5] Naval War College, "Lethal Autonomous Weapons - Ethical And Doctrinal Implications", Major Mike Guetlein
- [6] Wikipedia entry on "MQ-1 Predator". RQ-1 was the initial name of the unit, but was changed when a modified version, RQ-9 Reaper, appeared. (http://en.wikipedia.org/wiki/MQ-1_Predator#Yemen)
- [7] Jane's Defense Business News, "Yemen Drone Strike: just the start?", C. Hoyle, A. Koch (http://www.janes.com/aerospace/military/news/jdw/jdw021108_1_n.shtml)
- [8] Defense Update, "Low Cost Autonomous Attack System (LOCAAS)" (<http://www.defense-update.com/products/l/locaas.htm>)
- [9] US Navy, "Sea Predator A Vision for Tommorrow's Autonomous Undersea Weapons" (http://www.navy.mil/navydata/cno/n87/usw/issue_29/predator.html)
- [10] The Register, "Israel deploys robo-snipers on Gaza border", L. Page (http://www.theregister.co.uk/2007/06/05/israel_robo_sniper_gaza/)
- [11] The Register, "South Korea to field gun-cam robots on DMZ", L. Page (http://www.theregister.co.uk/2007/03/14/south_korean_gun_bots/)
- [12] Wikipedia entry on "Iran Air Flight 655". (http://en.wikipedia.org/wiki/Iran_Air_Flight_655)
- [13] Wikipedia entry on "AEGIS Combat System". (http://en.wikipedia.org/wiki/Aegis_combat_system#Iran_Air_Flight_655)
- [14] Danger Room from Wired.com, "Robot Cannon Kills 9, Wounds 14", N. Shachtman (<http://blog.wired.com/defense/2007/10/robot-cannon-ki.html>)
- [15] Wikipedia entry on the "Three Laws of Robotics". (http://en.wikipedia.org/wiki/Three_Laws_of_Robotics)
- [16] UNICEF.org , "Land-mines: A deadly inheritance"

(<http://www.unicef.org/graca/mines.htm>)

[17] The Guardian, "Robot wars are a reality", Noel Sharkey, from comments by readers
(<http://www.guardian.co.uk/commentisfree/2007/aug/18/comment.military>)

[18] The New Yorker, "Torture at Abu Ghraib", S. M. Hersh
(http://www.newyorker.com/archive/2004/05/10/040510fa_fact)