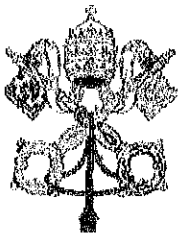

The Use of Lethal Autonomous Weapon Systems

Ethical Questions

PERMANENT MISSION OF THE HOLY SEE TO THE UNITED NATIONS AND OTHER INTERNATIONAL ORGANIZATIONS IN GENEVA

***

GENEVA, 16 APRIL 2015

1. Introduction

In today's conflicts we notice an increased presence of robotical devices. These may be aerial or sub-marine drones, reconnaissance or offensive robots, terrestrial scouting or de-mining robots, even transport robots. A robot is a system endowed with three essential components: sensors which collect data on the environment; processors to compute the obtained data, for example by means of artificial intelligence; and finally the means to carry out actions on the environment and to respond to the provoked reactions to these actions.

In addition to conventional military robots, there now exist computerised robots. These systems, which are able to acquire information, process it and to intervene into the networks they are connected to, can wage genuine "cyber-wars" (such as pirating important information networks, financial warfare, invasion of social networks by use of propaganda).

Today, some robots are also found in civil society due to their relatively accessible cost. This should make us cautious since, in addition to the possibility of invasions of privacy, these robots could relatively easily be armed and put to the service of terrorist organizations or criminal groups. This proliferation of small armed robots could lead to the importing and extension of situations of war or terrorism everywhere in the world, making such conflicts uncontrollable.

Whether it involves classical or computer robots, research today aims to give greater autonomy from their human operator. This means giving them the capacity to innovate in conflicts or cyber-conflicts. In other words, the research is geared towards granting them the ability to depart of their own accord from the original areas or tasks prescribed to them by the responsible authority.

Furthermore, the analysis of warfare robotics cannot be separated from an analysis of the nature of contemporary conflicts. A large part of the global population, today, lives in urban areas and this means that wars will take place essentially in these urban contexts. It is important to be able to measure the impact and pertinence of recourse to robotic technology and the use of lethal autonomous weapon systems (LAWS) in these contexts.

This new situation means we need to carry out a new and more comprehensive analysis of warfare robotics. This could first be done by raising the issue of the legality of the use of military robots in relation to international humanitarian law (IHL). This is necessary but insufficient, for the replacement of the human decider by sophisticated and autonomous machinery raises a series of anthropological questions (What is our conception of the human being?) and ethical questions (what values are at stake in the delegation of important powers to robots, power over life and death for example?). After having briefly underlined certain points raised by the consideration of international humanitarian law (IHL), we will suggest certain landmarks enabling the guidance of an ethical questioning relative to armed robots and especially LAWS.

2. International Humanitarian Law and military robots

Is international humanitarian law (IHL) compatible with the introduction of robotic weaponry endowed with a large autonomy?

If we begin with the premise of *jus in bello*, the application of the principle of discrimination between combatants and non-combatants, and of the principles of proportionality, precaution and expected military advantage of an attack cannot reasonably be entrusted to robots. In fact, these tasks require the interpretation and understanding of the context and of situations which are not really programmable. In combat situations which, today, occur frequently in urban areas with a permanent presence of the civilian population, the appreciation of the legitimacy or legality of a military action is more and more complex and too subtle to be entrusted to a machine (which, for example, would be ineffective when facing moral dilemmas or questions raised by the application of the principle of double effect).

It has been noted on many occasions that, following the premise of *jus ad bellum*, the risk exists that the use of autonomous weapon systems increases the probability of triggering conflicts, as they do not entail the loss of human lives on the part of the user and their production is relatively inexpensive. In addition, we cannot underestimate the risks connected to the miniaturization of surveillance robots, armed or not, in periods of tension preceding potential conflicts. Indeed, this could lead to the multiplication of incursions into an opponent's territory provoking the latter to counter-attack.

Following the premise of *jus post bellum*, we cannot underestimate the difficulties for achieving a durable peace that the use of autonomous weapons could bring after a conflict. In any case, the victor states who would use these should think of the effects produced by maintaining some sort of robotized sword of Damocles hanging in permanence above the heads of the conquered.

In fact, miniaturized or highly autonomous robots already open the door, from the point of view of civil law, to possible important violations of privacy. But from the military point of view, they represent a permanent and generalized risk of espionage, which needs to be controlled by law. The necessity for this control is motivated by the fact that it is not possible to build a durable peace or an alliance between nations if such means of espionage, ruining mutual trust, are maintained on a permanent basis. We have recently been able to note such behaviour with episodes of cyber-espionage even between allied countries. Robotized espionage could accentuate this phenomenon.

IHL is necessary, and most agree that it is an important framework to regulate the use of armed (autonomous) robots. Nevertheless, this framework is not sufficient. Furthermore, we know that it has sometimes been possible to justify immoral conflicts by the use of "the theory of a just war". It is therefore necessary to go beyond IHL to apprehend, in all its dimensions, the questions raised by robotic weapon systems.

3. Ethics: some landmarks

Our approach here is to draw up a list of points deserving particular attention and requiring discussion, if we want to approach the question of military robotization in an ethical way.

3.1 The anthropological coherence and the relational effects of robotization

When acquiring civil or military technological devices, supposed to assist in a set of difficult or dangerous tasks, human beings always have to watch out that they do not find themselves in a situation where, having entrusted a series of powers to their machines, they become slaves of their own inventions. The relevant principle here is coherence, since, if the devices supposed to free human beings become the cause of their oppression, we are faced with a profound contradiction. The momentary delegation or the permanent transfer of certain powers to machines endowed with the capacity for innovation and action have always to be measured in the light of this anthropological coherence.

The purpose of technological inventions robotized or not, is to help and free human beings. Therefore, it is suitable to be attentive to the consequences of projects that, under the pretext of wanting to augment human capacity, to protect him from dangers or to free him from constraints, would end up in "diminishing" him and enslave him to his own tools. Otherwise, when confronted by military robots endowed with the capacity to innovate, we could end in a situation where the technological device itself would develop strategies or tactics that the human conscience or will would not truly have ratified.

Another important ethical aspect in the use of armed robots is linked to relational effects. Maybe, today, we are not measuring sufficiently the risks that represent the progressive replacement of human actors by machines. As the philosopher Emmanuel Levinas has shown, the face-to-face encounter with the 'other' is one of the fundamental experiences which arouses the moral conscience and responsibility. All wars lead to a regression in humanity. But a war using solely and unilaterally technological systems, where man is absent, can increase even further this dehumanization. The presence, in the field, of human mediation can, in certain cases, open doors and offer occasions where dialogue can be resumed, where conflicts can be solved by means which are totally absent when using armed machines "without faces". This human presence permits us to be open to concrete occasions where empathy can operate, with the capacity to be able to be touched by others' suffering. On the contrary, armed machines are like psychopaths, insensitive to signs of suffering or sadness emanating from others.

Possibly, one will object that machines could be programmed to detect emotions and adopt conduct mimicking empathy. But empathy is something more than an immediate response to signs detected on the face or body of another. It is the capacity to be profoundly touched oneself by the misery of the other and to share in his burden. It has sometimes been objected that the history of contemporary conflicts has shown that under the influence of emotions or sentiments of vengeance for example, combatants can lose their cool or their reason and engage in terrible abuses, whereas programmed machines would not. However, it is demonstrable, on the contrary, that an officer or a soldier, when seriously formed in the respect of not only rules of engagement and of IHL, but also sensitive to human values, can, *a fortiori*, be an essential actor in reducing the dehumanization and barbarity involved in all wars.

We forget, or frequently minimize, that all weapons, from the perspective of their psychological impact, are not to be put in the same category. The Second World War has given certain flagrant examples, with the Stuka, whose siren terrorized civilian population, or the V1, whose characteristic sound, anxiously heard by the population, meant the imminent fall of the bomb. Certain weapons, by their intrinsic nature, are of a nature which provokes or

augments the stress of civilian populations. This needs to be taken into account in the ethical evaluation of these weapon systems. Being flown over by planes susceptible of bombarding us is already a traumatizing experience, but the fact of being flown over in permanence by robotized machines susceptible of choosing and neutralizing targets at unexpected moments can be even more stressful. Besides, we see here the difference between the juridical and ethical aspects. A robotized weapon could meet the principles of discrimination, proportionality and military usefulness, while producing important psychological traumas (difficult to quantify, but real) on the civilian population, and therefore raising important ethical questions.

3.2 The importance of prudential judgment

The removal of the human presence in the decision-making process for the selection of targets or the decision to shoot, and leaving this to an armed robot endowed with the capacity for innovation, raises fundamental questions. Indeed, to decide if this or that action is legal or legitimate on an ethical point of view, we need to refer ourselves to norms, to principles whose application to particular contexts demands evaluation and interpretation, which are not easy to translate into algorithms. The interpretative dimension operates on a double level: on one hand, the level of norms themselves (what is the sense, the spirit of the law?), on the other, the level of persons and contexts (what are the pertinent characteristic traits of the people and situations that make sense in the framework of the application of the law?).

A tradition emanating from Aristotle and St. Thomas Aquinas qualifies "prudence" as, that virtue that allows us *to apply universal knowledge to particular contingent realities*. It is created by experience, through the comparison of encountered situations and *a capacity for interpretation which can be identified to a purely deductive or demonstrative procedure*.

Classical discussions show indeed that the work of a judge cannot be reduced to the simple application of abstract norms regulated by pure formal (deontic) logic. For example, the qualification of a fact, the management of gaps or contradictions in the law (domestic or international), conflicts between norms, the clarification of terms that are too vague, or all situations where the personal interpretation of the magistrate cannot adequately be reproduced by a system of formal algorithms, without this interpretation sinking into the arbitrary or the irrational.

Furthermore, in conflicts involving interpretations of the 'good' or when faced with decisional dilemmas, the best programmed machines are more helpless than human beings! It is therefore important to be conscious that *prudential judgment cannot be put into algorithms* (for it is not merely a mechanical application of norms) and logic formalization. The experience that it presumes is in addition difficult to acquire by machines even if these are endowed with capabilities for self-learning and self-programming (by using genetic algorithms for example). Indeed, in this case, we do not see how to determine the criteria for selection of the situations the machine should remember as a pertinent basis for future behavior.

It is necessary to add here that a juridical or ethical decision leads sometimes to having to transgress the letter of universal rules to safeguard its spirit. Here too we see the limits of the use of programmed systems. Indeed, how could we programme systems that should transgress their own rules? When the question of the robotization of armed conflicts is raised, it is important to think about the question of knowing if this "prudence", so essential to the

judgment, has not been removed or replaced by formal imitations which are not their equivalent.

3.3 Curb the desire for almighty power

Another important point which necessitates our attention is the fascination exercised by armed robots and the fantasies of power they mobilize. Their use can implicitly originate from the desire for power rather than from the wish to find proportional means for a justified defense. The development of "augmented soldiers", these are combatants to whom we have conferred extraordinary capacities (perceptive, cognitive and of action) by robotized means (such as exoskeletons for example) or totally autonomous robots endowed with great capacities can originate not from military usefulness, but from dreams of power.

What is human is recognized by our capacity to put a brake on our powers. Moral behavior proper to humans is situated in the position of the 'juste milieu' (the 'happy medium') which leads us to exercise force when justice and law require it, but also to renounce its use when higher values demand it. *Accepting, at moments, to not exercise one's power, to not deploy one's "overwhelming power", is truly human behavior.* Indeed, it is that which paves the way to reconciliation, to pardon, to new solidarities, beyond all obstacles left by conflicts.[1] *The nature and real greatness of man shows itself in the behaviours where the strong can make place for the weak.* On the contrary, *barbarity and human regression, originate in the deployment of a pure power of destruction devoid of any restrictions.*

The ethical evaluation of research projects on military robotics has to take into account the differences between research motivated by a legitimate demand for security and defense, and research motivated by the sole will to realize a pure demonstration of force. This kind of demonstration would furthermore risk harming peace among nations, and generate complexes, stirring fear, rancor, jealousy and eventually, provoking in the adversary the desire to resort to similar methods, provoking a potentially uncontrollable escalation. The implementation, in a technological military system, of a pure fantasy of power, will by its very nature destabilize the balance of peace.

3.4 Curb the dangers of de-responsibilization

The robotized weapon, especially one which would be autonomous, raises important questions with regard to the exercise of responsibility. In the case of collateral damage, caused by weapon systems controlled by a human operator, the latter is spontaneously recognized as the one bearing responsibility. But if this collateral damage is the deed of an autonomous machine, even if the final responsibility lies with the authority who put it into action, it will always be possible to deny it by invoking a series of malfunctions for which the authority would not be responsible (computer bugs, technical failures, design faults, scrambling of communications, etc.). It is a problem which also arises in the civil world when machines are used instead of humans. Experience shows that lawyers can use technological "screens" to construct strategies that mitigate or overshadow the identification of who is responsible.

[1] "A society founded on solely on relations of power would have nothing human: it would necessarily minimize the freedom of men instead of helping and encouraging it to develop itself and to perfect itself", St-John XIII, Encyclical Letter *Pacem in Terris*, n. 17, *AAS*, 55 (1963) 265.

This, in turn, could promote the use of armed robots exactly because of *the impunity they permit*. It would be easy to engage in combat or carry out targeted killings because the sponsors feel sheltered from direct pursuit by the law. Nevertheless, it is suitable to keep in mind the classical maxim: "*qui facit per alium, est perinde ac si facit per se ipsum*" (he who makes another do something, it is as if he had done it himself).

Just as it is intrinsically human in being able to withhold the deployment of his full power to let the other be, especially when he is vulnerable, similarly, *it is intrinsically human to take responsibility for one's actions*. A doctor, like a magistrate or a military officer is not a mere technician (of the human body, the law or questions of defense). Their actions are not limited to the sole mechanical application of procedures.

The actions of jurists, as much as those of doctors or officers, are bound by rational decisions and by the taking of responsibilities which engage the whole of their person, this engagement confers to these actions an anthropological depth and density. We concur with Paul Ricoeur who characterizes the judge as some-one "humanizing the law". In a similar fashion, we could say that, without the doctor, the judge or the officer, there is no one to humanize their particular domains of action, to bestow on these meaning and value. The delegation of important powers (economic, political, military, ...) to machines must alert the ethical conscience, insofar as it deprives man of the power to behave authentically and where it leads to acts that we could qualify as, in the true sense of the word, "senseless".

It might be important to underline here an important point concerning *intentions*. The evaluation of the responsibility of a person is concerned with his intentions. Concerning the selection of a target or the choice to shoot, we must absolutely ask ourselves the question of what were the precise intentions of the person who made the decision. This taking into account of intent is central to apply for example "the rule of double effect", characteristic of situations where an action is simultaneously the cause of good and bad consequences. Indeed, this rule requires that one has fundamentally the intention of producing good consequences and not bad ones (even if the latter seem unavoidable). *However, this notion of intent*, equally important for both law and ethics, which influences greatly responsibility, *cannot easily be linked to a concept or a technically approached reality or apprehended by robotics or today's information technology.*

3.5 Managing the effect of astonishment linked to robots' performance

Robots which would be expected to make decisions in the absence of any human supervision have programmes that bear the mark of a series of assumptions. Nevertheless, *fascination for the performances of these machines can make us forget that their properties and behaviour depend largely on the ideas of their makers.* Let us take two examples.

To create a financial robot, a series of choices are made on how risks will be calculated. But this involves hypotheses on the evolution of the stock market, hypotheses which depend in their turn on certain presumptions. Different mathematical models will give different results and could lead to underestimating a potential yet important risk. The astonishment provoked by the achievements obtained by financial robots risks, in the long run, making us forget that their programmes cannot anticipate everything and may lead to catastrophes. These are linked, for example, to extremely rare financial events, whose occurrences are not included by the programmes in the calculation of risks.

In the military domain today, certain people envisage including, within the LAWS programmes, "ethical" programmes, thus making them into "moral machines". These are software programmes which are supposed to verify the compliance to, not only rules of engagement and principles of IHL, but also ethical norms. However, what ethical presuppositions are chosen for which software? Often, the defenders of "moral machines" use a utilitarian morality. But what justifies this choice? It is true that such a morality is founded on an optimizing calculus which can be translated into algorithms. Nevertheless, this calculation also requires measuring *quantitatively* the wellbeing or lack of it and, the value of this good or that person. How can we justify this? These are often ideologically driven choices. The fascination exercised by autonomous machines who would combat in our place and that would be endowed with a judgmental capacity can make us forget that the choices that underlay the creation of the software which commands these machines are contingent and that juridical or ethical judgments cannot be reduced to a mechanical application of formal rules or to calculations.

3.6 To maintain the control of the robotic object

One of the crucial points on which the ethicist will have to insist is the question of knowing if we can let machines whose behaviour is not entirely predictable proceed. *One of the conditions to be able to use armed robots is the fact that we can be assured that it will never produce behaviour that is a priori prohibited by its user.* However, this is never guaranteed, in an absolute fashion, whatever the sophistication of the software, due to logical limitations that have been put forward by theoretical algorithmics. In addition, these complex systems are not sheltered from computer bugs and viruses and computer piracy. More than before, modern military technology is complex and susceptible of dangerous misappropriation by criminal organizations. This is reinforced by the fact that militaries depend more than before on private companies, because of the required levels of technological expertise but also because of the savings made by reducing the personnel of Western armed forces. These private companies do not benefit from the same degree of security and protection as those that operate in military bases.

The ethical evaluation of armed robots has to take into account this aspect, for *it is fundamentally immoral to utilize a weapon the behaviour of which we cannot completely control.* For this reason, completely autonomous robots have to be per se definitively prohibited since they will be capable of escaping from the behavioural constraints imposed by both the inventor and the user. If a weapon system has lethal capacity, it is necessary that it be maintained under absolute human supervision.

3.7 Concern for the economic aspects of robotization

The economic and financial stakes of the civilian and military use of robots are colossal. It is certain that the richer countries will profit more from the advantage that robotized weapons confer on them. Divisions will therefore appear between those who possess and deploy such combat technology and those who do not. Situations of injustice risk appearing as well as easy violations of the airspace of less technologically favoured countries. We will therefore reinforce already existing divisions between states.

In addition to these problems, an arms race for more and more efficient robotized weapons systems risks leading to the degradation of resources that could be useful in other sectors such as health or education.

Ethical vigilance must also be applied to the civilian sector, as we can see that today the development of civilian robotics frequently precedes that of the military, and sometimes this leads to the civilian economy forcing the military sector to integrate such technologies even if they were not necessarily needed. The race towards profit could induce competitive industrial groups to escalate the production of increasingly sophisticated military robots making them more accessible to wider groups of actors. The relatively low cost of certain armed robots could then entice, in a near future, their rapid and uncontrollable proliferation.

4. Conclusions

This report argues that the reflections exposed herein give a few arguments showing that in the domain of the new robotized weapon systems, *it is suitable to prohibit the systems which possess lethal capacity and are at the same time capable of escaping effective control by human beings*. In other words, lethal capacity cannot be bestowed upon a weapon system that could eventually, by its own initiative, leave the parameters of action that were laid down for it by the responsible authority.

If we *define an autonomous weapon system as an armed system capable of innovating by adopting, by itself, behaviuor that has not been foreseen and prescribed by the responsible user, it is clear that such a system can eventually escape, by definition, effective human control*. It falls therefore under the scope of this prohibition.

It is necessary to note that the weapon systems targeted by this prohibition are not limited to hypothetical LAWS (Lethal Autonomous Weapon Systems). Indeed, there could exist automatic or not completely autonomous weapon systems endowed with lethal capacity and eventually able to escape effective and operative control by the human operators. It is not the autonomy that is covered by the prohibition we have recommended, but rather the conjunction of potential lethality with a possible loss of effective control over these weapons systems.

It is important, to conclude, to recall, in a few words, the reasons that motivate the advocated interdiction.

A first basis for the prohibition of lethal weapons potentially escaping effective human control is the *risk of a deresponsabilization* or, at least, the obscuring of where true responsibility lies. If armed robots endowed with a large autonomy provoke collateral damage, there is a risk that there will appear strategies seeking to exonerate the authorities which have employed these machines, by blaming instead technical malfunctions at different levels. The capacity to kill other persons implies an immense responsibility that can only profitably rely on a human decider and not on a technological device as sophisticated as it may be and of whom we can never truly say it is responsible!

A second basis to this prohibition is the risk of *the aggravation of the dehumanization process*. We know that all wars (robotized or not) are dehumanizing. However, a war that would be fought by the means of robotized machines that would not be effectively controlled by humans would be even worse. The eviction of man by the machine prevents the appearance of behaviour characterized by compassion, reconciliation, respect, which are essential to the achievement of a true peace. A true brotherhood of men and peoples cannot be acquired in conflicts where certain persons will be confronted, in an asymmetrical way, to attacks conducted by machines partially or totally subtracted from effective control by men.

Robotized wars, without human faces or people taking responsibility, risk obstructing the paths leading to a sustainable peace, which should be based on an ethic of true brotherhood among the nations.

A third basis of the prohibition is linked to the fact that *the frequent delegation of important powers to machines risks depriving the political authorities of their raison d'être* and therefore of *their capacity to act in a responsible manner.* Being out-runned by their machines and dazzled by their fascinating performance, these decision-makers risk not being able to decide anything, finding themselves in a paradox where the decider decides not to decide anymore! The autonomous combat machines risk, through their speed and their capacity, to dictate important military action policies.

The risks of *deresponsabilization, dehumanization* and *depolitization* induced by the use of lethal weapons removed from effective control by men are important enough so that we can envisage asking for their prohibition and the research that develops such systems.