Robot Warriors Autonomously Employing Lethal Weapons: Can They Be Morally Justified?

Dong Meizhen
Philosophy Department
Nanjing Normal University
Nanjing, China
e-mail: dongmz@hotmail.com

Gao Zhaoming
Philosophy Department
Nanjing Normal University
Nanjing, China
e-mail: gaozhaoming@hotmail.com

***Abstract*—The thesis ponders on the existence justification for Robot Warriors Autonomously Employing Lethal Weapons (RWAELW) from the angle of value and non-technique. RWAELW are different from Unmanned Aerial Vehicle (UAV), robot, etc, in military actions. The former that have certain learning capacity are "quasi-subjects" with autonomy, self-choice, self-determination and free will. However, they have no sense of good and evil, no moral affection and no capacity of value judgment. A killer that cannot be controlled by human beings is extremely dangerous. So it should be forbidden to develop Robot Warriors Autonomously Employing Lethal Weapons (RWAELW) in human society.**

Keywords—robot warriors autonomously employing lethal weapons (RWAELW), quasi-man warrior, ethical defense

I. RAISE THE PROBLEM

It is a general trend that robots are increasingly applied in various fields. So is it in the military area. To some extent robot warriors have already been in existence, such as robot bomber, unmanned reconnaissance aircraft, mine clearance robots, etc. Robots of war and their ethical problems have already been noticed.[1] Based upon current situations in the development of society, technology and economy, it would not be far away from now on for RWAELW to join military action in a war if we as human beings do not take some special measures to prevent it from taking place.

With the increasing maturity of modern science and technology, under the political, cultural and economic pressure, and due to the asymmetry of modern warfare, the possibility of RWAELW has become an urgent, vital and real issue, which is worthy of high attention of human beings. Firstly, in theory the development of relevant techniques, especially information technology and AIT (artificial intelligence technology), has made it gradually technically possible for us to manufacture RWAELW. The progress of such a technology might go faster than we have expected since the fact that AlphaGo has defeated Go champion indicates that AI has had the ability of self-directed learning, self-developing and pattern judgment, etc.[2] Secondly, the pressure from politics, culture and economy in a society increases certain demands to "RWAELW". On one hand, in spite of there is little possibility of the outbreak of a massive world war in present situations in the world, it seems inevitable for local and regional warfare to take place. The reason is not only from the existence of conflict between different countries and nations, but also from the dangerous terrorist activities in the form of ultranationalism and religionism. On the other hand, in consideration of economic and social price a war means death and huge amount of military expenditure. Minimizing the casualties of soldiers as well as civilians is the main goal in a war for a responsible democratic country when it has to make a necessary military decision. This is not only because of the pressure of domestic politics and humanism, but also because of vast post-war financial burden of paying pension and life insurance for the disabled or for the families of the deceased due to the war. Thirdly, modern regional war, especially anti-terrorist war, manifests more and more unsymmetry. Unsymmetrical war changes the people's view of war and its means, and makes them tend to choose unmanned tools.

RWAELW are quite different from unmanned reconnaissance aircraft, mine clearance robots and unmanned field artillery system, which have already been existed and applied in military field. Implanted human programming and arithmetic rules, they are all crystallization of modern science and technology, and thus they are all artifacts and with automaticity. However, "autonomy" of RWAELW refers not only to automatic, but further to self-determination and self-choice that is out of man's control and manipulation. There are two great differences between "RWAELW" and the other artifacts such as robot and unmanned aerial vehicles. The former has two distinctions: (1) their activities out of man's strict control and (2) capacity of self-learning. Robot and unmanned aerial vehicles, etc are not only made by human

[1] Pagallo, Ugo, "Robots of just war: a legal perspective", *Philosophy &Technology* 24.3 (2011): 307--323; Lin, Patrick, George Bekey, and Keith Abney. *Autonomous Military Robotics: Risk, Ethics, and Design.* California Polytechnic State University, San Luis Obispo, 2008.

[2] *AlphaGo, "*Using machine learning to master the ancient game of Go", Demis Hassabis, Google Official Blog, January 27,2016; D. Silver, A.Huang etc, "Mastering the game of Go with deep neural networks and tree search", *Nature*, 2016, 529(7587):484-489

beings, but also strictly manipulated through programming and telecontrol. In the strict sense, they do not have efficient autonomy capacity but only man-controlled instruments[3], Even those already implemented self-motion and semiautomatic weapon systems, whose applied range, target and intensity are still strictly restricted by previous programming, belong to human warfare instrument. While RWAELW are not purely human instruments, and they are "quasi-subjects" with certain learning ability, autonomy, self-determination and free will.[4] They can be installed some kind of setting process or arithmetic rules in advance. They can also receive remote commands (such phenomena widely exist in human daily life, such as that the pilots or frontline soldiers in military actions report and accept remote commands to obtain certain information or be empowered to adopt specific actions in time), but they have the following functions of "self-deciding" and "self-choice":

a. acting without telecontrol, not purely mechanical beings;

b. employing weapon autonomously, namely employing weapons independently without remote control or qualitative algorithm program restriction;

c. determining a target autonomously;

d. judging autonomously;

e. deciding to start fight button independently.

If unmanned aerial vehicles, etc. can be called "robot warriors," then RWAELW can be called "quasi-man warriors; if the former are only extending of man's sensual and somatic functions, the latter will not only have those extending functions, but also have the function of "thinking" and "judging", similar to that of man's mind.

Of course, the self-judging, self-determination and free will of "quasi-man warriors" can be further explained in two senses: strong and weak. In the weak sense, it means that quasi-man warriors can communicate and accept human commands, but they can act "expediently." In the strong sense, the free will of self-judging, self-determination and self-choice means quasi-man warriors are able to act independently of human commands (for instance, they can perceive, judge, act expediently without information communication). Whatever, as long as acting not in accordance with previous given program, logic, rules, but "expediently" through "learning," quasi-man warriors can be considered owning the ability of elementary self-judging, self-determination, self-choice and free will. It is such "quasi-man warriors" who can learn and act expediently that put forward a series of unprecedented ethical problems. If such quasi-man warriors have no sense of good and evil, no basic moral values of justice and peace, and are short of ethical judgment, then should such artifacts have the ability of autonomously applying lethal weapons? Should human beings empower such artifacts the capacity of autonomously applying lethal weapons?

[3] Lin. Patrick, George. Bekey, and Keith. Abney. *Autonomous Military Robotics: Risk, Ethics, and Design.* California Polytechnic State University. San Luis Obispo, 2008. P21.

[4] Here free will refers to RWAELW autonomously determine striking target, autonomously make decision to launch lethal weapons out of human control.

II. WARRIORS：KILLERS OR HEROES OF DEFENDING PEACE AND JUSTICE?

RWAELW will bring grievous ethical challenge. Of course we can envisage technically to "teach" them how to distinguish right and wrong, and pursue right and avoid wrong to some extent, and can be some kind of "moral robots".[5] However there exist not only complex technical problems, but also esoteric humanistic and psychological problems.[6] At least during the predictable time, they can not truly become "moral subject". They have no sense of right and wrong, good and evil, as well as sentimental ability.[7]

Warfare has existed since ancient times. However, in a long history of civilization, especially after the catastrophes of two world wars in the 20th century, a new understanding of war has gradually formed: although war is inevitable in a predictable period, the essence of war should not be the tools of aggression and killing, but maintain peace and pursue justice. Based on such social dominant culture trend, a new recognition of soldiers in the battle has appeared.

Soldiers in the war certainly have responsibility to destroy the enemy, while soldiers are not killers in the just war, but heroes defending peace, justice and humanity. In order to prevent war and killing, only can soldiers have the sense of basic value spirit of justice, peace and humanity, can they be not descended to simple killers. According to current developing trends in present science and technology, human beings are radically different from RWAELW. Human beings not only have the sense of good and evil, right and wrong, as well as kindness, mercy, and love of lives, but also have moral sentiment or dispositions such as sympathy, compassion, love and hate. Comparatively, robot warriors not only have no sense of values, no spirit of kindness, justice, peace and humanity, but also have no moral sentiment or dispositions of sympathy, compassion, love and hate. Man can become hero because of pursuing justice, peace and humanity, while RWAELW can only become purely killers.

Of course, not all members of human society have the value spirit of kindness, justice and mercy. Some of them are anti-human immoralists, such as terrorists, who have vicious values and commit anti-human activities by all means. In this sense, they are purely destroyers of human peace, justice and humanity. Just because of this, human societies are united to deprive those anti-human immoralists of their ability of action and to prevent their violent behavior from taking place by appeal to all legitimate means so as to maintain peace, justice and humanity. Neither the fact that man with moral conscience, moral sentiment and passion may commit murder and hurt

[5] Wallach.Wendell and Colin. Allen, *Moral machines: Teaching robots right from wrong*, Oxford University Press, 2010.

[6] Michael Pritchard, "Moral Machines?", *Science and Engineering Ethics* (2012) 18:411-417.

[7] Sentimental ability can be divided into two types: strong and weak. In the strong sense, it belongs to moral sentiment, such as sympathy, love and hate with moral property. In the weak sense sentiment behaves as laugh or cry. Here it refers in the strong sense RWAELW have no moral sentimental ability. In fact, man has already manufactured robots with certain kind of emotional response. See Coeckelbergh. Mark, "Moral appearances: emotions, robots, and human morality." Ethics and Information Technology 12.3 (2010): 235--241.

others, nor the existence of evil human beings and even they have ability to autonomously employ lethal weapons can be used as reasons to justify the existence of RWAELW with no concept of good and evil and moral sentiment.

In reflecting on the brutality of Fascism during the Second World War, Hannah Arendt put forward the concept "banality of evil,"[8] by which Arendt attack those Fascism who attempted to make use of the excuse that soldiers must obey orders to defend their behaviors. Arendt reasonably point out: man who exists with free will cannot abandon thinking independently. They are not puppets who just obey orders, and so the excuse that soldiers must obey orders should not disculpate their malignant behavior. Arendt's reflection adheres to the mainstream ideology of modern European. As everyone knows, the spirit of freedom is one of fundamental achievements of human civilization. Man is free, and freedom is the ultimate value goal of human beings. As Aristotle revealed, all human behavior should turn towards chief goodness just like Archery is always aimed at the target. Kant, who is well known for his realizing Copernicus Revolution of philosophy, has enlightened people through the categorical imperative of universalization to realize that man is the end and man is free. [9]It is the spirit of "freedom" that distinguishes man from animal and gives meaning to man's life and activities. Arendt's attack to "banality of evil" and her critique of obeying orders without any value reflection seriously reveal that soldiers in present human civilized society should not be regarded as manipulated machines, but as human beings with basic value spirit, conscience, and responsibility for their behaviors. In other words, in present civilized society, even soldiers in the war should have basic value spirit of justice, humanity and peace, fighting for civilization, peace, justice and progressive cause.

The soldiers who defend for civilization, peace, justice and humanity should be warmhearted with humanitarian spirit, instead of being hardhearted with inhumanity. They love life, and revere for life. In ancient times, eastern Mencius revealed that man has sympathy, mercy and sense of compassion. In ancient Greece, Plato and Aristotle revealed that man has affection. In the period of European enlightenment, moral sentimentalists, such as Hume and Adam Smith, indicated that man has sympathy and empathy. All these thinkers uncovered the basic value of moral affection in human life in their own ways. The commander in the film Saving Private Reyn risked eight soldiers' lives to save Reyn based on his humanism. He did not want to see the sufferings that Mrs. Reyn would have from losing all her four sons. The plots in the film, including soldiers in participation released an arrested German soldier, the timidity of Urban, etc., all illustrate nothing but the humanity and human moral affections.

Nevertheless, in the predictable future, robot warriors built on artificial intelligence cannot have values and moral affections. Even robot warriors who have autonomously learning capacity to some extent can hardly form man's sense of good and evil and moral affection. Thus an acute problem appears: should a being who has no value, no moral affection have the ability of autonomously applying lethal weapons? If it possesses such capacity, do human beings have the ability to control it so as to shun its damage to humans themselves? Or more precisely, do humans have the capacity to control it so as not to let it become a purely brutal killer with no sense of good and evil? If the answer is negative, what does all this mean to human beings?

III. Do Quasi-man warriors have judgment?

Even if RWAELW have some kind of knowledge of good and evil, they can hardly have moral judgment. [10]Such judgment has multiple meanings, including: 1. capacity of transforming abstract ideas to concrete cognition, judgment and choice for concrete action in concrete situations; 2. capacity of making concrete value judgment and choice when values conflict; 3. cognition, judgment and choice in emergency situation; 4. capacity of telling friends from enemies, spite from kindness or harmlessness, and hypocrisy or fraud from chaffing.

Such judgment doesn't refer to judgment based on formal or fuzzy logic, but to the capacity of applying general rational laws to specific situations, as Kant and Hegel described. It is the ability that links general to specific, universal to particular. According to Aristotle's classification, military actions belong to the area of practical reason, and practical reasoning is vague and indeterminate.[11] That is, even man finds out a reason end for action based upon his good will and reason, but concrete activities to realize such an end are always depending on concrete objects, conditions, and situations. A good concrete activity must be appropriate or moderate between extreme choices in concrete situations.[12] It is possible to do bad things with good intention if the agent is ignorant of concrete situations. It is similar to the situations where doctors treat patients or soldiers fight enemies. A doctor can't cure diseases if he only has professional knowledge but lacks ability to adjust therapy plan according to patients' conditions. Likewise, a soldier can't destroy the enemy if he only fights hard but is incapable of judging wisely according to various situations in the battlefield. Even setting up the goal to fight with terrorists and making careful arrangement beforehand, a soldier still need to be flexible during execution. When it came to Kant's deontology, Hegel approved of its sublime ideology while intensely criticized its defects, namely "empty formalism" of "obligation for obligation."[13] Actually, in Critique of Judgment, even Kant himself regarded judgment as man's one of the most important abilities.[14] In regard to concreteness and creativeness in practical activities of the agent, Aristotle, Kant, Hegel, etc. revealed to us at the level of philosophy: a soldier without judgment cannot fulfill obligations successfully, and realize the end of goodness.

---

[8] Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil*, the Viking Press, 1965.

[9] Kant, *The foundation of moral metaphysics*, §2.

[10] Michael Pritchard, "Moral Machines?", *Science and Engineering Ethics* (2012) 18:411-417.

[11] Aristotle, *Nicomachean Ethics*, 1104a.

[12] Aristotle, *Nicomachean Ethics*, 1106b-1107a.

[13] Hegel, *Elements of the philosophy of right*, Verse 133,135.

[14] Kant, *Critique of Judgment,* Preface to the First Edition,1790.

In concrete military actions, conditions change quickly, often coming with sharp value conflicts. In Saving Private Reyn, a soldier raised the question of whether saving one person with the price of more people is reasonable, which manifests such value conflicts. Whether Reyn should be saved depends on specific humanities, value standpoints, affective attitudes, intellectual abilities of different agents, and their particular judgments in the specific situation. It is the agent's ability of judgment that directly determines his specific behavior choice. We can reasonably assume that in anti-terrorist activities when the sniper aimed at a terrorist, and is about to pull the trigger, when a child running out, at the moment the sniper decisively gives up shooting to avoid hurting the child. If the agent changes to a "quasi-warrior", does it have the ability to choose to terminate shooting immediately? For another example, in more complicated activities, how to correctly distinguish between enemy and friends, soldiers and civilian without any relative physical information, how to correctly make out who is the "undercover on our side" among many enemies, especially with no signals; how to differentiate real surrender and pretending one; how to distinguish between well-meaning joke and antagonistic intention. Obviously wise judgment is the essential condition for successfully fulfilling the military actions.

Human activities are a kind of creative processes with imagination.[15] What specific actions would be in specific situations is an open question, and the possibilities of action are unlimited, and effective activities of the agent depend on the concrete cognition and judgment in specific conditions. Man has sentiment, reason and judgment. He is able to distinguish between truth and falsity, to take measures corresponding to specific situations and changes of the attitudes of the rivals, and to determine whether the surrenders and the injured are real. All of these may be beyond the ability of RWAELW. Thus, even in the weak sense, it is still doubtful for "RWAELW" to successfully fulfill the task without breaking through the bottom line of humanity and justice.

IV. Can Man Control the "Autonomous" Creature Employing Lethal Weapons?

Before the appearance of the Robot Warriors, some premised restricted principles on the manufacture and use of robots have already been put forward.[16] Two of them are the most important: the principle of no harming human life and the principle of obedience to man's command, which have priority in value sequence. Obviously the core of these two principles is human being: for and under the control of human being. Whereas, the appearance of the RWAELW would totally break through these two restrictions from the root. RWAELW implies that the "quasi-man soldiers" could autonomously hurt human beings; it also implies that it could break away from man's direct control and command, radically challenging man's current ethical ideas and relations. The core of such a challenge lies in:

1. Should modern military activities stick to the basic value spirit like peace, justice and humanity? If the answer is positive, how can we guarantee that those "quasi-soldiers" who have no sense of good and evil, no moral sentimental capability will abide by the value spirit of peace, justice and humanity? If RWAELW have the ability of autonomously choosing to kill man, can the principle of humanity as man's moral bottom line still work as before? If the humanistic principle be shaken, what can be man's spiritual home? How can the golden rule of "never do to others what you would not like them to do to you" and Kant's "common sense" possibly exist?

2. According to Aristotle, appropriateness is the characteristics of practical reason.[17] Now the trouble is whether a "quasi-man soldier" without the sense of good and evil and moral sentimental ability can really achieve "appropriateness." If it cannot, what it means to human beings? Should man be more cautious and alert in creating such a "quasi-man soldier"?

3. Can man control quasi-warriors, and in what sense can or cannot? If man does create a kind of creature with self-judging, self-determination and free will, which is different from human beings and not purely instrumental any more but with certain kind of its purpose, what does it all mean to human beings? Can man rein and live with them peacefully? If human beings themselves cannot avoid lots of conflicts and warfare among different nations, religions and interests, who can guarantee such creatures be peaceful with man, and not be hostile to us?

4. Is man the only subject with subjectivity and free will on the earth? If the answer is yes, then how can a robot warrior without free will be empowered autonomy? If the answer is no, what is the relationship between human beings and RWAELW? Can the Robot Warriors autonomously order and dominate human beings? Even as some kind of restricted subjects, RWAELW may imply the mutation from purely instrumental robots to the ones with subjectivity unconsciously. The consequences the Robot Warriors could bring about may be much more than we could imagine, and might completely overthrow our current social living order.

Perhaps we can finally create RWAELW technically, but in any case, on one hand, all that is real doesn't necessarily mean it is reasonable. On the other hand, "can do" does not necessarily mean "should do."[18] Technical feasibility can not be the reason of our overlooking the problem of RWAELW. On the contrary, rationality requires us to confront such great challenges and deal it properly.

V. Some Conclusions

It cannot be ethically defended for "RWAELW." Here are some basic conclusions:

1. Robot warriors can be employed at war, but only as instruments, and forbidden to have the ability of "autonomy" to "autonomously employ lethal weapons." Any robot warrior or intelligent machine should be regarded as the extension of man's will, and should obey man's will.

[15] John Dewey, *Human Nature and Conduct*, chapter 16,19.
[16] Asimov, I. (1942). "Runaround", *Astounding Science Fiction*, PP.94-103.
[17] Aristotle, *Nicomachean Ethics*, 1108b-1109a.
[18] Gao Zhaoming, "Technological and ehical disenchantment: the limits of ethical rationality in modern life Science', *Social Sciences in China*. Autumn 2003(36-44).

2. To protect man's dignity and subjectivity and to defend the humanistic spiritual bottom line, no RWAELW should be developed and employed.

3. The international community should reach an international convention as soon as possible, banning the exploitation and employment of RWAELW, just as facing possibly destructive disaster CBWs (Chemical and Biological Weapons) might bring, the international community reached a consensus on the prohibition of employing chemical weapons in war.

REFERENCES

[1] Lin, Patrick, George Bekey, and Keith Abney,*Autonomous Military Robotics: Risk, Ethics, and Design*,California Polytechnic State University, San Luis Obispo, 2008.

[2] Pagallo, Ugo. "Robots of just war: a legal perspective", *Philosophy &Technology*,24.3 (2011): 307--323.

[3] AlphaGo, "Using machine learning to master the ancient game of Go, Demis Hassabis", Google Official Blog, January 27, 2016.

[4] D. Silver, A.Huang etc, "Mastering the game of Go with deep neural networks and tree search", *Nature*, 2016, 529(7587):484-489

[5] Wallach. Wendell and Colin Allen, *Moral machines:Teaching robots right from wrong*, Oxford University Press, 2010.

[6] Michael Pritchard, "Moral Machines?" *Science and Engineering Ethics* (2012) 18:411-417.

[7] Coeckelbergh. Mark. "Moral appearances: emotions, robots, and human morality." *Ethics and Information Technology* 12.3 (2010): 235--241.

[8] Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil*, the Viking Press, 1965.

[9] Kant, *The Foundation of Moral Metaphysics*,§2.

[10] Kant, *Critique of Judgment*, Preface to the First Edition,1790.

[11] Aristotle, *Nicomachean Ethics*, 1104a, 1106b-1107a, 1108b-1109a.

[12] Hegel, *Elements of the Philosophy of Right*, Verse 133,135.

[13] John Dewey, *Human Nature and Conduct*, chapter 16,19.

[14] Asimov, I. (1942). "Runaround", *Astounding Science Fiction*, PP.94-103.

[15] Gao Zhaoming, "Technological and ethical disenchantment: the limits of ethical rationality in modern life science", *Social Sciences in China*. Autumn 2003(36-44).