Dilemmas in the Use of Autonomous Weapons

Gabi Siboni and Yoni Eshpar

This past year saw a heated debate in the international arena on the ethical implications of the use of autonomous weapons. The debate illustrates that at issue is a category of weapons that as it develops is expected to confront not only complex technological difficulties, but also challenges of moral and legal legitimacy.[1] Therefore, it is very important for decision makers today to prepare for these challenges on all levels, and a broad public discussion that encompasses the relevant technological, strategic, legal, and philosophical issues is a key component of this preparation. The current article is intended to contribute to such a discussion.

In the background of any discussion of autonomous weapon systems is a rich foundation of cultural anxiety about the moment when science and technology make it possible to produce humanoid machines. From the Golem to Frankenstein in the early nineteenth century to movies like *Metropolis*, *Blade Runner*, *The Terminator*, and *The Matrix*, the story remains the tale of unanticipated dangers inherent in the impulse to replace human beings with machines. Science fiction writer Isaac Asimov has called the fear of these dangers "the Frankenstein complex." He has also proposed a solution in the form of an ethical system with three laws, the first being that robots may not harm human beings. This is the rule that today, seventy years after its first mention in science fiction literature, is the focus of the debate on the possible future entry of real killer robots into the battlefield.

Dr. Gabi Siboni is a senior research fellow and head of the Military and Strategic Affairs Program at INSS. Yoni Eshpar is a political analyst and strategic consultant for Israeli and international organizations, specializing in security, human rights, public advocacy, and development issues.

This cultural background also perhaps explains why in the debate on the autonomy of weapon systems, there is an excessive focus on its most extreme implementation: fully autonomous offensive weapon systems, or "killer robots." This tendency and the horror scenarios it invites force a rigid dichotomy on the discussion – "autonomy in weapons: yes or no" – at the expense of shared thinking on autonomy as a capability that can be integrated into weapons to differing degrees, for specific applications and usually very partially.

This article will attempt to present a wider range of possible positions. It first defines a framework for discussion of autonomous weapons capabilities, and then presents several of the main operational and ethical-legal issues involved. In conclusion, it proposes a number of steps that decision makers in Israel should take today so that it will be possible, in the medium and long terms, to integrate autonomous capabilities in a manner that will effectively serve Israel's security needs without compromising its commitment to values and obedience to the law.

Autonomy in Weapons

The Red Cross defines autonomous weapons as having the ability to "search for, identify and attack targets, including human beings, using lethal force without any human operator intervening."[2] Such a system has still not been developed, and it is difficult to say with certainty if it will be possible in the future. Thus far, any assessment of the capabilities and limitations of such systems is merely speculative. However, autonomous capabilities can be partial and exist at lower levels. One of the basic divisions is a ten-rung hierarchy between automation and autonomy,[3] but for the purposes of the practical discussion of autonomous capabilities of unmanned systems, the autonomous hierarchy can be limited to three main levels.[4] The first level comprises systems that are fully operated by human beings. In general, these are standard platforms in which the steering system and operation of the weapons have been adapted for remote operation, such as engineering equipment and remotely operated armored combat vehicles.[5] The second level comprises semi-autonomous systems, some of whose tasks are operated autonomously with a "man in the loop" – a central term in the discussion – such as the ability to move from point to point on a predetermined axis while relying on a range of sensors. The scope of autonomous execution is expanded with the integration of human operation that relies on a situation snapshot

transmitted to the operator by means of sensors. This family of systems includes unmanned airborne vehicles, unmanned ground vehicles operated for defined tasks, and remotely operated marine vehicles. The third level comprises fully autonomous systems with fully autonomous ability to operate throughout the mission. In the military environment, there are no such weapons because of operational and technological limitations. The graduated manner in which autonomy is implemented leads to a situation in which the principle of "human in the loop" can be interpreted in a number of ways.[6]

An additional way to illustrate the hierarchy between automation and autonomy is to use the example of a mine[7] that is located at the edge of the spectrum, on the side of automation. It explodes when pressure is placed on its operating mechanism, and it cannot tell what sets it off. More sophisticated mines use more sensitive sensors and have a greater ability to analyze the information they receive and are therefore better at distinguishing between targets. They explode only when the vehicle or naval vessel that they sense meets predefined criteria such as weight, magnetism, or conductivity. We can say that these mines are on a higher level of autonomy. Continuing on the scale could lead, for example, to a future system that using a variety of sensors would be capable of identifying military targets entering a defined area and intercepting them with little or no supervision by a human operator. Such a system would meet the operational need to channel enemy movement in a defensive battle and free soldiers for other tasks, but unlike mines, could be deployed and removed from the territory relatively quickly.[8] At the same time, the more autonomous the imaginary system is, the more it raises sensitive ethical and legal questions that are explored below. Suffice it here, however, to illustrate two points: The first is that the transition from automation to autonomy is largely driven by the operational and ethical need for systems with greater powers of distinction. The second is that this transition can be expected to encounter more operational and ethical problems the more it gains higher levels of autonomy. These points reinforce the assessment that implementation

> Israel has a clear interest in promoting local and international mechanisms that will give legitimacy to the use of autonomous capabilities in weapon systems within the framework of the ethical restrictions to which it is committed.

of autonomous capabilities in weapons will ultimately remain at a midpoint between automation and full autonomy.

Operational Aspects

Much has already been said about the military advantages of increased use of unmanned systems: they reduce the danger to soldiers and carry out tasks not limited by fatigue, hunger, heat, cold, pressure (physical or psychological), the need for oxygen, anger, anxiety, and the like. Technological developments bolster the autonomous capabilities of these systems in a manner that frees the operators from carrying out some of the tasks. For some time, the IDF has been using systems with such capabilities for performing routine security operations along the border with the Gaza Strip.[9] The combination of autonomous capabilities and miniaturization technologies will allow the use of a variety of sizes and types of unmanned systems, including miniature tools that can move quickly in conditions in which remote operation by a human being is more difficult, such as dense urban areas and complex underground spaces.

The development of autonomous capabilities for weapons makes it possible to combine unmanned capabilities and platforms with conventional forces.[10] This would result in a hybrid force that uses unmanned systems and systems with autonomous capabilities of different kinds on a variety of platforms. An example is a combination of engineering systems tailored for remote operation that integrate autonomous movement capabilities into the fighting force. These systems are regularly operated by a human being. However, during an especially dangerous mission, such as clearing a route in a minefield, operation of the tool would exploit its autonomous capabilities but would include a remote human operator. Armored vehicles adapted for remote operation could also be integrated into the force and used in advance guard duties, for example, operating in a way that combines autonomous capabilities with human operators. An example would be securing off an area, with movement and information collection carried out autonomously, while the human operator monitors the operation and opens fire if necessary.

One of the ways to analyze operational benefit is to examine it through the principles of war. Such an analysis shows that the operational benefit of autonomous capabilities on the battlefield is built on strong foundations, given the principles of war of the IDF and other armies.[11]

Several principles that are relevant in the context under discussion are mentioned here.

Economy of force: The use of autonomous capabilities provides an additional layer for the fighting force and allows a significant expansion of its capabilities. It can improve its operational effectiveness and make execution of the mission more efficient, for example, while reducing the amount of collateral damage. In addition, it allows manpower to be utilized more effectively, with a limited number of operators using a large number of tools.

Initiative and offense: Systems with autonomous capabilities can provide a layer of operation in the context of patrols, intelligence gathering, logistical support, and deception operations while the fighting force organizes to carry out the next missions. The use of similar systems can produce offensive moves even when the force suffers from serious burnout and fatigue.

Subterfuge: The use of systems with autonomous capabilities for purposes of identifying the enemy's weak point and for missions that support subterfuge makes it possible to significantly expand the tools in the commander's possession. These tools will also help to enhance surprise by operating in areas − or from directions − that the adversary did not expect.

Concentration of effort: As in the principle of economy of force, here too the use of autonomous capabilities in the battlefield makes it possible to free up troops for the main operation.

Continuity of action: The use of unmanned tools that with the aid of autonomous systems can operate for long periods of time (for example, hovering in the air in the target area until conditions are ripe for an attack) will enhance the force's ability to produce ongoing and prolonged pressure on the enemy. In this context, see initiative and offense (above).

Depth and reserves: The contribution of autonomous capabilities to improved defensive systems and more effective use of manpower has direct implications for the depth and reserves of the military effort.

Security: The use of autonomous capabilities to protect the force, and in particular, for preliminary construction of an intelligence picture regarding the development of threats to the rearguard and the home front, makes possible better implementation of the principle of security.

Against this background, it is easy to understand the great interest among militaries in integrating autonomous capabilities into weaponry.

The technology provides the promise of dramatically reducing the risks to the fighting force while increasing its effectiveness and its advantage over the enemy in a wide range of scenarios. Yet along with these clear benefits, concerns arise about the implications of increasing dependence on technology at the expense of the human element. At the same time, however, those who today are most critical of the current trend speak less about the operational implications and more about the moral challenges that autonomous weapons could pose to humanity in the distant future, or more specifically, about key questions, including: Will this technology necessarily lead to the day when autonomous war machines make decisions about life and death without human intervention? Are we prepared for this to happen? The discussion below presents the main positions in this ethical and legal debate.

Ethics and Law

Discussion of the ethical and legal implications of autonomous weapon systems gained significant momentum following the publication of two documents in late 2012. The first was a position paper written by Human Rights Watch (HRW) in cooperation with the Harvard Law School Human Rights Clinic,[12] which called for an international treaty that would place a blanket ban on the development, sale, and use of autonomous weapons. Shortly thereafter, the US Department of Defense issued a directive for all offices under it, including those responsible for developing, testing, and approving weapons.[13] According to the directive, the Defense Department's policy is "that autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force."

Publication of these two documents gave rise to a heated public debate in international legal and media circles. While broad differences of opinion emerged, there was also widespread agreement about clear cases that are at the extremes of possible scenarios. For example, the authors of the HRW position paper would presumably agree that a naval defense system installed on a ship on the high seas that intercepts missiles aimed at it, even without waiting for approval from a human being, is not problematic from a legal or ethical point of view. On the other hand, even enthusiastic supporters of incorporating autonomous capabilities in weapons would express skepticism about the morality of a scenario whereby robots alone undertake a mission that involves fighting

in a built-up area in which combatants hide among a civilian population. It is agreed, then, that the increase in autonomy in weapons could also lead to dangerous situations, and that therefore, technical, ethical, and legal restrictions are in order.

First Approach: The Emergency Brake

For the purpose of this overview, we grouped the most prominent positions in the debate into three main approaches. The first, called the "emergency brake," is most clearly represented by the HRW report, which demanded that the speeding train of autonomy be stopped, immediately and definitively, through an international treaty that would ban the development, sale, and use of autonomous weapon systems. Christof Heyns, the UN special rapporteur on extrajudicial, summary, or arbitrary executions, examined the issue and reached a similar but somewhat more moderate conclusion. In his report, published in April 2013, he recommended a freeze on efforts to develop autonomous weapons all over the world until an agreed international framework on their future is formulated.[14]

From the philosophical point of view, the members of this camp echo Asimov's laws, arguing that there is an inherent moral flaw in granting a machine legitimacy to decide whether a human should live or die. Politically speaking, it is argued that technology would provide leaders with the possibility of fighting wars without risking the lives of their soldiers, and this would be an incentive to choose military options over a policy of dialogue and avoidance of conflict. This criticism is similar to that directed at the expanding use of remote controlled unmanned weapons, but with the added concern that adding autonomous capabilities to these systems would prompt increased use.

From the technological perspective, the main argument is that autonomous weapon systems will never be able to select and strike targets on the basis of an ability to analyze a complex situation, identify human nuances, and use basic instincts of mercy, identification, and morality, as human beings are able to do. In addition, some experts have warned that there is a large, difficult-to-predict area between the intentions of the developers and operators of autonomous weapon systems and their ultimate behavior in practice.[15] For these reasons, it is argued, these are offensive weapons that will never meet the standards of distinction and proportionality required by international law, and

therefore, by definition, they are not legal. Another legal argument is that with autonomous weapon systems, it is not clear who should be held responsible in the event of a glitch or a violation of the law. Consequently, a vacuum of responsibility or room for immunity could be created around autonomous offensive weapons.

The choice of the emergency brake is explained by saying that any other monitoring regime will not succeed in preventing illegal uses of this technology or its leakage to state and non-state actors that do not consider themselves bound by any restrictive norms of use. The main source of inspiration for this recommendation is the Ottawa Convention, signed in 1997, which bans the production, storage, transfer, and use of anti-personnel mines.[16]

The Second Approach: Wait and See

The second approach argues that at this stage the debate itself is pointless, since despite the rapid development of technology, fully autonomous offensive weapon systems are still far off. Any attempt to formulate ethical norms or binding legal tools on this issue today would be a speculative exercise in science fiction that could cause more harm than benefit. Therefore, the responsible and cautious position is to wait and see. In other words, first let the technology train arrive at the station called "fully autonomous weapon systems," and then see how to restrict its implementation and use in combat situations in order to effectively protect moral and legal principles.[17]

To proponents of the wait and see approach, it is the emergency brake approach that suffers from an inherent moral flaw. The demand to ban increasing autonomy in weapon systems may well prevent the development of tools that could carry out combat missions with less harm to civilians than human combatants would cause. Machines, according to this argument, are not afraid to die, and they are not likely to have a destructive desire for revenge when a friend next to them is killed in battle. They can be programmed so that in cases of doubt, they will respond only when fire is opened on them. It is possible that in the future, they will even be able to identify targets better than human beings and respond more rapidly and accurately and cause less collateral damage. In addition, increasing autonomy through an army's unmanned weapon systems will reduce the risk to its soldiers, which will serve the moral principle that aspires to minimize killing and human suffering in wars.

Another criticism of the opposing camp focuses on the idea of a treaty. It argues that there is no chance that countries will sign a sweeping treaty on such a wide range of weapons that do not yet exist. Past successful experiences prove that it is possible to formulate broad agreement on removing a type of weapon from use only when it is clearly defined and the results of its use are already known. Even if all the technology superpowers signed such an agreement, this would not stop the race to develop tools with autonomous capabilities for a wide range of civilian applications. Thus, the technology is likely to be available, and there will be those who will not hesitate to convert it for military uses as well. In such a case, a situation could develop in which terrorist organizations and rogue armies use autonomous tools in an unrestrained manner, while law-abiding armies are left without the ability to defend themselves using similar means.

The Third Approach: Controlled Containment

The third position in the debate is located between the two positions described above. It recognizes the destructive potential of autonomy in weapons, but also its positive military and moral potential. Its supporters agree that a sweeping and immediate ban on development of such a technology is excessive and has no chance. However, they also share the concern that if the ethical and legal discussion is postponed to an unknown time in the future when the technology "matures," it will already be too late to impose effective restrictions on it. Recently, the Red Cross has expressed such an opinion.[18]

The starting point of the third approach is the assessment that autonomous capabilities will affect every area of our lives, and that this will happen gradually and in a modular fashion that will make it difficult to draw a clear line between automation and autonomy.[19] All of us, apparently, will need to adjust our ethical conceptions to a world in which there are self-driving cars and other autonomous machines that travel in our blood vessels, look for survivors in ruins, and the like.[20] Therefore, if to continue with the metaphor of a technology train, while the train cannot be stopped completely, it needs a constant and controlled use of brakes that will prevent accidents. Thus, the focus must be on developing a tool for regulating military applications of autonomous systems in parallel with their development.

These control tools can take shape only as a result of a process in which, in combination with hardware, software, and combat doctrine development, the ethical boundaries are clarified and agreements are formulated nationally and internationally on norms of use of autonomous systems and legal frameworks that will enforce them. Supporters of the third approach recommend to countries developing autonomous weapon systems that they take the initiative even in arenas where they are generally hesitant to discuss technological developments with strategic sensitivity − the international arena and the public arena. If the only two positions shaping public opinion on the issue are the "emergency brake" and the "wait and see," the result is likely to be either unrealistic international norms or no such norms at all. In the first case, democracies will encounter, both at home and abroad, serious difficulties of legitimacy for the use or sale of autonomous weapons they have developed. In the second case, it will be easier for less democratic technology superpowers and private companies to develop autonomous weapons that are not subject to agreed ethical standards and sell them without restriction.

Conclusion

A review of the main positions in the discussion on autonomous capabilities in weapons shows that the question of killer robots is not the sum total. While there are those calling for a coordinated international move to stop the development of autonomous weapons and ban their use in the future, it appears that the more common opinion does not support such sweeping moves. According to this position, any excessively decisive ruling, positive or negative, concerning the ethics and legality of the use of autonomous capabilities would be rash and possibly even harmful at this stage.

The integration of autonomous capabilities into weapon systems can potentially bring tremendous military benefit while maintaining the accepted legal standards and sometimes even meeting higher moral standards. The harnessing of these operational and ethical benefits will be dependent on developing legal and political tools that will effectively curb dangerous technological developments and prevent immoral use of weapons with autonomous capabilities. This is a long process, and it must take place at the same time as, and with a deep connection to, the development of the technology, the design of its applications, and the formulation of the relevant combat doctrines.

Israel is one of the leading countries in the development and integration of autonomous capabilities in weapon systems. Therefore, it already has a clear interest today in promoting local and international mechanisms that will give legitimacy to the use of such systems within the framework of the ethical restrictions to which Israel is committed, and at the same time, will make it difficult for other actors to develop, sell, and use similar technologies in a manner that is not bound by those restrictions. This will meet Israel's determination, like that of any democracy, to maintain a technological, military, and moral advantage over its adversaries. In the spirit of the third approach described here, the public and legal discussion on autonomous weapons should be directed to the position between the emergency brake and wait and see. In order to do so, it is not necessary to reveal technology secrets or operational plans. The following are a number of points with recommended action items.

First, domestically, the security establishment should set guidelines for all officials under its supervision who deal with operational specifications for the development of weapons with autonomous capabilities. This is similar to the directive from the US Department of Defense. In addition, it would be desirable to include in this process experts in law and ethics so that considerations from those perspectives will be inherent to the process of specification, development, and operational integration of these systems. Second, we suggest initiating cooperation with international security and legal officials who deal with the subject in other democratic countries, which share the same values and have a similar interest in formulating consensual approaches to the development of autonomous capabilities in weapons and to their operational use. Israel could even take the lead on some of these processes internationally. Finally, the value of the public debate must be emphasized. Therefore, it is proposed that the security establishment make public some of the security debates by taking the initiative and launching open discussions on the various meanings of increased autonomy in weapons and the ways of confronting the ensuing challenges and dangers. This would make it possible to strive for a constructive debate with critical positions in order to seriously address the deep and longstanding fears aroused by the idea of autonomous robots.

Notes

1 Liran Antebi, "Who Will Stop the Robots?" *Military and Strategic Affairs* 5, no. 2 (2013): 61-77.

2 ICRC, "Autonomous Weapons: States Must Address Major Humanitarian, Ethical Challenges," September 2013, http://www.icrc.org/eng/resources/documents/faq/q-and-a-autonomous-weapons.htm.

3 Raja Parasuraman and Thomas B. Sheridan, "A Model for Types and Levels of Human Interaction with Automation," IEEE Transactions on System, Man and Cybernetics" − Part A: Systems and Humans, Vol. 30, May 2000.

4 There are also other divisions. See US Department of Defense, "Unmanned Systems Integrated Roadmap FY2011-2036," p. 46.

5 Another example of this family of systems is aircraft controlled from the ground. The inaugural flight of an F-16 removed from operational service and converted by Boeing so that it could be flown without a pilot from a control room on the ground was recently reported. In this case, the purpose was to train pilots in live fire at a target plane in an aerial battle.

6 For a detailed explanation of the term "loop" in the context of the hierarchy between automation and autonomy in weapons, see William C. Marra and Sonia K. McNeil, "Understanding 'the Loop': Regulating the Next Generation of War Machines," *Harvard Journal of Law and Public Policy* 36, no. 3 (2013). Available at SSRN, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2043131.

7 Alan Backstrom and Ian Henderson, "New Capabilities in Warfare: An Overview of Contemporary Technological Developments and the Associated Legal and Engineering Issues in Article 36 Weapons Reviews," *International Review of the Red Cross*, October 22, 2012. Available at SSRN, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2198826.

8 The complexity of the use of mines and the difficulties in clearing them years later is described in John Walker, "Minerats: Moore's Law in the Minefield," IEEE Asilomar Microprocessor Workshop, April 19, 1995, https://www.fourmilab.ch/minerats/asilomar95.html.

9 Matthew Humphries, "Israel Patrols Its Borders with 10 Autonomous Off-road Vehicles," *Geek*, March 20, 2013, http://www.geek.com/news/israel-patrols-its-borders-with-10-autonomous-off-road-vehicles-1543460.

10 Steve Jameson and Jerry Franke, Lockheed Martin Advanced Technology Laboratories, and Robert Szczerba and Sandy Stockdale, Lockheed Martin Systems Integration − Owego, "Collaborative Autonomy for Manned/Unmanned Teams," American Helicopter Society 61th Annual Forum, Grapevine, TX, June 1-3, 2005.

11 There is a great deal of similarity between the IDF's principles of war and those of other armies. The main similarity concerns the universality of these principles. See Yaakov Amidror, "The Principles of War in Asymmetric Combat," *Maarachot*, No. 16 (December 2007): 4-11, http://maarachot.idf.il/PDF/FILES/1/112271.pdf.

12 Human Rights Watch and International Human Rights Clinic of the Human Rights Program at Harvard Law School, *Losing Humanity: The Case against Killer Robots*, November 2012. Available at http://www.hrw.org/reports/2012/11/19/losing-humanity-0.

13 US Department of Defense, "Department of Defense Directive: Autonomy in Weapons Systems," November 21, 2012. Available at http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf.

14 United Nations, "Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns," April 9, 2013, http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf.

15 Noel Sharkey, "Americas Mindless Killer Robots Must Be Stopped," *The Guardian*, December 3, 2012, http://www.theguardian.com/commentisfree/2012/dec/03/mindless-killer-robots.

16 The treaty is considered a success story, partly because within a relatively short time, a large number of states – to date, 161 – have signed it. There are other historical examples of the effective use of international treaties and agreements to prevent the use of weapons that cause unnecessary suffering, such as exploding bullets and blinding lasers.

17 Michael N. Schmitt and Jeffrey S. Thurnher, "'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict," *Harvard Law School National Security Journal* 4, no. 2 (2013): 231-81, http://harvardnsj.org/wp-content/uploads/2013/05/Vol.4-Schmitt-Final.pdf.

18 ICRC, "Autonomous Weapons: States Must Address Major Humanitarian, Ethical Challenges."

19 Marra and McNeil, "Understanding 'the Loop.'"

20 Research firm Gartner believes that "smart machines" with autonomous capabilities will have a deep impact on society and the economy by the end of the current decade. See Gartner Inc., "Gartner Says Smart Machines Will Have Widespread and Deep Business Impact Through 2020," Press Release of October 10, 2013, http://www.gartner.com/newsroom/id/2605015.