When Asimov's Robots Encounter the Laws of War

by Michael H. Hoffman

It's been culturally ingrained since 1942 that robots should never harm human beings. Isaac Asimov first introduced his famed laws of robotics in a science fiction story published that year. They stand the test of time as an influence on popular thinking. The modern, "transhuman" movement is pressing for artificial enhancement of natural human abilities. The view that robots should do no harm is now complemented by an emerging view that engineering of human beings should do no harm either.

In consequence, military legal and ethical standards are undergoing healthy scrutiny to determine if they are sufficient to address emerging artificial intelligence (AI) capabilities, and whether these will be complemented by a system sufficient to maintain command and control over them.[1] Not yet getting as much attention are ethical implications should AI and transhuman warfighters gain some measure of unplanned for autonomy. Beyond that, other challenges calling for attention are ethical implications should feedback from AI and transhuman warfighters adversely influence military decision making. The ethical implications of decisions and actions taken by autonomous AI and transhuman military actors, and their potential influence on military decision making, is the focus of this paper.

Norms for Military Artificial Intelligence

The cultural foundation for modern exploration of ethics and robotics first appeared in 1942 in Asimov's story "Runaround" which ultimately found its way into his famed novel *I Robot*. The rules are as follows.

> "We have: One, a robot may not injure a human being, or, through inaction, allow a human being to come to harm."
>
> "Right!"

Michael H. Hoffman is an associate professor with the U.S. Army Command and General Staff College. He holds a J.D. from Southern Methodist University School of Law. His current research focuses on the legal and ethical implications of advanced military and space technologies.

"Two," continued Powell, "a robot must obey the orders given it by human beings except where such orders would conflict with the First Law."

"Right!"

"And three, a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws."

"Right! Now where are we?"[2]

Where we are, almost 80 years later, is the rapid integration of artificial intelligence and robotics into military capabilities, art and science. Asimov's own stories demonstrated potential problems in the implementation of these rules. However, in spirit, his Rules of Robotics are still alive and well, with debate underway as to how such principles can be prospectively modified for application in the employment of military technology and in compliance with the modern law of armed conflict.

The fundamental summary of the law of armed conflict is found in "Common Article 3" found in each of the four Geneva Conventions of 1949. Though this article is commonly interpreted as applying to internal armed conflicts within states, the principles apply to all military armed conflict. It sets out requirements that endure in all forms of warfare including machine augmented military operations.

ART. 3. In the case of armed conflict not of an international character occurring in the territory of one of the High Contracting Parties, each Party to the conflict shall be bound to apply, as a minimum, the following provisions:

1) Persons taking no active part in the hostilities, including members of armed forces who have laid down their arms and those placed *hors de combat* by sickness, wounds, detention, or any other cause, shall in all circumstances be treated humanely, without any adverse distinction founded on race, colour, religion or faith, sex, birth or wealth, or any other similar criteria.

To this end, the following acts are and shall remain prohibited at any time and in any place whatsoever with respect to the above-mentioned persons:

a) violence to life and person, in particular murder of all kinds, mutilation, cruel treatment and torture;

b) taking of hostages;

c) outrages upon personal dignity, in particular, humiliating and degrading treatment;

d) the passing of sentences and the carrying out of executions without previous judgment pronounced by a regularly constituted court, affording all the judicial guarantees which are recognized as indispensable by civilized peoples.

2) The wounded and sick shall be collected and cared for.

An impartial humanitarian body, such as the International Committee of the Red Cross, may offer its services to the Parties to the conflict.

The Parties to the conflict should further endeavour to bring into force, by means of special agreements, all or part of the other provisions of the present Convention.

The application of the preceding provisions shall not affect the legal status of the Parties to the conflict.[3]

These principles of the law of war are the foundation for a substantial body of law. The law of armed conflict is sometimes applied in two categories; "Geneva law" applying to care and protection for the wounded, sick, and shipwrecked of armed forces, prisoners of war, and civilians, and "Hague law" applying to regulate means and methods of war, to include weapons systems and targeting. Harmonizing the

use of military robotics and artificial intelligence in a manner complimentary with Hague law is a challenge generating heated international debate.[4]

That debate goes beyond the scope of this paper, but it's important to note that it encompasses questions relating to use of artificial intelligence to make judgment calls on identification of military targets and risk to civilians. Potential uses for autonomous military machines (with use of unmanned aerial vehicles being the first category to gain attention in this dispute) generate particularly fierce controversy. Some commentators question whether machine intelligence is capable of decision making consistent with the law of armed conflict.[5] Also calling for consideration are the challenges of AI that gets loose and makes its own decisions, even with no intent by its designers to build in such capabilities.

Some commentators question whether machine intelligence is capable of decision making consistent with the law of armed conflict.

If military AI assets become truly autonomous, our closest analogy will come from the ethical and legal challenges presented by operations with coalition partners. Law of war treaties do point towards some obligation by armed forces to take measures ensuring humanitarian compliance and restraint by allies and coalition partners. Similar requirements need to be anticipated when armed forces launch AI that, with or without planning by its designers and operators, takes itself out of the human decision making process. Specific law on continuing responsibility for law of war violations, after a handover of capabilities or responsibilities to an ally or coalition partner is limited, but some guidance is available for analogous situations that could arise involving artificial intelligence.

The Hague Convention of 1907 Respecting the Laws and Customs of War on Land remains a foundation for the modern law of war. That Convention's Regulations Respecting the Laws and Customs of War on Land establish that any armed force, constituting part of an army legally qualified to wage war, is obligated to follow the rules. "The laws, rights, and duties of war apply not only to armies, but also to militia and volunteer corps…In countries where militia or volunteer corps constitute the army, or form part of it, they are included under the denomination 'army.'"[6] The Hague Regulations of 1907 were certainly adopted without reference to responsibility for the actions of artificial intelligence, but international law advances by custom as well as treaty. Thus, by analogy to established legal responsibilities, powers that build and launch autonomous systems platforms would be well advised to anticipate that they will be legally culpable for attacks conducted by that technology in contravention to the laws of war.

Another suggestion of continuing responsibility comes from the Geneva Convention relative to the Treatment of Prisoners of War of 12 August 1949. When a Detaining Power transfers prisoners of war to the custody of another power, this does not free it from all obligations for the protection of those POWs. If the transferee power "fails to carry out the provisions of the Convention in any important respect, the Power by whom the prisoners of war were transferred shall, upon being notified…take effective measures to correct the situation or shall request the return of the prisoners of war. Such requests must be complied with."[7] This provision also points towards a broader principle likely to be applied in future machine warfare. Ceding or losing control over autonomy capable machines does not absolve combatants from responsibility to ensure that the technologies make decisions and act in conformance with the laws of war.

Norms for Transhuman Warfighters

While unmanned aerial vehicles are already bringing forward legal and ethical issues in the employment of artificial intelligence, scrutiny of such issues in relation to the employment of transhuman warfighters-artificially augmented combatants-still lags behind. Though real problems will likely emerge soon enough, science fiction still remains a source for cautionary tales on the perils of artificial human augmentation. This paper does not address the many benefits flowing from therapeutic medical technologies employed to treat injuries and illness. Rather, it focuses on artificial augmentation designed to boost natural human capabilities. There is a history of combatants using drugs to maintain energy and keep their mental edge, with dubious clinical results and ethical implications that have been explored elsewhere.[8] The prospect of using medical technology to "build" warfigthers with enhanced physical and mental capabilities is a new one and we do not have a long history to draw upon for insights.

However, a far seeing cautionary tale first published in 1948 did anticipate some of the more extreme transhuman prospects taking form in the non-fictional 21st century. In *Scanners Live in Vain*, Cordwainer Smith depicted an elite outer space security force, composed of re-engineered, artificially enhanced members who suffered devastating psychological effects and dislocations from reality in consequence of being cocooned in enhanced artificial casings. "Martel noticed that he alone relaxed. The others could not know the meaning of relaxation with the minds blocked off up there in their skulls, connected only with the eyes, and the rest of the body connected with the mind only by controlling non-sensory nerves and the instrument boxes on their chests."[9] This vision may be moving towards reality with the prospect that bio-engineering could foster a merging of biology and technology to create transhuman warfighters.[10]

If transhuman combatants disengage psychologically from their units and other service members, we need to consider the prospect of distorted decision making, and actions taken in defiance of Rules of Engagement and obligations under the laws and customs of war. Such scenarios would not absolve their governments from obligations to ensure transhuman combatant compliance with the laws of war[11] but does make it more difficult. We will now consider some of the problems inherent in the advance of military AI and artificially augmented transhuman warfighters.

The prospect of using medical technology to "build" warfigthers with enhanced physical and mental capabilities is a new one...

Military Ethical and Legal Implications of Autonomous Artificial Intelligence and Transhuman Warfighters

Closest to home, in our comprehension, are likely to be issues concerning the deployment of artificially augmented warfighters. Challenges will begin in the recruitment phase. Here are some of them.

The prospect of government paid transhuman augmentation may attract many recruits. There will be need to assess personality traits to determine if likely recruits for augmentation are also a good bet for disciplined, ethical conduct on the battlefield. Once they are accessioned, other issues will arise.

Artificially enhanced capabilities may raise the bar on expectations of situational awareness that supports rapid ethical decision making in fast changing operational environments. If so, this also raises the prospect of legal liability exceeding that of other combatants. Other issues could also proliferate.

Transhuman warfighters may require

unique medical support taking forms that do not yet exist. When their deployment or service ends, weighty ethical and legal questions await on when, if, or how they can be "unplugged" from enhanced capabilities. Other questions will arise on the impact such decisions have on these augmented service members and their wider communities. Stranger scenarios involve the prospect that transhuman warfighters will engage, intellectually and emotionally, with AI systems and disengage from human ones.

...there is a risk that human operators will give up their own ethical and legal reasoning in deference to artificial intelligence...

Transhuman warfighters who have integrated into machine centric systems may decide, whether by accurate assessment or delusion, that machines rather than humans are their reference group. They may require tracking to determine that they are still engaged with other humans ethically as well as operationally. Though the capacity to maintain control over military AI is already under consideration, it needs to be asked whether efficiencies found in military AI might turn this around and begin influencing *all* human operators. Researchers have already identified the phenomena of "automation bias," meaning human deference to decision making conducted by automated systems, even when human operators are presented with evidence that those systems are in error.[12]

Systems should be designed to function in conformance with the laws of war even if human direction is cut off. AI may well sometimes generate superior analysis and decision-making. However, there is a risk that human operators will give up their own ethical and legal reasoning in deference to artificial intelligence, and AI generated ethical and legal problem solving outcomes may conflict with our own.

Conclusions

Care must be taken to ensure that AI takes direction from humans and does not turn this assumed paradigm around, becoming a negative influence on ethical human reasoning. However, AI must be designed to maximize the likelihood that if such systems do break loose from human control they will still function in compliance with the laws of war. Without careful monitoring and control, the introduction of AI warfare and transhuman warfighters may trigger negative effects difficult to contain.

A cautionary tale on ethically toxic systems that gain an enduring life of their own is set not far from our conference rooms at the Lewis & Clark Center. In August, 1863 a notorious massacre of civilians took place in Lawrence, Kansas during the American Civil War. The ethical consequences of that mass atrocity also shaped another massacre 70 years later. In June 1933, gunmen killed or wounded six law enforcement officers, and killed their prisoner in Kansas City's infamous Union Station massacre. Historian Paul Wellman noted "there is a weird sort of historical connection between the two crimes so far removed from each other in time, though so near in distance… not by blood, but by a long and crooked train of unbroken personal connections, and a continuing criminal heritage..."[13]

The prospect of autonomous military thinking machines and transhuman warfighters who drift from the orbit of military control, and unaugmented combatants who come under the dazzling influence of thinking machines raises the prospect of systems out of control. If this process starts, it may continue unabated and create its own human and ethical devastation reaching across generations. Sufficiently sophisticated systems may surprise us by splitting away in an operational sense even though that was never the intention for them.

We must build compliance with military ethics and the laws of war into AI functions, and ensure its prominence in transhuman military training and decision-making so that new systems do not over-ride the wisdom and functionality of many centuries-worth of military law and ethics. ***IAJ***

NOTES

1 Amitai Etzioni and Oren Etzioni, "Pros and Cons of Autonomous Weapons Systems," *Military Review*, Volume 97, No. 3, May-June 2017, pp.72-80. See also "Chances of UN Banning Killer Robots Looking Increasingly Remote," (March 25, 2019), accessed March 27, 2019, https://www.voanews.com/a/chances-of-un-banning-killer-robots-looking-increasingly-remote-/4847546.html.

2 Isaac Asimov. *I, Robot*. Del Rey Books. p. 37.

3 See, e.g. Geneva Convention (I), for the Amelioration of the Condition of Wounded and Sick in Armed Forces in the Field of 12 August 1949, art. 3, see also Geneva Convention (II), art. 3, Geneva Convention (III), art. 3, and Geneva Convention (IV), art. 3.

4 Jeremy Rabkin and John Yoo, *Striking Power: How Cyber, Robots, and Space Weapons Change the Rules for War*, Encounter Books, New York, 2017, pp. 141-152.

5 Ibid, p. 152.

6 Hague Convention (IV) Respecting the Laws and Customs of War on Land of 18 October 1907, Annex to the Convention, art. 1.

7 Geneva Convention (III) Relative to the Treatment of Prisoners of War of 12 August 1949, art. 12.

8 See, e.g., Norman Ohler, *Blitzed: Drugs in the Third Reich*, Boston: Mariner Books Houghton Mifflin Harcourt, 2018.

9 Cordwainer Smith, 1948. "Scanners Live in Vain" in *The Science Fiction Hall of Fame, Vol. 1, 1929-1964*, edited by Robert Silverberg, pp. 290-322. New York: ORB, 2003.

10 Nayef Al-Rodhan, "Transhumanism and War," Global Policy Journal, 18 May 2015, accessed Feb. 22, 2019, https://www.globalpolicyjournal.com/blog/18/05/2015/transhumanism-and-war.

11 See, e.g., Geneva Convention (I) for the Amelioration of the Condition of Wounded and Sick in Armed Forces in the Field of 12 August 1949, art. 49.

12 Linda J. Skitka, Kathleen Mosier, and Mark D. Burdick, "Accountability and automation bias," *Int. J. Human-Computer Studies* 52 (2000): 701, accessed March 27, 2019, https://lskitka.people.uic.edu/IJHCS2000.pdf.

13 Paul L. Wellman, *A Dynasty of Western Outlaws*, Lincoln, Nebraska: University of Nebraska Press, 1986, pp. 18-20.