Overlooked Implementation Issues with Autonomous Operations

By Major Scott E. Bruck, U.S. Army

Introduction

In February 2000, Congress mandated that one-third of all deep strike aircraft and one-third of all ground combat vehicles would be autonomous by 2015.[1] The law implementing autonomous and semi-autonomous vehicle development was intended to: ensure the safety of soldiers by sending robots into harm's way instead of soldiers; reduce soldier deaths due to improvised explosive devices (IEDs); decrease the need for "boots on ground;" and in some cases, create weapons that could outperform their human counterparts (e.g., some aerial drones can stay airborne for up to twenty-four hours).[2]

The problem with these seemingly perfect weapons is the lack of human decision-making. Although humans require more time than robots to process information, human decisions are still currently the safest decisions. The decision as to whether an incoming aircraft is friend or enemy needs an answer within a split second, and there are instances of air defense systems making these decisions autonomously with disastrous results. For example, the Russian Buk missile system that downed Malaysia Airlines Flight MH17 allegedly was equipped with an autonomous operations mode.[3] The Dutch Safety Board Chairman, responsible for investigating the crash, emphasized his shock when he stated, "The reasons are as straightforward as it is disquieting, nobody thought civil aviation was at risk."[4]

Removing humans from the warfare equation also has grave moral implications because an autonomous or semi-autonomous robot cannot take personal responsibility for, nor be personally accountable for, any atrocities it commits. That leaves the original computer programmer or remote operator or leader on the ground to assume such responsibility and accountability; however, this "blame line" would be both subjective and contentious with legal implications and one that soldiers would need to agree to in advance. For these and other reasons, the physical and moral boundaries must be addressed for the Army's use of semi-autonomous robotics in combat

Army Press Online Journal is published bi-monthly by The Army Press to provide cutting edge content on topics related to the Army and national defense. The views expressed belong to their authors, and do not necessarily represent the official view of the Department of the Army, the Department of Defense, or any other government institutions or agencies.

vehicles as well as other war fighting functions. First, in the physical domain, a boundary must be set so that all actions require a human decision, thus protecting soldier life. Secondly, a moral boundary must be set defining the chain of command and ensuring the informed consent of soldiers.

Physical Boundaries – Human Decision

The first physical boundary that needs addressing, for the Army's use of autonomous robotics is that human decision is required. The defense community has quantified this exact moment in doctrine through the observe-orient-decide-act (OODA) loop.[5] The OODA loop is a technique that quantifies a decision cycle. Both large-scale and small-scale processes can benefit from the OODA loop framework. For the purposes of this article, the OODA loop serves as a vehicle for showing exactly where a human decision needs to reside in the process.

An example of robotics technology conducting the observation, orientation and decide sub-processes is the software in the F-35 joint strike fighter. According to Deputy Defense Secretary Robert Work, in National Defense, "The F-35 is a flying sensor-computer that sucks in an enormous amount of data, correlates it, analyzes it and displays it for the pilot on his helmet."[6] The F-35 is capable of making decisions, but the pilot remains the sole decision maker of the "act" sub-process. The F-35 is an example of the current state of military robotics in that subprocess are carried out autonomously, while the "act" portion still rests with a human.

By default, when one supports the automation of the act sub-process they are assuming that every eventuality can be pre-programmed. Within the constraints of existing technology, artificial intelligence cannot truly think for itself. Currently, artificial intelligence is limited to data analysis such as correlation software. What one does with this data is still the purview of the human programmer and requires human interaction. In the "Unmanned Systems Integrated Roadmap", the Department of Defense states:

> Today's iteration of unmanned systems involves a high degree of human interaction. DoD must continue to pursue technologies and policies that introduce a higher degree of autonomy to reduce the manpower burden and reliance on full-time high-speed communications links while also reducing decision loop cycle time.[7]

An example of required human interaction is the Squad Mobile Support System (SMSS), an autonomous wheeled vehicle the size of a golf cart that can follow a squad while carrying their supplies. The SMSS deployed to Afghanistan, but was not completely successful. Because of the infinite amount of data required to negotiate every conceivable terrain feature, human assistance was required at times.[8]

Just like difficult physical terrain, the complex human terrain associated with modern conflict is challenging to adequately program for. Because of this difficulty robots currently operate under a high degree of human interaction. Regardless, every programing decision has the potential to increase danger for soldiers. Current trends in software engineering tend to include extremely large numbers of people working on the same project, which causes unintended interactions (emergence). Furthermore, an instance of emergence may "hide" throughout product testing and take time to manifest itself. Emergence is a normal part of daily life and joking for civilians, it is an inevitable result of complex programming. Emergence may not be a problem when it appears as a bug in a favorite video game; however, when discovery occurs during combat emergence may prove fatal.

Through proper utilization and planning, the powerful phenomenon of emergence can also be an asset. For example, programming each leg of the aforementioned SMSS system to follow basic functional rules, and not programming for every eventuality, would allow for each leg to act independently. This independence could allow each leg to deal with specific terrain it encounters regardless of what the other legs are encountering. In short, emergent actions would occur as each leg "learned" how to negotiate an obstacle independently. Harnessing emergence could relieve the need for thousands of lines of code and make the SMSS more attuned to the outside environment.

Another problem with automating the "act" portion of the OODA loop deals with security concerns. In an era where computer hacking is becoming increasingly prevalent, the Army will need to stop enterprising hackers from turning its military robots against itself. The ramifications of hacking credit card information pales in comparison to the possibility of hacking the Army's military robots. If the decision to engage is automated, a situation in which a line of code is changed could cause the Army's own military robots to support the enemy.

Peter Singer, in an interview with Karen Paul, summarizes this possible scenario:

> A human would think, 'That makes no sense, I'm questioning that order,' but a computer, if you have that access, will follow that instruction. It's a whole new realm where you've never been able to convince a bullet or an arrow to change direction in mid-flight, you can with this kind of system. That points to both possibilities and perils.[9]

The perils of which Mr. Singer states emphasize the ease with which binary decisions are manipulateable. Information security needs to become just as important as engagement criteria, if not more so. As the battlefield becomes more complex, the Army will continue to need a person to question any autonomous or semi-autonomous orders. Due to advances in networking, said person does not need to be necessarily on the same continent as the robot. The ease with which engagement criteria are reverseable makes it imperative that a human view these choices.

Physical Boundaries – Protection of Friendly Troops

The second physical boundary that must be addressed for the Army's use of autonomous robotics is the protection of friendly troops. A robot will have to decide who is friend or foe. Even today, we have instances of fratricide that occur when humans are making that decision. In the future, the Army needs to create a method where friendly troops are safe from engagement with their own tools. The Army could do this in one of two ways: first, internal code in the robot's software could identify friendly troops; or second, the troops themselves could wear interrogation markers. Until technology matures, markers represent an adequate option for protection.

An internal way to accomplish the identification of friendly troops would be to involve a technology referred to as computer vision. According to Wikipedia, "computer vision is a field that includes methods for acquiring, processing, analyzing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, *e.g.*, in the forms of decisions."[10] For example, soldiers' faces could be input into the robot's software and thus used as references for engagements. Unfortunately, face recognition software is not perfect and a soldier's face is rarely visible in combat. Regardless of the current state of face recognition software, it offers potentially extraordinary advantages for autonomous operations.

A second, much more practical method would be for friendly soldiers to wear a device identifying them as friendly, such as a radio frequency identification (RFID) tag. A military robot would interrogate the RFID tag before it would engage a target. Other

examples for marking friendly troops include everything from bar codes to certain camouflage patterns. As previously stated, much care would have to be taken in the prevention of possible hacking attempts before one of these techniques could be instituted.

Protection of friendly troops operating with military robots is not only limited to defining engagement criteria. If a military robot were incorporated into a squad, troops would need to be physically protected from a lumbering military robot weighing upwards to 1,000 pounds. Serious review, including doctrinal review, would need to occur before any type of autonomous or semi-autonomous military robot moves down to the tactical level. Because robotics is such a new technology, it needs testing in every possible environment. If the Army is to successfully incorporate military robots into its operating framework, soldier safety needs to be the number one concern, not only from engagement, but also from other less obvious threats.

Moral Boundaries – Accountability

The first moral boundary that must be addressed for the Army's use of autonomous robotics deals with the definition of accountability. The Pentagon just recently released a 1,204-page manual attempting to quantify the laws of war.[11] A reflection of society's morals is seen within the new manual's pages. Not written directly for military robots, the manual will nonetheless be a standard to measure their actions. Responsibility for a robot's actions, which contradict the manual, will need delineation in doctrine. Current precedent holds that the commander on the ground is ultimately responsible for all actions within his battle space. Robots have the potential to change this paradigm. The fielding of such a complex entity as a robot, with the ability to make decisions, involves a large number of personnel who are possibly responsible. People such as software engineers, manufacturers, or acquisition officials may legitimately be responsible for unanticipated deaths on the battlefield.

Defining who is morally responsible for military robots' actions is difficult. Holding robots to the same moral standards as soldiers will most likely reinforce current doctrine. Conversely, if robots begin to be seen as their own specific type of moral entity, their place in the chain of command will need to be re-evaluated. Before accountability can be defined, the Army will need to decide what exactly a moral machine is. According to Wallach and Allen, in their 2009 book *Moral Machines*, both a high level of autonomy and a

high level of ethical sensitivity are required for a machine to be moral (see Figure 1). Modern robots have low-level proficiency in both autonomy and ethical sensitivity.

For example, the lowest end of the spectrum is a basic machine, which Wallach and Allen describe as having "operational morality." A rifle with a safety mechanism is a machine with operational morality. Though the rifle does not have autonomy or ethical sensitivity per se, its mechanics reflect the ethical sensitivity of the creators to safety. The mere fact that a basic machine cannot end human life through accident of the user recognizes sensitivity to the fallability of human behavior.

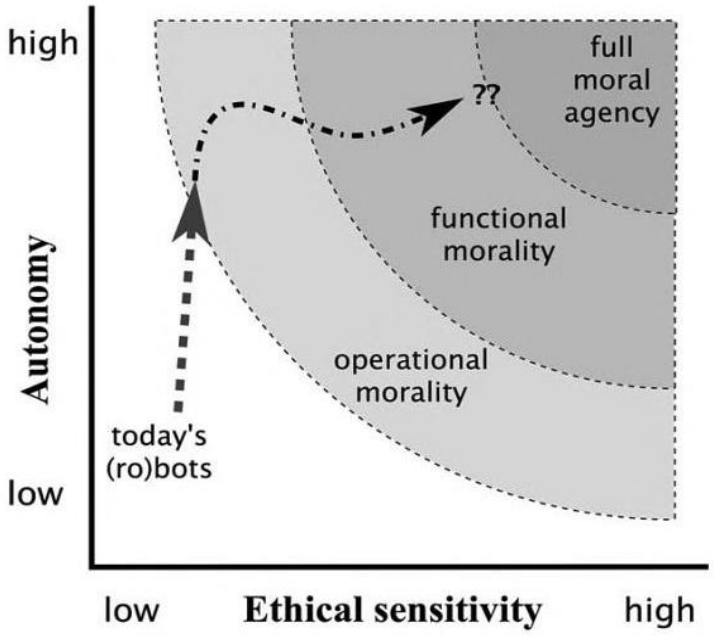

Figure 1

The next level of "functional morality" applies to machines that have either high levels of autonomy or high levels of ethical sensitivity, but not both. An example of a functionally moral machine, which demonstrates high levels of autonomy, would be an aircraft autopilot.[12] An aircraft under autopilot control continually makes decisions with no human input. This high level of machine autonomy is standard in the aviation community, but at no point is the autopilot making an ethical decision.

An example of a machine with high ethical sensitivity, but no autonomy, would be a clinical decision support system. An example of a clinical support system would be MEDETHEX, which is:

> [...] an ethical advisor that provides guidance to health care workers faced with ethical dilemmas. *MedEthEx* is an implementation of Beauchamp's and Childress' Principles of Biomedical Ethics that harnesses machine learning techniques to

> abstract decision principles from cases in a particular type of dilemma with conflicting *prima facie* duties and uses these principles to determine the correct course of action in similar and new cases [...].[13]

MEDETHEX attempts to use the power of past positive examples and apply them to current situations. When a doctor confronts a complex problem, he or she can use MEDETHEX to formulate an opinion. MEDETHEX's algorithms compare previous situations, which ended positively, with the current situation. The human doctor still has responsibility, but he or she also has a tool at their disposal that can quantify an ethical outcome. A large error in these types of ethically sensitive systems is that they only rely on past precedence for their decisions.

These types of applications already exist in the military domain. The Georgia Institute of Technology created a software package, which takes large-scale plans and executes them in a virtual environment. In order to advise the user on ethical conflicts, the Georgia Institute of Technology created an add-on application, which counsels users regarding the mission's ethical obligations and prohibitions.[14] So far, ethical support applications have not gained widespread acceptance throughout the Department of Defense.

It is the nature of the current complex environment to be difficult to predict. Black swan events, or events that cannot be adequately forecasted, have become much more the norm than the exception. Instilling morality into a program does not have to occur from only past precedence. Theoretical research also has used both the top-down implementation of morals as well as bottom-up. Top down implementation deals with comparing a current situation to a precedent, then making a decision. In bottom-up implementation, machine learning occurs, whereas the robot learns morality from doing.

The introduction of a safe military robot will require the machine to have both high levels of autonomy and high levels of ethical sensitivity. This level is called "full moral agency." Fully moral agents are the only truly safe option and one in which no one other than the military robot itself is responsible. Due to the current nature of autonomous and semi-autonomous technology, these fully moral agents do not exist yet.

Until fully moral agents exist, commanders on the ground are ultimately morally responsible for all military robots' actions. As such, commanders need all information pertinent as to the current state of military robotics. All too often commanders are interacting with a field service representative from a manufacturer (FSR). The FSR may not necessarily have the soldiers' best interest in mind, their

unstated goal is to sell more robots. An example of the current state of ignorance regarding military robotics entails combatant commanders sending urgent operational needs statements which detail technology that does not exist.[15] The current ad hoc incorporation of military robots into the force has not fully leveraged the doctrine, organization, training, materiel, leadership & education, personnel, facilities, and policy (DOTMLPF-P) model.

Understandably, there is a rush to incorporate anything that may save soldier lives. Commanders' good intentions may be doing more harm than good when they hastily incorporate robotic solutions onto the battlefield. It is because of a lack in such things as doctrine and training that current commanders are uninformed, which results in making them dangerously liable.

Moral Boundaries – Soldier Consent

The second moral boundary that must be addressed for the Army's use of autonomous robotics is the requirement for soldier consent. Currently, all dangerous occupations in the Army require a volunteer statement or an instance of informed consent. When soldiers willingly enter a combat zone, they understand the abstract ideas of imminent danger and enemy will. When soldiers volunteer for jump status, they understand the risks associated with aircraft and parachutes. At this time, however, the Army does not fully understand the risks to soldiers from incorporating autonomous military robots throughout the force.

The use of case studies and known fatal accidents involving autonomous robotics should be used for both training and consent. Mandatory basic computer science classes for soldiers in close contact with military robots might also be required. This training could help soldiers think like programmers and understand that robots' actions are derived from a programming language. Soldiers would also learn how the robot sees the world via sensors. Understanding input and output of autonomous robotics will undoubtedly make those who work with military robots safer and more confident.

Finally, true moral consent will only come with knowledge. Until soldiers can show a deep understanding and appreciation of the benefits and hazards of robotics - and in particular, a deep understanding of the robots they will be working with - an informed consent would be meaningless. Additionally, having soldiers complete an informed consent without the requisite knowledge base could derail the program.

Conclusion

The physical and moral boundaries that must be addressed for the Army's use of autonomous or semi-autonomous robotics include physical boundaries such as human decision and friendly troop protection, and moral boundaries, such as accountability and soldier consent. Due to the young age of robotics science, programming, development, training, and policy, the Army may not truly master or implement autonomous robotics in the near future, nor truly understand its related dangers.

NOTES

1. Floyd D. Spence National Defense Authorization Act for Fiscal Year 2001, Pub. L. No. 106-398, H.R. 4205 (2001), accessed 17 June 2015, https://www.congress.gov/bill/106th-congress/house-bill/4205.
2. Peter Singer, *Wired for War: The Robotics Revolution and Conflict in the Twenty-First Century.* (New York: Penguin Press, 2009), Chapter 1, Section 5, para. 4.
3. Bill Sweetman, "Buk Missile System Lethal, But Undiscriminating," *Aviation Week* online, July 23, 2014, http://aviationweek.com/defense/buk-missile-system-lethal-un2.
4. Tony Osborne, "MH17 Final Report: Investigators Confirm Russian-Made Buk Missile Shot Down 777," *Aviation Week* online, Oct 15, 2015, http://aviationweek.com/commercial-aviation/mh17-final-report-investigators-confirm-russian-made-buk-missile-shot-down-777discriminating.
5. Robert Coram, "OODA Loop," In *Boyd: The fighter pilot who changed the art of war*, (Boston, MA: Little, Brown, 2002), Chapter 24, Section 3.
6. Sandra I. Erwin, "Pentagon Will Develop 'Thinking Machines' to Defeat Future Enemies," *National Defense Magazine* online, Nov 8, 2015, http://www.nationaldefensemagazine.org/blog/Lists/Posts/Post.aspx?ID=2011.
7. James A Winnefeld, Frank Kendall, "Unmanned Systems Integrated Roadmap FY2011-2036," Office of the Undersecretary of Defense for Acquisition, Technology, & Logistics (August 2011): VI.
8. Rob Waugh, "What could possibly go wrong? U.S. Army Wants to Give War Robots More Power to 'Make their own Decisions,'" *Daily Mail* online, May 9, 2012, http://www.dailymail.co.uk/sciencetech/article-2141694/What-possibly-wrong-U-S-Army-wants-war-robots-power-make-decisions.html.
9. Kari Paul, "When the Killer Robots Arrive, They'll Get Hacked" *Vice News* online, 24 February 2015, http://motherboard.vice.com/read/when-the-killer-robots-arrive-theyll-get-hacked
10. Wikipedia contributors. "Computer vision." *Wikipedia, The Free Encyclopedia*, (Wikipedia, The Free Encyclopedia, 5 Nov. 2015) accessed 19 2015.
11. Ross Cypher-Burley, "Pentagon Unleashes Behemoth Bible Laying Out the Laws of War for US Soldiers," *Vice News* online, June 18 2015, https://news.vice.com/article/pentagon-unleashes-behemoth-bible-laying-out-the-laws-of-war-for-us-soldiers
12. Wendell Wallach, Colin Allen, *Moral machines: Teaching robots right from wrong* (Oxford: Oxford Univ. Press, 2010), 27
13. Michael Anderson, Susan L. Anderson & Chris Armen, "MedEthEx: a prototype medical ethics advisor," *Proceedings of the National Conference on Artificial Intelligence*, (MIT Press, 2006).
14. MissionLab v7.0. (n.d.), accessed 21 November 2015, http://www.cc.gatech.edu/aimosaic/robot-lab/research/MissionLab/

15. Singer, Chapter 3, Section 8, para. 19.

Fig. 1. Two Dimensions of AMA Development. Graph from Wallach, Wendell., & Allen, Colin.