Autonomous military robots: A short survey

Jason Lomberg, Technical Editor

The four soldiers move from their concealed position and “stack” themselves, one behind the other, parallel to the door. The #2 man throws an M84 grenade (flashbang) into the room, then yells “frag out!” The flashbang detonates, and the fire team storms through the “fatal funnel” (door). In those first moments, the #1 man must instantaneously: 1) Distinguish between combatants and non-combatants, and 2) Determine the most immediate threat. Failure to react properly could have dire consequences. This high-risk procedure is an integral part of MOUT [1] (Military Operations on Urban Terrain), which begs the question: In the future, could procedures like “room clearing” be performed by autonomous military robots?

At present, the only true autonomous military robot is the Navy's Phalanx Close in Weapon System (CIWS), seen here.

The benefits are clear: robots wouldn't experience the "fog of war," or fall prey to emotions. They wouldn't have an itchy trigger finger, experience fear, or get jittery. This is the ideal scenario, anyway; the one theorized by scientists, roboticists, intellectuals, and science fiction writers for decades, arguably centuries. In reality, aside from the Navy's Phalanx Close-In Weapon System (CIWS [2]), there are no true autonomous military robots. Some feel the concept of Artificial Intelligence (AI) is inherently flawed. The "first computer programmer," Ada Lovelace [3], observed in her notes [4] to the analytical engine [5], "The Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform. It can follow analysis; but it has no power of anticipating any analytical relations or truths." Lovelace's observation was remarkably prophetic.

True robot "consciousness" may be eons away, or impossible, but that hasn't stopped its development. The end stage includes robots acting as "full ethical agents," i.e. "those that can make explicit moral judgments"[1]. Some feel this is unnecessary. In a report prepared for the Office of Naval Research, Patrick Lin, George Bekey, and Keith Abney remarked [6] that, "an ethically-infallible machine ought not to be the goal now (if it is even possible); rather, our

goal should be more practical and immediate: to design a machine that *performs better* than humans do on the battlefield, particularly with respect to reducing unlawful behavior or war crimes" (emphasis theirs).

The challenge in creating artificial moral agents (AMAs) is threefold. It includes: ethical considerations, programming approach, and inherent risks. Do we really want robots to have free will (assuming it's possible)? Do we want military robots that can disobey orders? Frag a superior officer? Voluntarily defect? All these actions, though highly-illegal, are within the realm of human soldiers. How do we create AMAs? Would a deontological "top-down" programming approach work? Asimov's famous "Three Laws of Robotics [7]" springs to mind. Would an evolutionary "bottom-up" approach (i.e. a "learning computer") suffice? Or should we strive for a hybrid system? Finally, what dangers would autonomous military robots pose to non-combatants and friendly forces?

As the Naval Report [6] points out, "'robot' is derived from the Czech word 'robota' that means 'servitude' or 'drudgery' or 'labor.'" It's important that we retain this "slave morality" in the creation of autonomous robots. Such a robot could, "make autonomous choices about the means to carrying out their pre-programmed goals, but not about the goals themselves; they could not choose their own goals for themselves." Without a "slave morality," a robot could be unpredictably dangerous (picture The Terminator or The Matrix). Robots must follow their programmed goals. This should be the overriding principle in constructing autonomous robots. Autonomy should not extend past the commander's orders.

Most people are familiar with Asimov's "Three Laws of Robotics [7]," seen below.

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

A Library of Congress portrait of Isaac Asimov, author of "I, Robot," and "The Naked Sun," among others.

Asimov's Three Laws represent a deontological programming approach; the robot's decisions are governed entirely by codified rules. "Conditional statements," a fixture of computer programming, are another example. But Asimov's writings demonstrated the fallacy of such an approach. All laws contain loopholes. No one can possibly foresee every eventuality. A robot could follow its programming and still cause human harm. "The Naked Son" illustrates the Three Laws' key weakness—absence of the word "knowingly." Thus, a robot could unknowingly perform an action that causes human injury.

The opposite, a "bottom-up" approach, focuses on evolution and learning. Alan Turing, often considered the "father of modern computer science," posed the question [8], "Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain." But as Messrs Lin, Bekey, and Adney point out, "the algorithms and techniques presently available are not robust enough for such a sophisticated educational project." Without this education, such a robot would lack foundational understanding. Picture a child tasked with running a Fortune 500 company.

Human learning combines the top-down and bottom-up approaches. Our parents/educators/religion/society set out rules ("Don't lie, cheat, or steal," "Thou shalt not kill," etc.) and through experience, we gain "wisdom." Autonomous robots should follow this hybrid model. It should have both preprogrammed rules, and the capacity for learning. Without a mechanism to speedily "educate" robots, experience will be paramount. Should we wait until robots are fully-autonomous, with ample "wisdom," before deploying them? This could take decades, possibly centuries. Controversy would arise as to how many lives could have been saved by

deploying robots in place of humans.

But as the report suggests, it isn't necessary for autonomous robots to exhibit a perfect track record. Rather, they need only perform "better than humans do on the battlefield, particularly with respect to reducing unlawful behavior or war crimes." Inevitably, first-generation autonomous robots will be buggy. Indeed, buggy software is as ubiquitous as the sun rising.

A solution would be to "operate combat robots only in regions of heavy fighting, teeming with valid targets…the Rules of Engagement are loosened, and non-combatants can be reasonably presumed to have fled, thus obviating the issue of discriminating among targets…providing a training ground of sorts to test and perfect the machines." The Phalanx CIWS is a perfect example. The Phalanx operates in a limited environment and does not need an identification, friend or foe (IFF) system. Instead, it uses multiple criteria (heading, velocity, and capability for maneuver, among others) to identify proper targets.

What about the danger to civilians and friendlies? Could robots ever grasp the nuances of the Laws of War (LOW) and the Rules of Engagement (ROE)? An immobile enemy soldier, stripped of weapons, is not considered a threat. If he's crawling away, he is considered a threat (and thus subject to lethal force). "Protected" locations, such as schools and mosques, lose their protected status if used for improper purposes (such as stockpiling weapons). Could robots appreciate these subtleties? Or will they only be useful in "target rich" environments?

A potential incongruity is "collateral damage." According to the DOD, "Such damage is not unlawful so long as it is not excessive in light of the overall military advantage anticipated from the attack." Would a robot, programmed to avoid civilian casualties, refuse any order that put civilians in danger? Could they be programmed to understand "acceptable" collateral damage? Should the bar for acceptable collateral damage be raised?

The risk of "fratricide" is potentially more troubling than collateral damage. In 2007, a semi-autonomous cannon deployed by the South African Army malfunctioned, resulting in nine "friendly" deaths. The Naval report indicates the need for extensive, realistic testing before the deployment of fully-autonomous robots. Combined with initial deployment in "target-rich" environments, the danger to friendlies should be minimal. Nonetheless, friendly-fire incidents are inevitable.

The military is always careful to delineate "acceptable" risk. In the case of autonomous military robots, I believe the potential benefits outweigh the risks. War is a nasty business, and the military is a dangerous profession. Deploying autonomous robots won't eliminate all risk, but it will significantly reduce it. But if one's overriding goal is to reduce risk, then remote-controlled devices would suffice. A crucial aspect of the LOW and ROE is having "eyes on target." The primary benefit of autonomous robots (over remote-controlled) is their ability to "see" the situation more clearly than humans operators could, seated thousands of miles away.

In the future, robotic systems will play an integral role in our military. Congressional

mandate has set the wheels in motion. By 2010, one-third of all operational deep-strike aircraft must be unmanned, and by 2015, one-third of all ground combat vehicles must be unmanned [National Defense Authorization Act, 2000]. It's our job as engineers, technologists, and observers to ensure the responsible development of autonomous robots.

Works Cited

1. Wallach, Wendell; and Allen, Colin. Moral Machines: Teaching Robots Right from Wrong. Oxford: Oxford University Press, 2009

Source URL (retrieved on *06/20/2013 - 1:51am*):
http://www.ecnmag.com/articles/2009/04/autonomous-military-robots-short-survey?qt-most_popular=0

Links:
[1] http://www.globalsecurity.org/military/ops/mout.htm
[2] http://www.raytheon.com/capabilities/products/phalanx/
[3] http://en.wikipedia.org/wiki/Ada_Lovelace
[4] http://www.fourmilab.ch/babbage/sketch.html
[5] http://en.wikipedia.org/wiki/Analytical_Engine
[6] http://ethics.calpoly.edu/ONR_report.pdf
[7] http://en.wikipedia.org/wiki/3_laws_of_robotics
[8] http://loebner.net/Prizef/TuringArticle.html