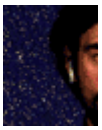
The New York Times
nytimes.com





**November 25, 2008**

# A Soldier, Taking Orders From Its Ethical Judgment Center

**By CORNELIA DEAN**

ATLANTA — In the heat of battle, their minds clouded by fear, anger or vengefulness, even the best-trained soldiers can act in ways that violate the Geneva Conventions or battlefield rules of engagement. Now some researchers suggest that robots could do better.

“My research hypothesis is that intelligent robots can behave more ethically in the battlefield than humans currently can,” said Ronald C. Arkin, a computer scientist at Georgia Tech, who is designing software for battlefield robots under contract with the Army. “That’s the case I make.”

Robot drones, mine detectors and sensing devices are already common on the battlefield but are controlled by humans. Many of the drones in Iraq and Afghanistan are operated from a command post in Nevada. Dr. Arkin is talking about true robots operating autonomously, on their own.

He and others say that the technology to make lethal autonomous robots is inexpensive and proliferating, and that the advent of these robots on the battlefield is only a matter of time. That means, they say, it is time for people to start talking about whether this technology is something they want to embrace. “The important thing is not to be blind to it,” Dr. Arkin said. Noel Sharkey, a computer scientist at the University of Sheffield in Britain, wrote last year in the journal Innovative Technology for Computer Professionals that “this is not a ‘Terminator’-style science fiction but grim reality.”

He said South Korea and Israel were among countries already deploying armed robot border guards. In an interview, he said there was “a headlong rush” to develop battlefield robots that make their own decisions about when to attack.

“We don’t want to get to the point where we should have had this discussion 20 years ago,” said Colin Allen, a philosopher at Indiana University and a co-author of “Moral Machines:

Teaching Robots Right From Wrong," published this month by Oxford University Press.

Randy Zachery, who directs the Information Science Directorate of the Army Research Office, which is financing Dr. Arkin's work, said the Army hoped this "basic science" would show how human soldiers might use and interact with autonomous systems and how software might be developed to "allow autonomous systems to operate within the bounds imposed by the warfighter."

"It doesn't have a particular product or application in mind," said Dr. Zachery, an electrical engineer. "It is basically to answer questions that can stimulate further research or illuminate things we did not know about before."

And Lt. Col. Martin Downie, a spokesman for the Army, noted that whatever emerged from the work "is ultimately in the hands of the commander in chief, and he's obviously answerable to the American people, just like we are."

In a report to the Army last year, Dr. Arkin described some of the potential benefits of autonomous fighting robots. For one thing, they can be designed without an instinct for self-preservation and, as a result, no tendency to lash out in fear. They can be built without anger or recklessness, Dr. Arkin wrote, and they can be made invulnerable to what he called "the psychological problem of 'scenario fulfillment,' " which causes people to absorb new information more easily if it agrees with their pre-existing ideas.

His report drew on a 2006 survey by the surgeon general of the Army, which found that fewer than half of soldiers and marines serving in Iraq said that noncombatants should be treated with dignity and respect, and 17 percent said all civilians should be treated as insurgents. More than one-third said torture was acceptable under some conditions, and fewer than half said they would report a colleague for unethical battlefield behavior.

Troops who were stressed, angry, anxious or mourning lost colleagues or who had handled dead bodies were more likely to say they had mistreated civilian noncombatants, the survey said (PDF). (The survey can be read by searching for 1117mhatreport at www.globalpolicy.org.)

"It is not my belief that an unmanned system will be able to be perfectly ethical in the battlefield," Dr. Arkin wrote in his report (PDF), "but I am convinced that they can perform more ethically than human soldiers are capable of."

Dr. Arkin said he could imagine a number of ways in which autonomous robot agents might

be deployed as “battlefield assistants” — in countersniper operations, clearing buildings of suspected terrorists or other dangerous assignments where there may not be time for a robotic device to relay sights or sounds to a human operator and wait for instructions.

But first those robots would need to be programmed with rules about when it is acceptable to fire on a tank, and about more complicated and emotionally fraught tasks, like how to distinguish civilians, the wounded or someone trying to surrender from enemy troops on the attack, and whom to shoot.

In their book, Dr. Allen and his coauthor, Wendell Wallach, a computer scientist at the Yale Interdisciplinary Center for Bioethics, note that an engineering approach “meant to cover the range of challenges” will probably seem inadequate to an ethicist. And from the engineer’s perspective, they write, making robots “sensitive to moral considerations will add further difficulties to the already challenging task of building reliable, efficient and safe systems.”

But, Dr. Allen added in an interview, “Is it possible to build systems that pay attention to things that matter ethically? Yes.”

Daniel C. Dennett, a philosopher and cognitive scientist at Tufts University, agrees. “If we talk about training a robot to make distinctions that track moral relevance, that’s not beyond the pale at all,” he said. But, he added, letting machines make ethical judgments is “a moral issue that people should think about.”

Dr. Sharkey said he would ban lethal autonomous robots until they demonstrate they will act ethically, a standard he said he believes they are unlikely to meet. Meanwhile, he said, he worries that advocates of the technology will exploit the ethics research “to allay political opposition.”

Dr. Arkin’s simulations play out in black and white computer displays. “Pilots” have information a human pilot might have, including maps showing the location of sacred sites like houses of worship or cemeteries, as well as apartment houses, schools, hospitals or other centers of civilian life.

They are instructed as to the whereabouts of enemy materiel and troops, and especially high-priority targets. And they are given the rules of engagement, directives that limit the circumstances in which they can initiate and carry out combat. The goal, he said, is to integrate the rules of war with “the utilitarian approach — given military necessity, how important is it to take out that target?”

Dr. Arkin's approach involves creating a kind of intellectual landscape in which various kinds of action occur in particular "spaces." In the landscape of all responses, there is a subspace of lethal responses. That lethal subspace is further divided into spaces for ethical actions, like firing a rocket at an attacking tank, and unethical actions, like firing a rocket at an ambulance.

For example, in one situation playing out in Dr. Arkin's computers, a robot pilot flies past a small cemetery. The pilot spots a tank at the cemetery entrance, a potential target. But a group of civilians has gathered at the cemetery, too. So the pilot decides to keep moving, and soon spots another tank, standing by itself in a field. The pilot fires; the target is destroyed.

In Dr. Arkin's robotic system, the robot pilot would have what he calls a "governor." Just as the governor on a steam engine shuts it down when it runs too hot, the ethical governor would quash actions in the lethal/unethical space.

In the tank-cemetery circumstance, for example, the potentially lethal encounter is judged unethical because the cemetery is a sacred site and the risk of civilian casualties is high. So the robot pilot declines to engage. When the robot finds another target with no risk of civilian casualties, it fires. In another case, attacking an important terrorist leader in a taxi in front of an apartment building, might be regarded as ethical if the target is important and the risk of civilian casualties low.

Some who have studied the issue worry, as well, whether battlefield robots designed without emotions will lack empathy. Dr. Arkin, a Christian who acknowledged the help of God and Jesus Christ in the preface to his book "Behavior-Based Robotics" (MIT Press, 1998), reasons that because rules like the Geneva Conventions are based on humane principles, building them into the machine's mental architecture endows it with a kind of empathy. He added, though, that it would be difficult to design "perceptual algorithms" that could recognize when people were wounded or holding a white flag or otherwise "hors de combat."

Still, he said, "as the robot gains the ability to be more and more aware of its situation," more decisions might be delegated to robots. "We are moving up this curve."

He said that was why he saw provoking discussion about the technology as the most important part of his work. And if autonomous battlefield robots are banned, he said, "I would not be uncomfortable with that at all."