Industrial Challenges of Military Robotics[1]

Professor Dr. George R Lucas, Jr.
Professor of Ethics & Public Policy
Naval Postgraduate School (Monterey CA)
&
Class of 1984 Distinguished Chair in Ethics
Stockdale Center for Ethical Leadership
U.S. Naval Academy (Annapolis MD)

Baron Karl von Clauswitz once cautioned that "every age had its own kind of war, its own limiting conditions and its own peculiar preconceptions." Ours is the era of "irregular" or unconventional war, together with the revolution in military affairs, including emerging new military technologies, that go along with it. I hope in this paper to marry concerns of military ethics and business ethics and address challenges to industry – not, of course, the engineering, technical, and production challenges of robotics generally, which are considerable – but rather the unique ethical challenges facing industry as a result of the demand for robots and unmanned systems with military applications.

The chief criticism leveled at robotics, or for that matter, all of the recent technological innovations in war, ranging from robots to non-lethal weapons, biological warrior enhancement and nanotechnology, and cyber warfare, is what I term the "threshold" problem. That is, inasmuch as all such technological innovations tend to render war itself less destructive and costly in terms of loss of life and injury to persons and property, these innovation also might inadvertently *lower the threshold* for resorting to war, or undermine the principle that war is to be resorted to as a method of conflict

resolution only as a last resort. The "threshold problem" infuses many of the arguments of critics of these innovations (e.g., Sparrow 2007, Lin *et alia* 2008, Singer 2009; 2010).[2]

That prospect, however, is primarily a policy question, not directly an issue for industry alone. The defense industry is, to a large extent, a servant or handmaiden of the policy decisions of its government. CEOs, engineers, and employees are, of course, also citizens who might be concerned to avoid increasing the incidence of war through their efforts, even as they labor to lessen its most damaging effects through their inventions. So this "threshold concern" is not utterly irrelevant as a consideration for them as well.

The ethical challenges I wish to consider for industry specially, however, come under the headings of *reckless endangerment* and *criminal negligence*. The associated legal challenge is "product liability," the normal commercial arrangements in domestic law under which a manufacturer is held financially responsible for damages or injury due to the malfunction of its products. There are several variant conceptions of product liability, and we might wonder which best captures the risks and concerns pertaining to military uses of robotics.

The classical, time-worn (rather than time-honored) conception of product liability is *caveat emptor*, "let the buyer beware." In this conception, producers and end users are considered equally knowledgeable and competent moral agents, and accordingly, are given equivalent responsibility for the safe use of industrial products. If a consumer, like the military, orders a robot product, and proceeds to misuse it, or to discover that its use has unintended negative consequences that are not strictly due to defects of manufacture or misrepresentation by the producer, then the responsibility for the results devolves strictly to that end user. The company is blameless.

From the producer or manufacturer's perspective, a more stringent concept of product liability, and one now widely adopted in many nations, is known as "due care." In this conception, the producers or manufacturers are recognized as inherently more knowledgeable about the products they design, manufacture, and distribute than the typical consumer or end user. They are responsible for anticipating reasonable risks of harm due to inadvertent improper or even incompetent use of their products, as well as harm or injury stemming from possible malfunctions or undetected product defects that subsequently come to light under normal operating conditions. They are responsible for warning consumers about these risks, and the dangers of improper use, and are liable for the harm that might otherwise result from the purchase and use of their products in ways not specifically cautioned against.

This second concept is the origin of the familiar, lengthy warning labels on even the most routine household products that manufacturers sometimes despairingly refer to as "idiot labels:" "warning, do not drink this household cleaning product as a beverage; do not place your head inside the washing machine during operation," or, with robots, "warning, do not use your lethally-armed autonomous security robot to entertain children at a birthday party." The manufacturer's frustrations stem from the inability within the law to distinguish the sublime or the sophisticated from the ridiculous, and so the customer or consumer must be presupposed an idiot (as some, regrettably, are), and warned against virtually every conceivable misuse (however unlikely or non-culpable these extremes may seem), lest the manufacture be held liable by some angry consumer for having "failed to warn" against even the most absurdly inappropriate misuses of their product.

A more controversial conception of product liability is known as strict or "no fault" liability. In this case, favored by populist consumer advocates, but not widely accepted in law, all of the responsibility for harm or injury in the use of a product devolves to its manufacturer. No attempt is made to determine fault. Instead, it is no longer risk of harm, but full or strict liability for harm, including injury or death, resulting from any use of the product whatever that falls to the producer or manufacturer. It no longer matters, in this conception, whether the risks of harm from unintentional or unforeseeable malfunction or misuse were reasonable or unreasonable, or whether or not the resulting harm was due to manufacturing defects. All liability falls directly to the producer, who ordinarily is presumed to build the considerable resulting legal and financial liability into the cost of the product. In conjunction with lethally-armed, autonomous robots, however, this seems a chilling prospect (as well as prohibitively expensive) for industrial producers to contemplate.

We might now consider how these background concerns manifest themselves variously in the debates about the feasibility and desirability of military robotics that have emerged in the recent debates about ethics among philosophers, computer scientists, and robotics experts. Their positions frame the context within which these industrial questions ought to be examined.

At one extreme, we have the position of computer scientist Noel Sharkey concerning the military uses of robotics, and the development and deployment of autonomous robots armed with lethal weapons in particular (e.g., Sharkey 2007, 2008; 2010). We may array alongside him many other prominent critics of these proposed military uses, including the Australian philosopher, Rob Sparrow (Sparrow 2007, 2008,

2009), and the eminent American political scientist, P.W. Singer (2009; 2010). In one way or another, all of these critics believe a great many of the proposed military uses of robotics are reckless and unwise in the extreme. They represent poorly thought-through and badly conceived military and public policy that needlessly and carelessly poses a grave danger to the public. I think it would not be a stretch to claim that these critics, collectively, believe such developments and military applications as currently proposed by enthusiastic or unreflective proponents to constitute something bordering on criminal negligence.

Rob Sparrow in particular claims product liability, even in its most stringent forms, is an absurd and inadequate conception of "accountability" to bring to bear upon mistakes and accidents that might occur in the international arena with autonomous lethal robots in combat. What sense would it make, he argues, in the aftermath of civilian casualties sustained during misuse or malfunction, to force the manufacturer to compensate the victims' families financially, and otherwise simply "recall" or withdraw the robot product? Such a mode of accountability and redress seems incommensurate with the degree of threat of harm posed, let alone damage and injury that might actually be done. Lacking any appropriate or meaningful conception of accountability, Sparrow maintains that any company that manufactured such instruments, or any government who deployed them, would be guilty of violations of the underlying provisions of international humanitarian law (IHL). This is because, at bottom, according to Sparrow, IHL holds generally that no nation or its militaries may purposively and knowingly develop a weapon over whose potentially indiscriminate or disproportionate destructive capacities

they would have no ultimate control, and for which no accountability for mistakes or accidents arising from their use can plausibly be assigned.

At the other end of this spectrum of opinion is the eminent American computer scientist and roboticist, Prof. Ronald C. Arkin (e.g., Arkin 2007; 2009; 2010), who maintains that the resort to armed autonomous robots would result in greater compliance with the laws of war, especially those laws pertaining to the protection of civilian noncombatants, known as the laws pertaining to discrimination or "distinction." Robots, in contrast to human warriors, Arkin argues in sum, don't get scared, don't get angry, and don't try to "get even" or exact revenge for the injury or loss of comrades in combat. As a result, they are more likely (when properly programmed) to comply with the laws of armed conflict than their human counterparts.

More generally, with regard to military necessity, force protection, and the economy of force, robots are more likely than terrified, angry, or just plain uncaring human agents, to adhere to the principles of "last resort" and "least possible force," and hence reduce the prospects for inadvertent and unintended noncombatant casualties. Robot warriors would be programmed with a "proportionality algorithm" that would, Arkin argues, permit a cool and calculated assessment of the least amount of force required to achieve a legitimate military objective without undue concern for friendly force protection. This would help sharply reduce the incidence of inadvertent noncombatant casualties arising from the use of excessive force, or from resorting to force too quickly out of fear or from an overly strict and conservative doctrine of force protection, and so minimize deaths due to mistakes and accidents in comparison with human counterparts.

In the middle, we have engineers and roboticists – like John Canning at the U.S. Navy's Surface Warfare Research Center in Dahlgren, VA (e.g., Canning 2008, Canning *et alia* 2004) – who are the "hedge funds" for this debate. Canning argues, in agreement with Arkin, in favor of greater robot autonomy, but wants to leave humans "in the loop" with executive oversight, and to arm robots either strictly with non-lethal weapons, or else with lethal force that would be directed primarily against an opponent's or enemy's weapons systems rather than its combatants themselves, aiming at disarming and rendering them ineffective, rather than destroying them. This, he argues, would reduce the likelihood of disproportionate harm done to enemy combatants, let alone mistaken casualties among noncombatants.

For engineers and policymakers in the U.S., in particular, this range of opinion does not leave much of a choice. Their decisions are ultimately driven by fiscal necessity. In 2001, the U.S. Congress ordered that 33% of all combat aircraft be unmanned by 2010, and that 1/3 of all combat vehicles be unmanned by 2015 (see OSD 2009). They have not mandated "autonomy" specifically in these cases, but as P.W. Singer argues, this is the only meaningful implication of such requirements, since otherwise we have merely increased the distance between operators and machines at a one-to-one ratio. The shortages of manpower and money that these new technological developments would address only make sense if robots are "force multipliers," thereby magnifying, through semi and fully autonomous function, what each individual soldier or sailor can accomplish. Dr. Siva Banda, Senior Scientist for Control Theory with the U.S. Air Force Research Laboratory, and a world-renowned UAV control expert, recently summed up the matter this way:

> "The military's use of UAVs is escalating. . .from a quarter of a million cumulative UAV flight hours during the twelve-year period from 1995 to 2007, to an additional quarter million flight hours in merely six months (from May to November, 2007). . .predominately [driven by] cost savings over manned systems. The main reason is the increasing cost to train pilots, but it also costs more and more to manage the escalating UAV flight hours. The solution is increased autonomy. We have to become less and less dependent on humans, to decouple UAVs from their human operators. . .To achieve this, UAVs in the future will need to be endowed with smart sensors and with tactical reasoning and decision-making capabilities. Without intelligence, which is the key enabler, we will remain tied to human operators for sensor interpretation and have limited influence." (Banda 2010)

For a policymaker, the only ethical question is thus a utilitarian one: the prudent expenditure of public funds upon the public welfare, including its defense. If it appears likely that robots "do it better, faster, and cheaper" than humans, or than trained military personnel specifically – or alternatively, if we will in the future simply lack the financial resources to provide adequately for national defense solely through reliance on human agents and operators – then, in either case, that appears to end the debate for policy makers.[3] If building armed, autonomous combat robots is what it takes to accomplish the mission with increasingly limited resources, and if the numbers are correct, it is a "no brainer."

Such reasoning does not even bother to dignify the challenges posed by the Sharkey faction, and this wholly utilitarian, instrumental policy perspective certainly does not begin to exhaust all the relevant moral and legal considerations. For my part, I tend to favor the middle position, the Canning faction, and I'll explain why, in terms of my doubts about the feasibility of Arkin's project. Let me return to this project.

Readers are doubtless familiar with the so-called "Turing Test" for artificial intelligence: we will have achieved the goal of simulating intelligence when the behavior

of a machine is indistinguishable from that of a human under similar circumstances. Note that Arkin has proposed (albeit inadvertently) a similar test for robot morality, or artificial conscience: we will have succeeded in developing a moral and legally reliable autonomous combat robot whenever the rate of compliance of the robot with the relevant laws of armed conflict equals or exceeds that of human warriors under similar circumstances. I have come to call this the "Arkin test" for robot morality, though Ron didn't much like the term when I first introduced it. There are several challenges with realizing this goal. Chief among them is the "character recognition" problem – what DoD scientist Siva Banda described as "smart sensors [coupled with] tactical reasoning and decision-making capabilities." The robot's sensors must be able to recognize and distinguish among several different kinds of inhabitants of any prospective battle space – soldiers, civilians, children, for example – as well as be able to recognize morally significant features of the behavior of each – say, an enemy combatant attempting to surrender (something for which there is, at present, no universally recognized sign). As challenging as this would be in itself, I do not doubt Arkin's, Siva's, and their colleagues' ability to one day address the character recognition problem.

But all this begs an even more difficult question, that is outside the scope of engineering *per se*. The Arkin test presupposes that the "laws of war" and even a specific military force's ROE's are inherently programmable. But they are not, at least not in their entirety, and certainly not in their most essential features. The problem lies not with Arkin and his colleagues in engineering and computer science. Rather, the fault is mine and my colleagues in law and ethics.

We don't have or use anything like a formal "proportionality algorithm," for one thing. Historically in just war doctrine and in IHL (more readily known among military and defense personnel as "the Law of Armed Conflict," or LOAC), we have proclaimed our commitment to a pseudo-quantitative sounding principle that, in declaring war in the first place, and then carrying out legitimate and justifiable military maneuvers, the collateral damage done, and damage done overall to everyone and everything by the war, must be found "proportional" to the good accomplished. There are a host of difficulties, however, that we encounter in the practice of operationalizing such principles, or in carrying out the required calculations, all of which are well known to just war theorists, and even more to the men and women we send into combat. In the *jus ad bellum* case, for example, the entire deliberation is a utilitarian estimate of future events, in which the tendency is always to minimize the negative consequences and overestimate the good results that will be realized from the pursuit of the preferred policy. This is an inherent problem with utilitarian calculus as a guide to human decision-making generally, and it is pervasive in policy calculations about war.

In the *jus in bello* case of otherwise legitimate and presumably necessary military maneuvers, furthermore, the case is even more problematic, because it requires military commanders in the field to weigh and compare incommensurable goods in order to determine the presumed quantitative ratio that any algorithm would require. How many children's lives, for example, is a given military objective worth? We have no precise idea, only some "sample scenarios" to refine prudent judgment. Consider the following two controversial examples of how difficult calculations of "proportionality" actually are to carry out.

Early in the morning on September 4, 2009, over 100 children and elderly villagers were accidentally killed in an air strike on two fuel trucks stolen by Taliban forces. The strike was called in by a commanding officer in the German contingent of NATO/ISAF forces. The order for the attack violated strict rules for engagement designed to prevent accidental killing of civilians, and thus was reckless, if not criminally negligent. Even had the threat assessment been more accurate, however, the collateral damage sustained in the attack seems wildly disproportionate, reasonably enough so to conclude that the attack should never have been ordered, even had the fuel trucks turned out to be a legitimate target, and even had the proper rules of engagement for calling in the air strike otherwise been properly followed.[4]

By way of comparison, the release by "WikiLeaks" of classified U.S. Army video showing at attack carried out by two Apache helicopter gunships on July 12, 2007 on fourteen alleged "insurgents" in Iraq (two of whom turned out to be Reuters News Agency employees) prompted international outrage in light of the uncertain identify of the victims and a resort to deadly force that was denounced as "disproportionate."[5] I find the second less grievous a violation of the algorithm than the first, both because of the numbers, and also because of the kind of reasoning involved. The Americans apparently thought they were attacking enemy insurgents, and that the journalist's large camera was a rocket-propelled, shoulder launched grenade. These are the kinds of mistakes that occur in war, and this incident was subsequently and thoroughly reviewed and investigated. By contrast, Col. Georg Klein and his contingent of German NATO troops apparently ordered the attack in deliberate disregard for protective procedures in place to order deadly force strikes, apparently motivated by the desire to impress their allies with

their decisiveness. That is reprehensible. But that is my view, in contrast to widespread public perception that the second event (judging by international public reaction to the revelation of the engagement video released on "Wiki-Leaks" in April, 2010) was somehow "worse than" the first, which has faded from public view since December, 2009. Regardless of how we view the degree of negligence involved in ordering and carrying out the two attacks, however, it remains the case that some fourteen persons, including two children, were mistakenly wounded or killed in the second incident, while over 100 Afghan civilians, including many more children, were mistakenly wounded or killed in the first. While both are terrible tragedies, certainly if "proportionality" has any kind of intelligible meaning in law or morality, the German case is several times (at least seven times) worse than the American Apache helicopter incident.

Such wildly inconsistent public perception and evaluation, even among scholars, is hardly reassuring about our ability to judge infractions of the law, let alone to establish reliable quantitative benchmarks that might give consistent estimates of proportionality in wartime. In this, we scholars and the public generally have not given Arkin anything remotely susceptible to algorithmic, systematic reasoning of a sort that would be programmable, even in principle.

The situation is similar with respect to the bedrock principles of noncombatant immunity, discrimination, and what international law calls "distinction." What you find enshrined in legislation is not a precise set of regulations that might be programmed for compliance, (try though we might to accomplish this objective for puzzled human combatants in our more detailed "Rules of Engagement" governing specific conflicts). Instead, in all such guidance we find a stated commitment to lofty principle, with little

concrete (let alone programmable) guidance as to who the noncombatants are in specific cases, or how reliably to distinguish non-combatants from combatants and enemy hostiles.

Thus, in a now-famous incident in the summer of 2005, four members of U.S. Special Forces "SEAL Team 10" carrying out a dangerous reconnaissance mission in Taliban-held territory south of Kandahar, encountered two very unfriendly shepherds accompanied by a 14-year old boy and their flock of 100 goats. Were they "enemy hostiles" intent on discovering and reporting the position of the team to their Taliban leader, or were they merely local residents going about their business in their own land? The debate rages to this day (Lucas 2009). The result is invariably, despite the best of intentions among the military forces engaged, imprecise and inconsistent application of both quantitative and qualitative considerations, with absolutely no guidance whatever in the law as to comparisons or ratios that are reasonable.[6]

Thus, even if computer scientists and roboticists solve the character recognition problem, and are one day reliably able with vanishing error to distinguish between an apple and a tomato, or between a young soldier and an older child, we would have to "formalize" laws of war in computer language in a manner that seems to me precluded in principle by the vagueness with which the governing conceptions are currently framed in international law. Far too much is, finally, left up to the prudent [qualitative] judgment of commanders in the field in the legal case. How do you program that? In AI, this is known as the "frame problem." That is, in order to mimic or duplicate human behavior, one has first to specify relevant operational parameters and boundary conditions for a given problem or scenario (such as ethical behavior in combat) that are programmable in

principle. My argument is, we have not done this, not through flaws in programming, or gaps in engineering, but through flaws and inherent ambiguity in LOAC itself.

Indeed, LOAC is written so as to "offload" most of the troubling questions onto the practitioners, and hold them responsible for our *post facto* review of their decision-making under stringent circumstances in the field. That is pretty shoddy legal guidance, and highly questionable from a moral point of view, but it is how things work in the human case. I don't think much of this is subject to meaningful programming, and even if so, much in the way of meaningful accountability could be attached to robot behavior under such circumstances. We have enough trouble holding human beings accountable for their actions in these cases. And this problem returns us back to Rob Sparrow's point: if we can't solve the accountability problem, even in principle, then we are prohibited under international law from developing, let alone deploying the weapons systems in question.

In addition, there appears to be an additional, conceptual problem hard-wired into Arkin's project. The problem is: how do we determine, or what do we mean by saying, that the robot complies "more closely" with LOAC than humans? We might discover empirically, for example, that the "failure rate" of humans confronting decision dilemmas involving potential noncombatants and use of deadly force is, say, 10% (or, alternatively, that our soldiers are found to comply with the requirements of LOAC in 90% or more of the altercations involving deadly force). However we chose to phrase this, it means that, in all altercations in which some dilemma over discrimination occurs, for example, the humans are not perfect. Some fail to comply with the legal requirements, and we can track or measure that benchmark. This operationalizes Arkin's proposal, by suggesting

that robots should have a lower rate of failure than humans under the same circumstances. That is, if, in a given region of conflict, humans can be shown to have a "LOAC failure rate" of 10% or less, or alternatively be said to have a positive LOAC compliance rate of 90% or better, we would demand that our robot warrior perform as well or better when compared to this benchmark.

But there is a very important conceptual confusion at work in stating the nature of the robot challenge. In the human case, we do not "program" for, or aim to attain, say, a 90% LOAC compliance rate: rather, we aim at, and require 100% compliance. That is, we expect and demand that every human warrior will comply fully with the law, *even though we realize statistically that not all will* as a matter of experience. But, when those failures are detected, they are not "tolerated" as falling within the approved margin of error! Each infraction is recorded, and its perpetrator subject to investigation and discipline. Each individual human warrior, that is, is held to a 100% compliance standard, and held strictly accountable for failing to meet it. That is the meaning of accountability.

Thus, achieving the same or better empirical rate of LOAC compliance with robots is meaningless by itself, unless it is, as in the human case, linked with a procedure to investigate and "punish" the failures. And that accountability and punishment procedure for robots would *likewise* have to be at least as effective as that in place for human warriors. Hence, even if robot warriors as a group have a better rate of compliance (or a lower failure-of-compliance rate) than their human counterparts, we cannot deliberately aim at, or deliberately decide to tolerate, less than 100% compliance. And in cases of failure of individual robot prototypes to meet this standard, we must

likewise be able to hold these individual failures accountable. Unless Arkin and computer engineers can plausibly claim 100% perfection, and/or show how to "punish" or otherwise hold failures meaningfully accountable, then it seems to me that the "Arkin test" of artificial "conscience," and the consequent claims for "machine morality," fail. Ethical robots, or as Wendell Wallach terms them, "moral machines" (Wallach 2009), thus entail a two-fold test that has not, to this point, been clearly defined. The first is the empirical component, that autonomous platforms must be shown to meet or exceed human performance under similar conditions. The second is that incidents of robot failure, just as with incidents of human failure, must be subject in principle to a meaningful procedure of review, accountability, and punishment.

So now let me return to the industrial development and production challenges. If I were engaged in industry, I would want some legal clarification and moral assurances on the foregoing range of matters before proceeding, lest I find public dissatisfaction and uncertainty in these matters serving to indict me, my engineers, my company, and its products. I would not wish to find myself, *post facto*, accused of having endangered the public recklessly and thoughtlessly, merely through accepting and fulfilling a contract to develop and manufacture the autonomous combat platforms in question. I would want to know what conception of product liability we were presupposing. Will it prove sufficient to demonstrate that my engineers and mechanics exercised "due care" during design and production by attempting to anticipate and prevent common and foreseeable misuses or malfunctions of my product? I would want to know whether I was instead going to be held strictly accountable under product liability law for what, during development and manufacture, were wholly unforeseeable, not to mention unintended, uses or abuses of

my product. I would want to know how, exactly, would we were to apportion responsibilities for mishaps among the designer, builder and producer, and operator in the field.

The concept of "due care" in industrial development and manufacture that I summarized earlier can be usefully compared to what P.W. Singer termed the "precautionary principle" in scientific research and development generally (see his conclusion at Singer 2009). A workable concept of due care, for example, likely means that we would not blindly, thoughtlessly, or unreflectively rush into, or forge ahead with, research, development, and production of lethal and/or autonomous unmanned systems, without having thoroughly explored the most likely consequences of our efforts, including a reasonable assessment of risks, to include having in turn reasonably foreseen and prevented obvious problems. Eagerness on the part of defense industries to secure lucrative government contracts ought not to blind such organizations to the potential downsides of commercial success.[7]

Here I think we can learn by analogy, tracing the similarities of robot questions with those raised regarding private military contractors (PMCs) generally, and armed private security contractors (APSCs) in particular, where liability and accountability for mistakes and accidents is (or, until recently, was) likewise unclear. In both cases, as mentioned, the need for development and deployment is driven by resource scarcity: PMCs (and now, presumably, unmanned armed autonomous systems) are thought to "do it better, faster, and cheaper" than humans, or than trained military personnel specifically. What we were slow to appreciate fully prior to the infamous Blackwater incident in Nisoor Square, Baghdad in September 2007, however, is that, for PMCs, the "corporate

profit" and customer service motives prompt very different conceptions of what seems rational, reasonable, and even responsible behavior than does the "public service" vector of military personnel under certain circumstances. Frustrated Blackwater personnel in the first instance argued in their defense that they were doing a "good job" of protecting U.S. State Department personnel from harm, which is precisely what they were hired to do. They served their customers and their paying clients well. For the military personnel on site, however, that "corporate success" was irrelevant, in light of the means employed to achieve it: namely, killing and wounding civilian by-standers indiscriminately, resorting to force too readily and roughly, and so alienating the very population whose allegiance constituted the chief objective of the counterinsurgency effort.

That illustrates how dramatically conceptions of rationality and reasonable behavior can differ, based upon motivation and perspective of the mission. What is the analogue in the robot case? Arkin rightly informs us that robots don't get angry or frightened. That emotional difference undeniably effects what is meant by rational behavior in context. How differently might robots view their combat environment in other relevant respects, when compared with humans? Would they lack compassion? Would they be capable of capturing and properly detaining surrendering enemy soldiers, or even surrender themselves when circumstances warranted? How well might they otherwise demonstrate a capacity to compromise or revise mission goals when that seemed to us to be the reasonable course of action? What would such behaviors even mean, let alone how are they programmable? We have yet to begin to reflect on this larger perspective of robot rationality beyond mere compliance with LOAC, and its implications for their effective use in combat.

I return now to the "threshold problem" with which I began this essay. Note that this not only infuses arguments about military technology, but other innovations, like the increasing resort to private military contractors. The objection in all these cases is that such measures hide the true costs of war from the public, or lower those costs for at least one side in the conflict, and so make it easier in general, or for the side utilizing these measures, to resort to war before exhausting prospects for alternative means of conflict resolution. Robots certainly do promise to lower the casualty rate for the side that employs them, and so the threshold problem might be invoked. But they also, as Arkin, Canning, and others claim, promise to lower the destructiveness for all sides, by avoiding wanton violence, negating the need for excessive "force protection," and generally for being more discriminate and proportional in their use of force and threat of harm to noncombatants than human warriors. Indeed, if we follow Canning's logic, they would be less destructive to enemy combatants, principally by aiming to disarm and capture, rather than to destroy them.

Robots in particular (unlike PMCs or other innovations that threaten to lower the threshold and hide war's true costs) might actually make war itself less destructive and costly. That would force us to re-examine the nature of the "last resort" principle itself. Is it a true (deontological) principle, like the principle of noncombatant immunity? Or is it merely the summation of a utilitarian or prudential judgment: to whit, "war is always destructive and costly, and so must never be resorted to, save as a last resort?" If the latter, then the advent of robots (alongside highly discriminate weapons, non-lethal weapons, and other technological innovations) might force us to re-examine the principle of last resort, rather than simply to condemn the emerging technology as "destabilizing"

or “war-provoking.” If war, as a result of these innovations, becomes less costly and destructive than its alternatives (including costly and harmful blockades, trade sanctions, financial embargos, or efforts at political destabilization), then it would no longer be wrong to resort to it *other* than as a “last resort.” That criterion would no longer have its traditional force. These revolutions in military technology, that is to say, may well force us to re-examine our settled principles about the presumed evils of war, at least when compared with its alternatives.

References

Anderson, Mark. "The New Robotics." *Discovery Magazine* (10 May, 2010): 37-41; http://www.discoverymagazine.com

Arkin, Ronald C. (2007). *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*, Report GIT-GVU-07-11, Atlanta, GA: Georgia Institute of Technology's GVU Center: http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf

Arkin, Ronald C. (2009). *Governing Lethal Behavior in Autonomous Robots*. Boca Raton, FL: Chapman & Hall/Taylor & Francis Group.

Arkin, Ronald C. (2010). "The Case for Ethical Autonomy in Unmanned Systems," G.R. Lucas, Jr., ed. "New Warriors and New Weapons: Ethics & Emerging Military Technologies," *Journal of Military Ethics* 9, no. 4 (December 2010).

Banda, Siva (2010). "The Challenges of Achieving UAV Autonomy." Graduate School of Engineering and Applied Sciences Distinguished Lecture Series, Naval Postgraduate School (2 December 2010). Excerpted in "Update: Naval Postgraduate School Newsletter" (January 2011), p. 5: http://www.nps.edu/About/News/World-Renowned-UAV-Control-Expert-Presents-GSEAS-Distinguished-Guest-Lecture.html [accessed 17 January 2011].

Bekey, George (2005). *Autonomous Robots: From Biological Inspiration to Implementation and Control*, Cambridge, MA: MIT Press.

Canning, John, Riggs, G.W., Holland, O. Thomas, Blakelock, Carolyn (2004). "A Concept for the Operation of Armed Autonomous Systems on the Battlefield", *Proceedings of Association for Unmanned Vehicle Systems International's (AUVSI) Unmanned Systems North America*, August 3-5, 2004, Anaheim, CA.

Canning, John (2008). "Weaponized Unmanned Systems: A Transformational Warfighting Opportunity, Government Roles in Making it Happen", *Proceedings of Engineering the Total Ship (ETS)*, September 23-25. 2008, Falls Church, VA.

Krishnan, Armin (2009). *Killer Robots: Legality and Ethicality of Autonomous Weapons*. London: Ashgate Press.

Lin, Patrick; Bekey, George; Abney, Keith (2008). *Autonomous Military Robotics: Risk, Ethics, and Design*. U.S. Department of the Navy, Office of Naval Research (20 December 2008): 112 pp.

Lucas, George R. Jr. (2009). "'This is Not Your Father's War': Confronting the Moral Challenges of 'Unconventional' War," *Journal of National Security Law and Policy*, 3, no. 2: 331-342.

Lucas, George R. Jr. (2010). "New Warriors and New Weapons: Ethics & Emerging Military Technologies," *Journal of Military Ethics* 9, no. 4 (December 2010).

OSD (2009). Office of the Secretary of Defense. FY 2009-2034 Unmanned Systems Integrated Roadmap, Second Edition. http://www.acq.osd.mil/psa/docs/UMSIntegratedRoadmap2009.pdf

Rowe, Neil C. (2007). "War Crimes from Cyberweapons," *Journal of Information Warfare*, 6: 3, 15-25.

Rowe, Neil C. (2008). "Ethics of Cyber War Attacks", in Lech J. Janczewski and Andrew M. Colarik (eds.) *Cyber Warfare and Cyber Terrorism*, Hershey, PA: Information Science Reference

Rowe, Neil C. (2009). "The Ethics of Cyberweapons in Warfare," *International Journal of Cyberethics*, 1: 1.
Sharkey, Noel (2007a). "Robot Wars are a Reality", *The Guardian* (UK), August 18, 2007, p. 29. http://www.guardian.co.uk/commentisfree/2007/aug/18/comment.military
Sharkey, Noel (2007b). "Automated Killers and the Computing Profession", *Computer* 40: 122-124.
Sharkey, Noel (2008a). "Cassandra or False Prophet of Doom: AI Robots and War", *IEEE Intelligent Systems*, July/August 2008, pp. 14-17.
Sharkey, Noel (2008b). "Grounds for Discrimination: Autonomous Robot Weapons", *RUSI Defence Systems*, 11:2, 86-89.
Sharkey, Noel (2010). "Just Say 'No!' to Autonomous Lethal Systems," in G.R. Lucas, Jr., ed. "New Warriors and New Weapons: Ethics & Emerging Military Technologies," *Journal of Military Ethics* 9, no. 4 (December 2010).
Singer, P. W. (2009). *Wired for War*. New York: Penguin Press.
Singer, P.W. (2010). "The Ethics of 'Killer Apps'," in G.R. Lucas, Jr., ed. "New Warriors and New Weapons: Ethics & Emerging Military Technologies," *Journal of Military Ethics* 9, no. 4 (December 2010).
Sparrow, Robert (2007). "Killer Robots", *Journal of Applied Philosophy*, 24: 1, 62-77.
Sparrow, Robert (2008). "Building a Better WarBot: Ethical Issues in the Design of Unmanned Systems for Military Applications," *Science and Engineering Ethics* 15, no. 2: 169-187.
Sparrow, Robert (2009). "Predators or Plowshares? Arms Control of Robotic Weapons," *IEEE Technology and Society Magazine*, 28, no. 1: 25-29.
Wallach, Wendell (2009). *Moral Machines: Teaching Robots Right from Wrong*. New York: Oxford University Press.

---

[1] This paper was originally delivered at a conference on robotics and ethics sponsored by the French Military Academy at the *Ecole Militaire*, Paris (17 June 2010), and at the annual meeting of the International Society for Military Ethics (ISME) in San Diego, CA: 26 January 2011.

[2] Most recently, the "threshold" and last resort" problem is cited as a chief objection to further development of military robotics by a number of leading authorities in John Markoff, "War Machines: Recruiting Robots for Combat," *New York Times* "Science" section (28 November 2010): http://www.nytimes.com/2010/11/28/science/28robot.html?emc=eta1 [accessed 29 November, 2010].

[3] I use the catch-phrase "better, faster, cheaper" advisedly, as this phrase, originally coined by the National Aeronautics and Space Administration for the U.S. manned-space program, is now the motto of private military contractors, whose resultant use has also prompted some moral and legal soul-searching.

[4] An account of the incident and subsequent investigation can be found in *Der Spiegel* (12 December 2009): http://www.spiegel.de/international/world/0,1518,667476,00.html. For a discussion of proportionality in the incident, see Susanne Koebl, "The Dilemma of the Kunduz Bombing: How Much is a Human Life Worth?" *Der Spiegel* (15 December 2009): http://www.spiegel.de/international/world/0,1518,667123,00.html

[5] A disputed analysis, along with footage from the video itself, can be found in an account of the incident for the *Manchester Guardian* by Washington correspondent, Chris McGreal (5 April 2010): http://www.guardian.co.uk/world/2010/apr/05/wikileaks-us-army-iraq-attack [accessed 17 January 2011]. For many in the wider public, this was their first encounter with WikiLeaks. Correspondent McGreal's report of the incident has since been much debated and contested in news analysis and internet blogs.

[6] Interestingly, in subsequent discussion with me about this case, Arkin suggested that it might instead serve as an illustration of the potential superiority of robot warriors, who might have been nearly as able to carry out the required reconnaissance, without their accidental discovery or disclosure having triggered anything like the panicked, defensive response of the humans, that might have tempted the latter to commit war crimes.

[7] A useful analogy of how badly wrong commercial success in securing contracts for risky ventures can go, is British Petroleum and the Gulf of Mexico oil disaster in the U.S. Given how thoroughly uncertain and risky deep-water drilling was thought to be, most persons now agree that reasonable precautions should have been taken, based upon realistic scenarios of error. The company's former CEO, Tony Hayward, confessed in the midst of the unfolding calamity that his company simply failed to consider what they might do in what he termed this "high risk, low probability" scenario. It would likewise be difficult for us to say that we have done other than likewise thus far in the case of autonomous military robots. That failure on BP's part proved extremely expensive, and did extraordinary damage to the corporation itself. Such grave outcomes ought to serve as an object lesson in the "precautionary principle," and inform the judgment of any corporate ventures into autonomous unmanned systems as well.