



# **Autonomous Military Robotics: Risk, Ethics, and Design**

Prepared for: US Department of Navy, Office of Naval Research

Prepared by: Patrick Lin, Ph.D.  
George Bekey, Ph.D.  
Keith Abney, M.A.

Ethics + Emerging Sciences Group at  
California Polytechnic State University, San Luis Obispo

Prepared on: December 20, 2008

Version: 1.0.9

*This work is sponsored by the Department of the Navy, Office of Naval Research,  
under awards # N00014-07-1-1152 and N00014-08-1-1209.*

## 7. Robot Ethics: The Issues

---

From the preceding sections, it should be clear that there are myriad issues in risk and ethics related to autonomous military robotics. In this section, we will pull together and organize these various strands, as well as raise additional ones to provide a single, full view of the challenges facing the field.<sup>6</sup> These challenges are organized in thematic sub-groups: legal, just war, technical, robot-human, societal, and other and future challenges.

This is not meant to be an exhaustive list, as other issues certainly will emerge as the technology develops and field use broadens.<sup>7</sup> The value of this section again is to help anticipate the challenges facing the development and deployment of autonomous military robots, in order to proactively address them in both the design or application phases. Further, they may help to inform ethical and risk issues related to non-military robots, given the close historical relationship between defense technologies and consumer or public technologies, such as the evolution of ARPANET into the Internet.

### 7.1 Legal Challenges

---

1. *Unclear responsibility.* To whom would we assign blame—and punishment—for improper conduct and unauthorized harms caused by an autonomous robot (whether by error or intentional): the designers, robot manufacturer, procurement officer, robot controller/supervisor, field commander, President of the United States...or the robot itself? [Asaro, 2007; Sparrow, 2007; Sharkey, 2008a] We have started an inquiry into this critical issue in section 5: The law offers several precedents that a robotics case might follow, but given the range of specific circumstances that would influence a legal decision as well as evolving technology, more work will be needed to clarify the law for a clear framework in matters of responsibility.

---

<sup>6</sup> We thank and credit Ron Arkin for his discussions on many of these issues presented here.

<sup>7</sup> As an example of an unexpected policy change, when German forces during World War II recognized the impracticality of using naval submarines to rescue crews of sinking enemy ships—given limited space inside the submarine as well as exposure to radar and attacks when they surface—they issued the *Laconia* Order in 1942, based on military necessity, that released submarines from a long-standing moral obligation for sea vessels to rescue survivors; other nations soon followed suit to effectively eliminate the military convention altogether [Walzer, 1977, pp. 147-151].

In a military system, it may be possible to simply *stipulate* a chain of responsibility, e.g., the commanding officer is ultimately responsible. But this may oversimplify matters, e.g., inadequate testing allowed a design problem to slip by and caused the improper robotic behavior, in which case perhaps a procurement officer or the manufacturer ought to be responsible. The situation becomes much more complex and interesting with robots that have greater degrees of autonomy, which may make it appropriate to treat them as quasi-persons, if not full moral agents some point in the future. We note that Kurzweil forecasts that, by the year 2029, “[m]achines will claim to be conscious and these claims will be largely accepted” [Kurzweil, 1999].

2. *Refusing an order.* A conflict may arise in the following situation, among others: A commander orders a robot to attack a house that is known to harbor insurgents, but the robot—being equipped with sensors to ‘see’ through walls—detects many children inside and, given its programmed instruction (based on the ROE) to minimize civilian casualties, refuses the order. How ought the situation proceed: should we defer to the robot who may have better situational awareness, or the officer who (as far as she or he knows) issues a legitimate command? This dilemma also relates back to the question of responsibility: if the robot refuses an order, then who would be responsible for the events that ensue? Following legitimate orders is clearly an essential tenet for military organizations to function, but if we permit robots to refuse an order, this may expand the circumstances in which human soldiers may refuse an order as well (for better or worse).
3. *Consent by soldiers to risks.* In October 2007, a semi-autonomous robotic cannon deployed by the South African army malfunctioned, killing nine ‘friendly’ soldiers and wounding 14 others [Shachtman, 2007]. It would be naive to think such accidents will not happen again. In these cases, should soldiers be informed that an unusual or new risk exists, e.g., when they are handling or working with other dangerous items, such as explosives or even anthrax? Does consent to risk matter anyway, if soldiers generally lack the right to refuse a work order? We discussed the notion of consent and informed in the previous section.

## 7.2 Just-War Challenges

---

1. *Attack decisions.* It may be important for the above issue of responsibility to decide who, or what, makes the decision for a robot to strike. Some situations may develop so quickly and require such rapid information processing that we would want to entrust our robots and systems to make critical decisions. But the LOW and ROE generally demand there to be human ‘eyes on target’, either in-person or electronically and presumably in real time. (This is another reason why there is a general ban on landmines: without eyes on target, we do not know who is harmed

by the ordnance and therefore have not fulfilled our responsibility to discriminate combatants from non-combatants.) If human soldiers must monitor the actions of each robot as they occur, this may limit the effectiveness for which the robot was designed in the first place: robots may be deployed precisely because they can act more quickly, and with better information, than humans can.

However, some military robots—such as the Navy’s Phalanx CIWS—seem to already and completely operate autonomously, i.e., they make attack decisions without human eyes on target or approval. This raises the question of how strictly we should take the ‘eyes on target’ requirement. One plausible argument for stretching that requirement is that the Phalanx CIWS operates as a last line of defense against imminent threats, e.g., incoming missiles in the dark of the night, so the benefits more clearly outweigh the risks in such a case. Another argument perhaps would be that ‘eyes on target’ need not be *human* eyes, whether directly or monitoring the images captured by a remote camera; that is, a human does not necessarily need to directly confirm a target or authorize a strike. A robot’s target-identification module—assuming it has been sufficiently tested for accuracy—programmed by engineers is arguably a proxy for human eyes. At least this gives the system some reasonable ability to discriminate among targets, in contrast to a landmine, for instance. A requirement for 100% accuracy in target identification may be overly burdensome, since that is not a bar we can meet with human soldiers.

2. *Lower barriers for war.* As raised in section 4, does the use of advanced weaponry such as autonomous robotics make it easier for one nation to engage in war or adopt aggressive foreign (and domestic) policies that might provoke other nations? If so, is this a violation of *jus ad bellum*? [Asaro, 2008; Kahn, 2002] It may be true that new strategies, tactics, and technologies make armed conflict an easier path to choose for a nation, if they reduce risks to our side. Yet while it seems obvious that we should want to reduce US casualties, there is something sensible about the need for some terrible cost to war as a deterrent against entering war in the first place. This is the basis for just-war theory, that war ought to be the very last resort given its horrific costs [Walzer, 1977].

But the considered objection—that advanced robotics immorally lowers barriers for war—hides a logical implication that we should not do anything that makes armed conflict more palatable: we should not attempt to reduce friendly casualties, or improve battlefield medicine, or conduct any more research that would make victory more likely and quicker. Taken to the extreme, the objection seems to imply that we should *raise* barriers to war, to make fighting as brutal as possible (e.g., using primitive weapons without armor) so that we would never engage in it unless it were truly the last resort. Such a position appears counterintuitive at best and dangerously foolish at worst, particularly if we expect that other nations would not readily adopt a policy of relinquishment, which would put the US at a competitive disadvantage.

3. *Imprecision in LOW and ROE.* Asimov's Laws appear to be as simple as programmable rules can be for autonomous robots, yet they yielded surprising, unintended implications in his stories [e.g., Asimov, 1950]. Likewise, we may understand each rule of engagement and believe them to be sensible, but are they truly consistent with one another and sufficiently clear—which appears to be a requirement in order for them to be programmable? Much more complex than Asimov's Laws, the LOW and ROE leave much room for contradictory or vague imperatives, which may result in undesired and unexpected behavior in robots.

For instance, the ROE to minimize collateral damage is vague: is the rule that we should not attack a position if civilian deaths are expected to be greater than—or even half of—combatant deaths? Are we permitted to kill one (high-ranking) combatant, even if it involves the death of five civilians—or \$10M in unnecessary damage? A robot may need specific numbers to know exactly where this line is drawn, in order to comply with the ROE. Unfortunately, this is not an area that has been precisely quantified nor easily lends itself for such a determination.

### 7.3 Technical Challenges

---

1. *Discriminating among targets.* Some experts contend that it is simply too difficult to design a machine that can distinguish between a combatant and a non-combatant, particularly as insurgents pose as civilians, as required for the LOW and ROE [e.g., Sharkey, 2008a; Sparrow, 2007; Canning et al., 2004]. Further, robots would need to discriminate between active combatants and wounded ones who are unable to fight or have surrendered. Admittedly, this is a complex technical task, but we need to be clear on how accurate this discrimination needs to be. That is, discrimination among targets is also a difficult, error-prone task for human soldiers, so ought we hold machines to a higher standard than we have yet to achieve ourselves, at least in the near term?

Consider the following: A robot enters a building known to harbor terrorists, but at the same time an innocent girl is running toward the robot (unintentionally) in chasing after a ball that happens to be rolling in the direction of the robot. Would the robot know to stand down and not attack the child? If the robot were to attack, of course that would cause outrage from opposing forces and even our own media and public; but this scenario could likely be the same as with a human soldier, adrenaline running high, who may misidentify the charging target as well. It seems that in such a situation, a robot may be less likely to attack the child, since the robot is not prone to overreact from the influence of emotions and fear, which afflict human soldiers. But in any event, if a robot would likely not perform worse than a human soldier, perhaps this is good enough for the moment until the technical ability to discriminate among

targets improves. Some critics, however, may still insist on perfect discrimination or at least far better than humans are capable of, though it is unclear why we should hold robots to such a high standard before such a technology exists (unless their point is to not use robots at all until we have perfected them, which is also a contentious position).

One proposed ‘workaround’ solution is to permit robots to target only weapons, including any hostile robots, rather than the human soldiers wielding such weapons [Canning, 2008]. Thus if an enemy combatant fails to relinquish his weapon in the presence of a robot, then he significantly increases his risk of being unintentionally harmed as the robot proceeds to disable the weapon. However, while this seems reasonable in principle, other experts continue to point to the technical challenge of discrimination: given current and foreseeable limitations in AI, a robot still may not be able to reliably target only a weapon and not the person, nor even reliably identify weapons from non-weapons, e.g., a child pointing her ice cream cone at an urban patrol robot [Sharkey, 2007b]. If this is true, then the considered solution merely postpones the discrimination problem, though it does create an extra layer of protection against inappropriate harm to humans; so the solution merits further consideration.

Another possible solution, which avoids the above programming issues, may be to simply operate combat robots only in regions of heavy fighting, teeming with valid targets [Sharkey, 2008b]. In these zones—sometimes called ‘kill boxes’ or ‘engagement regions’—the Rules of Engagement are loosened, and non-combatants can be reasonably presumed to have fled, thus obviating the issue of discriminating among targets (and assuming none of our own troops is in the kill box or at least can be easily identified, e.g., by wearing some sensor). Using combat robots, at least initially, in only such zones might help to solve the first-generation problem described below, providing a training ground of sorts to test and perfect the machines. Or even in regions without heavy fighting but in need of tight security, e.g., guarding perimeters, armed sentry robots could operate in those designated zones, as long as it is clear to everyone that trespassers will be presumed to be enemy combatants and sufficient safeguards or deterrents to entry are in place to prevent, say, an accidental trespass by a child. (The risk of harm to a non-combatant here seems to be the same as with using guard dogs to protect property today.)

2. *First-generation problem.* We previously mentioned that it would be naive to believe that another accident with military robots will not happen again. As with any other technologies, errors or bugs will inevitably exist, which can be corrected in the next generation of the technology. With Internet technologies, for instance, first-generation mistakes are not too serious and can be fixed with software patches or updates. But with military robotics, the stakes are much higher, since human lives may be lost as a result of programming or other errors. So it seems that the prudent or morally correct course of action is to rigorously test the robot before deploying it, as discussed in section 6.

However, testing already occurs with today's robots, yet it is still difficult if not impossible to certify any given robot as error-free, given that (a) testing environments may be substantially different than more complex, unstructured, and dynamic battlefield conditions in which we cannot anticipate all possible contingencies; and (b) the computer program used in the robot's on-board computer (its 'brain') may consist of millions of lines of code.

Beta-testing of a program (testing prior to the official product launch, whether related to robotics, business applications, etc.) is conducted today, yet new errors are routinely found in software by actual users even after its official product launch. It is simply not possible to run a complex piece of software through all possible uses in a testing phase; surprises may occur during its actual use. Likewise, it is not reasonable to expect that testing of robots will catch any and all flaws; the robots may behave in unexpected and unintended ways during actual field use. Again, the stakes are high with deploying robots, since any error could be fatal. This makes the first-generation problem, as well as ongoing safety and dependability, an especially sensitive issue [e.g., Van der Loos, 2007].

3. *Robots running amok.* As depicted in science-fiction novels and movies, some imagine the possibility that robots might break free from their human programming through methods as: their own learning, or creating other robots without such constraints (self-replicating and self-revising), or malfunction, or programming error, or even intentional hacking [e.g., Joy, 2000]. In these scenarios, because robots are built to be durable and even with attack capabilities, they would be extremely difficult to defeat—which is the point of using robots as force multipliers. Some of these scenarios are more likely than others: we wouldn't see the ability of robots to fully manufacture other robots or to radically evolve their intelligence and escape any programmed morality for quite some time. But other scenarios, such as hacking, seem to be near-term possibilities, especially if robots are not given strong self-defense capabilities (see below).

That robots might run amok is an enhanced version of the worry that enemies might use our own creations against us, but it also introduces a new element in that previous weapon systems still need a human operator which is a point of vulnerability, i.e., a 'soft underbelly' of the system. Autonomous robots would be designed to operate without human control. What precautions can be taken to prevent one from being captured and reverse-engineered or reprogrammed to attack our own forces? If we design a 'kill switch' that can automatically shut off a robot, this may present a key vulnerability that can be exploited by the enemy.

4. *Unauthorized overrides.* This concern is similar to that with nuclear weapons: that a rogue officer may be enough to take control of these terrible weapons and unleash them without

authorization or otherwise override their programming to commit some unlawful action. This is a persistent worry with any new, devastating technology and is a multi-faceted challenge: it is a human problem (to develop ethical, competent officers), an organizational problem (to provide procedural safeguards), and technical problem (to provide systemic safeguards). So there does not yet appear to be anything unique about this worry that should hinder the development or deployment of advanced robotics, to the extent that the concern does not impact the development of other technologies. But it nevertheless is a concern that needs to be considered in the design and deployment phases.

5. *Competing ethical frameworks.* If we seek to build an ethical framework for action in robots, it is not clear which ethical theory to use as our model [e.g., Anderson and Anderson, 2007]. In section 3, we have argued for a hybrid approach related to virtue ethics, as the theory that seems to lead to the fewest unintuitive results, but any sophisticated theory seems to be vulnerable to inconsistencies and competing directives (especially if a three- or four-rule system as simple as Asimov's cannot work perfectly). This concern is related to the first technical challenge described here, that it is too difficult to embed these behavioral rules or programming into a machine. But we should recall our stated mission here: our initial goal ought *not* be to create a perfectly ethical robot, only one that acts more ethically than humans—and sadly this may be a low hurdle to clear.
  
6. *Coordinated attacks.* Generally, it is better to have more data than less when making decisions, particularly one as weighty as a military strike decision. Robots can be designed to easily network with other robots and systems; but this may complicate matters for robot engineers as well as commanders. We may need to establish a chain of command within robots when they operate as a team, as well as ensure coordination of their actions. The risk here is that as complexity of any system increases, the more opportunities exist for errors to be introduced, and again mistakes by military robots may be fatal.

#### 7.4 Human-Robot Challenges

---

1. *Effect on squad cohesion.* As a 'band of brothers', there understandably needs to be strong trust and support among soldiers, just as there is among police officers, firefighters, and so on. But sometimes this sense of camaraderie can be overdeveloped to the extent that one team member becomes complicit in or deliberately assists in covering up an illegal or inappropriate action of another team member. We have discussed the benefits of military robots with respect to behavior that is more ethical than currently exhibited by human soldiers. But robots will also likely be equipped with video cameras and other such sensors to record and report actions on the battlefield. This could negatively impact the cohesion among team or squad members by



eroding trust with the robot as well as among fellow soldiers who then may or may not support each other as much anymore, knowing that they are being watched. Of course, soldiers and other professionals should not be giving each other unlawful ‘support’ anyway; but there may be situations in which a soldier is unclear about or unaware of motivations, orders, or other relevant details and err on the side of caution, i.e., not providing support even when it is justified and needed.

2. *Self-defense.* Asimov’s Laws permitted robots to defend themselves where that action did not conflict with higher duties, i.e., harm humans (or humanity) or conflict with a human-issued order. But Arkin suggests that military robots can be more conservative in their actions, i.e., hold their fire, because they do not have a natural instinct of self-preservation and may be programmed without such [Arkin, 2007]. But how practical is it, at least economically speaking, to not give robots—which may range from \$100,000 to millions of dollars in cost—the ability to defend themselves? If a person, say, a US civilian, threatens to destroy a robot, shouldn’t it have the ability to protect itself, our very expensive taxpayer-funded investment?

Further, self-defense capabilities may be important for the robot to elude capture and hacking, as previously discussed. Robots may be easily trapped and recovered fully intact, unlike tanks and aircraft, for instance, which usually sustain much if not total damage in order to capture it. These considerations are in tension with using robots for a more ethical prosecution of war, since a predilection to hold their fire would be a major safeguard against accidental fatalities, e.g., mistakenly opening fire on non-combatants; therefore, a tradeoff or compromise among these goals—to have a more ethical robot and to protect the robot from damage and capture—may be needed.

3. *Winning hearts and minds.* Just-war theory, specifically *jus post bellum*, requires that we fight a war in such a manner that it leaves the door open for lasting peace after the conflict [Orend, 2002]. That is, as history has shown, we should not brutalize an enemy, since that would leave ill-feelings to linger even after the fighting has stopped, which makes peaceful reconciliation most difficult to achieve. Robots do not necessarily represent an immoral or overly brutal way of waging war, but as they are needed for urban operations, such as patrolling dangerous streets to enforcing a curfew or securing an area, the local population may be less likely to trust and build good-will relationships with the occupying force [Sharkey, 2008a]. Winning hearts and minds is likely to require diplomacy and human relationships that machines would not be capable of delivering at the present time, as we previously discussed in section 4.
4. *‘Comfort’ robots.* Ethicists are already talking about the impact of robots as lovers or surrogate relationship partners [Levy, 2007]. This does not seem so unthinkable, considering that some people already have ‘relationships’ with increasingly-realistic sex dolls, so robotics appear to be

a natural next step in that industry; indeed, people today engage in sexual activities online, i.e., without a partner physically present.

In previous wars, women have been taken by the military to provide ‘comfort’ to soldiers or, in other words, forced into sexual slavery or prostitution. In World War II, women were most infamously used by the Japanese Imperial Army to satiate the pent-up carnal desires of its soldiers, ostensibly to prevent possible riots and discontent among the ranks; Nazi Germany reportedly also used women to stock their ‘joy divisions’ at labor or concentration camps. And instances of rape have been reported—and continue today—in armed conflicts from Africa to the Americas to Asia.

Robots, then, may be able to serve the same function of providing ‘comfort’ to the troops in a much more humane way, i.e., without the exploitation of women and prisoners of war. However, it is unclear that this function is truly needed (to the extent that most militaries today do not employ military prostitutes and seem to be operating adequately) or can overcome existing public inhibitions or attitudes on what is mostly a taboo subject of both sex in the military and sex with non-human objects.

## 7.5 Societal Challenges

---

1. *Counter-tactics in asymmetric war.* As discussed in the previous issue of lowering barriers to war or making war more risk-free, robots would help make US military actions more effective and efficient, which is exactly the point of deploying those machines. Presumably, the more autonomous a robot is, the more lethal it can be (given requirements to discriminate among targets and so on). This translates to quicker, more decisive victories for us; but for the other side, this means swifter and perhaps more demoralizing defeats. We can reasonably expect that a consequence of increasing the asymmetry of warfare in our favor will cause opposing forces to engage in even more unconventional strategies and tactics, beyond ‘terrorist’ acts today as necessitated by an overwhelming superiority of US troop numbers and technologies [e.g., Kahn, 2002]; few nations could hope to successfully wage war with the US by using the same methods we use.

This not only involves how wars and conflicts are fought, but also exposes our military as well as public to new forms of attack which may radically change our society, as the events of 9/11 have already. For instance, more desperate enemies may resort to more desperate measures, from intensifying efforts to acquire nuclear or biochemical weapons to devising a ‘scorched earth’ or ‘poison pill’ strategy that strikes deeply at us but at some great cost to their own forces or population (a Pyrrhic victory).

2. *Proliferation.* Related to the previous issue, history also shows that innovations in military technologies—from armor and crossbows to intercontinental missiles and ‘smart’ bombs—give the inventing side a temporary advantage that is eroded over time by other nations working to replicate the technologies. Granting that modern technologies are more difficult to reverse-engineer or replicate than previous ones, it nevertheless seems inevitable or at least possible that they can be duplicated, especially if an intact sample can be captured, such as immobilizing a ground robot as opposed to shooting down an unmanned aerial vehicle. So with the development of autonomous military robots, we can expect their proliferation with other nations at some future point. This means that these robots—which we are currently touting as lethal, difficult-to-neutralize machines—may be turned against our own forces eventually.

The proliferation of weapons, unfortunately, is an extremely difficult cycle to break: many nations are working to develop autonomous robotics, so a unilateral ban on their development would not accomplish much except to handicap that nation relative to the rest of the world. So the rush to develop this and other emerging technologies is understandable and irresistible, at least in today’s world. One possible defense for our pursuit, apart from self-interested reasons, is that we (the US) want to ensure we develop these commanding technologies first, after which we would have more leverage to stop the proliferation of the same; further, because we occupy the higher moral ground, it would be most responsible for the US to develop the technologies first.

The problem, of course, is that every nation thinks of itself as moral or ‘doing the right thing’, so it would be difficult to objectively assign a moral imperative to any given nation, including the US. Solving this problem, then, would seem to require additional legal and ethical theorizing, likely resulting in new international treaties and amendments to the Laws of War.

3. *Space race.* As on earth, autonomous robots may hold many benefits for space exploration [Jónsson et al., 2007]. Proliferation also has significant financial and environmental costs, particularly if military robotics technology is developed for outer space. First, launch costs are still astronomical, costing thousands of dollars *per pound* to put an object into low Earth orbit, and several times more per pound for geostationary orbit (not to mention periodic replacement costs and in-orbit repairs). An unlikely ‘star wars’ scenario aside—which would create countless pieces of space debris that would need to be tracked and threaten communications satellites and so on—even using robots for research purposes, e.g., to explore and develop moons and other planets, may spark another space race given the military advantages of securing the ultimate high ground. This not only opens up outer space for militarization, which the world’s nations have largely resisted, but diverts limited resources that could make more valuable contributions elsewhere.

4. *Technology dependency.* The possibility that we might become dependent or addicted to our technologies has been raised throughout the history of technology and even with respect to robotics. Today, ethicists worry that we may become so reliant on, for instance, robots for difficult surgery that humans will start losing that life-saving skill and knowledge; or that we become so reliant on robots for basic, arduous labor that our economy is somehow impacted and we forget some of those techniques [Veruggio, 2007]. In the military, some soldiers already report being attached to the robot that saved their lives [Garreau, 2007].

As a general objection to technology, this concern does not seem to have much force, since the benefits of the technology in question often outweigh any losses. For instance, our ability to perform mathematical calculations may have suffered somewhat given the inventions of the calculator and spreadsheets, but we would rather keep those tools even at that expense. Certainly, it is a possible hypothetical or future scenario that, after relying on robots to perform all our critical surgeries, some event—say, a terrorist attack or massive electromagnetic pulse—could interrupt an area’s power supply, disabling our machines and leaving no one to perform the surgery (because we forgot how and have not trained surgeons on those procedures, since robots were able to do it better). But as abilities enhanced by technology, such as performing numeric calculations, have not entirely disappeared from a population or even to a life-impacting degree in individuals, it is unclear why we would expect something as artful as brain or heart surgery to be largely lost. Similarly, in the case of relying on robots for manual labor, technology dependency would not erase our ability to, say, dig holes to plant trees to any impacting degree.

5. *Civil security and privacy.* Defense technologies often turn into public or consumer technologies, as we previously pointed out. So a natural step in the evolution of military robots would seem to be their incarnation as civil security robots; they might guard corporate buildings, control crowds, and even chase down criminals. Many of the same concerns discussed above—such as technical challenges and questions of responsibility—would also become larger societal concerns: if a robot unintentionally (meaning that no human intentionally programmed it to ever do so) kills a small child, whether by accident (run over) or mistake (identification error), it will likely have greater repercussions than a robot that unintentionally kills a non-combatant in some faraway conflict. Therefore, there is increased urgency to address these military issues that may spill over into the public domain.

And while we take it that soldiers, as government property, have significantly decreased privacy expectations and rights, the same is not true of the public at large. If and when robots are used more in society, and the robots are likely to be networked, concerns about illegal monitoring and surveillance—privacy violations—may again surface, as they have with most other modern

technologies, from DNA testing to genome sequencing to communications-monitoring software to nanotechnology. This raises the question of what kind of consent we need from the public before deploying these technologies in society.

## 7.6 Other and Future Challenges

---

1. *Co-opting of ethics effort by military for justification.* A possible challenge that does not fit neatly into any of the above categories is the following political concern. Defense organizations may be aware (now) of the above concerns, but they may still not choose to address the issues to mitigate risk by absolving themselves of this responsibility: they may simply point to ethicists and robot scientists working on related issues as justification for proceeding ahead without any real plan to address at least some of these risks [Sharkey, 2007b].

This is an interesting *meta*-issue for robot ethics, i.e., it is about the study and aims of robot ethics and not so much about an issue directly related to the use of autonomous robots. While it is certainly a possibility that organizations may only pay ‘lip-service’ to the project of robot ethics to appease critics and watchdogs, it does not take much enlightenment or foresight to see actual, real-world benefits from earnestly addressing these challenges. Further, we might measure the commitment that organizations have to robot ethics by the funding levels for such research. And it would be readily apparent if, for instance, defense organizations ignored the counsel and recommendations of experts engaged in the field. This is to say that co-opting is a relatively transparent activity to identify, although the point is more that it could be too late (for those harmed or society in general) by then.

2. *Robot rights.* For now, robots are seen as merely a tool that humans use, morally no different (except in financial value) than a hammer or a rifle—their only value is instrumental, as a means to our ends. But as robots begin to assume aspects of human decision-making capabilities, the question may arise of their *intrinsic value*: do they deserve moral consideration of their own (beyond their financial or tactical value), and at what point in their evolution will they achieve this intrinsic value (as human lives seem to have)? When they become Kantian autonomous agents, making their own goals for themselves? Or would intrinsic value also require consciousness and emotions?

Some technologists have suggested that, by 2029, robots will demand equal treatment before the law with humans—and believe that this demand will be granted [e.g., Kurzweil, 1999]. The only guarantee of avoiding this outcome appears to be a prohibition on programming robots with anything other than a ‘slave morality’, i.e., simply not allowing a Kantian-autonomous robot to ever be programmed or built (though such bans, especially when applied internationally, have

been notoriously difficult to enforce). It will require careful consideration in the future as to whether such a prohibition should ever be lifted. Fortunately, even ‘technological optimists’, such as Kurzweil, do not expect this to be an issue until at least the 2020s.

Thus far, we have not discussed the possibility of giving rights to robots, not so much that it is farfetched to do so (e.g., we give rights to non-living entities such as corporations) or to consider them as persons (philosophically-speaking; e.g., again corporations or ships or some animals such as dolphins), but that the prerequisites for rights seem to require advanced software or artificial intelligence that is not quite within our foreseeable grasp. Specifically, if our notion of personhood specifies that only persons can be afforded rights and that persons must have free will or the capacity for free will, then it is unclear whether we will ever develop technologies capable of giving free will or full autonomy to machines, and, indeed, we don’t even know whether any other *biological* species will ever have or is now capable of such full autonomy; thus, we do not want to dwell on such a speculative issue here. That said, we will leave open the possibility that we may someday want or be logically required to give rights to robots [e.g., Kurzweil, 1999], but much more investigation is needed on the issue.

3. *The precautionary principle.* Given the above laundry list of concerns, some may advocate following a precautionary principle in robotics research—to slow or halt work until we have mitigated or addressed possible catastrophic risks—as critics have done for other technologies, such as bio- and nanotechnologies. For instance, those fearful of ‘Terminator’ scenarios where machines turn against us lesser humans, current research in autonomous robotics may represent a path towards possible, perhaps likely, disaster; thus a cautious, prudent approach would be to ban or at least significantly slow down research until we can sufficiently think about these issues before technology overtakes ethics. While we believe that a precautionary principle may be the appropriate course of action for some technology cases, many of the issues discussed above do not appear imminent enough to warrant a research moratorium or delay, just more investigation which may be sufficiently conducted in parallel to efforts to develop advanced robotics.

Furthermore, a cautionary approach in the development of advanced systems is inherently in tension with both the approaches taken by the scientists and engineers developing robots and with the outlook of military planners, rapidly searching for more effective tools for the task of waging war. We will again leave open the possibility that someday we may have to seriously consider the role of the precautionary principle in robotics, but that day appears to be in the distant horizon and does not demand an extensive discussion here.

## 7.7 Further and Related Investigations Needed

---

Again, we do not intend the above to capture all possible issues related to autonomous military robotics. Certainly, new issues will emerge depending on how the technology and intended uses develop. In preceding sections, we have started to address what we seemed to be the most urgent and important issues to resolve first, especially as related to responsibility, risk, and the ability of robots to discriminate among targets. This is only the beginning of a dialogue in robot ethics and merits further investigations.

Moreover, our discussion here may be helpful in informing ethics research related to non-military robots, such as security, labor, and sex robots previously mentioned in the public domain. For instance, robots are already being used to care for the elderly, but are we merely pawning off our obligations to care for the elderly to machines who may be unable to provide the emotional content that seems to be needed in human relationships (even though there are significant advances in enabling robots to display ‘emotions’)? What are the benefits and risks of using robots as teachers, domestic help, or even as researchers, e.g., exploring difficult and alien environments? Do robotic planes, trains, and automobiles pose any special issues? These and other questions will need to be addressed; and as is often the case with public technologies, they have their roots in military innovations, so we may have a separate but related responsibility to begin looking ahead to these non-military questions as well.