Ethica Themen

Institut für Religion und Frieden

Gerhard Dabringer (Hg.)

Ethical and Legal Aspects of Unmanned Systems
Interviews

Institut für Religion und Frieden

http://www.irf.ac.at

IMPRESSUM
Amtliche Publikation der Republik Österreich/ Bundesminister für Landesverteidigung und Sport
MEDIENINHABER, HERAUSGEBER UND HERSTELLER:
Republik Österreich/ Bundesminister für Landesverteidigung und Sport, BMLVS, Roßauer Lände 1, 1090 Wien
REDAKTION:
BMLVS, Institut für Religion und Frieden,
Fasangartengasse 101, Objekt VII, 1130 Wien, Tel.: +43/1/512 32 57,
Email: irf@mildioz.at
ERSCHEINUNGSJAHR:
2010
DRUCK:
BMLVS, Heeresdruckerei, Kaserne Arsenal, Objekt 12, Kelsenstraße 4, 1030 Wien
ISBN: 978-3-902761-04-0

Ethica Themen

Institut für Religion und Frieden

Gerhard Dabringer (Hg.)

Ethical and Legal Aspects of Unmanned Systems
Interviews

Institut für Religion und Frieden

http://www.irf.ac.at

TOC

John Canning, Gerhard Dabringer: Ethical Challenges of Unmanned Systems

Introduction

The word “robot” has been in public use since the Czech writer Karel Čapek introduced it in his play R.U.R. (Rossum’s Universal Robots), published in 1920[1]. Karel claims that his brother, Josef Čapek , actually coined the word, stemming from the Czech word “robota” refering to work, labor or serf labor, and figuratively "drudgery" or "hard work.”[2] In the play, these were creatures that could be mistaken for humans, and seemed happy to serve. The issue in Karel’s play was whether the robots were being exploited. Thus was born, not only the term “robot,” but also the first ethical question involving them. It should come as no surprise, then, that questions involving the ethics of using robots have not gone away.

For many years the public’s frame of reference for robotic ethics were taken from Isaac Asimov’s Three Laws of Robotics, which he penned in 1942 in his science fiction short story “Runaround.[3]” (Asimov later added the less well-known Zeroth Law to this collection as well.[4]) But this was all from science fiction, since there were no real robots, and thus no real robotic ethics. Today, we stand on the threshold of the emergence of real robots, although not as Karel Čapek first envisioned them. So it is time to consider the real ethical (and legal) issues that come with them.

The Spread of Robotics

Today, we see the widespread commercial sale and use of such products as the iRobot Roomba and Scooba carpet and floor cleaners[5], with other products coming, but more importantly to our discussions in the military arena, we have such items as the HELLFIRE missile-armed Predator and

[1] An English translation of the book under the Creative Commons Licence is available: http://ebooks.adelaide.edu.au/c/capek/karel/rur/complete.html.
[2] Lidové Noviny, 24.12.1933, translation at: http://capek.misto.cz/english/robot.html.
[3] Published in: Isaac Asimov, I, Robot, New York, 1950.
[4] Isaac Asimov, Robots and Empire, New York 1985.
[5] According to iRobot, the manufacturer of Roomba, more than 2 million units have been sold worldwide until 2008 (http://www.irobot.com/sp.cfm?pageid=74).

Reaper Unmanned Air Systems (UAS). While the commercial products can make your life easier, the military ones could end your life!

Since 1994, when the U.S. Department of Defence commissioned the production of ten Predators of which the first ones were deployed in Bosnia in July 1995[6], the number of UAS has risen steadily. In total there are over seven thousand UAS in service in the U.S. Armed Forces in 2010 as opposed to 167 in 2001.[7]

The spread of robotic systems is not merely a military phenomenon but constitutes a trend of the society as a whole. According to the Statistical Department of the International Federation of Robotics, in 2007 6.5 million robots were in use worldwide with 18 million predicted for 2011[8], ranging from industrial robots to service and entertainment robots. Industrial robots, numbering approximately 1 million[9] as of today, have been growing steadily at about 100.000 per year.[10] In contrast, service robots for professional use, such as military robots, but also entertainment robots, are seen as the field where most of the growth will be located in the near future. [11]

The history of the use of UAS by the military goes back as far as the 19th century, with the Austrian Army under Franz von Uchatius using unmanned balloon bombs in 1849 in the siege of Venice. Similar concepts had also been developed in the American Civil War, though they were not deployed. [12] The development has been driven on by Nikola Tesla, Archibald Low and many others to the point that over the period of the Second World War that U.S. Forces had produced almost 1.000 units of the Radioplane OQ-2A UAV model alone.[13]

---

[6] http://www.af.mil/information/transcripts/story.asp?storyID=123006556 and Statement of John F. Tierney, Chairman, Subcommittee on National Security and Foreign Affairs, Committee on Oversight and Government Reform, U.S. House of Representatives: Hearing on "Rise of the Drones: Unmanned Systems and the Future of War "http://www.oversight.house.gov/images/stories/subcommittees/NS_Subcommittee/3.23.10_Drones/3-23-10_JFT_Opening_Statement_FINAL_for_Delivery.pdf.

[7] http://www.nytimes.com/2009/03/17/business/17uav.html?_r=1&hp.

[8] http://www.worldrobotics.org/downloads/2008_Pressinfo_english.pdf.

[9] http://www.ifrstat.org/downloads/2009_First_News_of_Worldrobotics.pdf.

[10] In 2007 118.000 additional units have been produced. (http://www.ifrstat.org/downloads/Pressinfo_11_Jun_2008_deutsch.pdf).

[11] Growth rate from 33% in the sector of service robots (http://www.ifrstat.org/downloads/2009_First_News_of_Worldrobotics.pdf).

[12] http://www.ctie.monash.edu.au/hargrave/rpav_home.html.

[13] http://www.nationalmuseum.af.mil/factsheets/factsheet.asp?id=486.

Unmanned Systems and the Military

Why is it, that a technology that has been used by the military for decades, should now revolutionize warfare itself? There are a number of aspects, which are to be considered.

Firstly, war spurs the development of militarily relevant technology. This has been true for centuries, and remains so today. Looking at the ongoing Operation Iraqi Freedom, and the widespread adoption of Explosive Ordnance Disposal (EOD) robots, we see them dealing with the emergence of the Improvised Explosive Device threat. At the start of the conflict, there were virtually none of these systems in use. Today, they number in the thousands, and the EOD technicians know that every mangled robot that comes into the repair facilities represents at least one life saved. [14]

If we shift our view to Operation Enduring Freedom in Afghanistan, and neighboring Pakistan, we see the same sort of thing with the increased use of surveillance, and armed Predators, and now the armed Reapers. The US administration would not have moved in these directions if there wasn't a clear benefit in doing so, and the pressure to add more systems to inventory show that the demand for this benefit hasn't been met.

What should we draw from this? First, it is obvious that these systems are saving lives. Second, it is clear that the "persistent stare" that these systems provide, coupled with weapons, is providing increased knowledge of the battlespace, and the ability to strike time-critical targets. Thirdly, there is no reason to believe that the push to develop more capable systems will drop off anytime soon, since these conflicts are continuing.

This brings us to the consideration of how future war may be conducted, and possibly in the not-too-distant future at that: Today's unmanned systems are not what most people think of as really being robots. For the most part, they operate with "man-in-the-loop remotely" control. This is particularly true for the use of weapons by one of these systems. We can expect to see a push to develop higher-level autonomy for operations by these machines to include the autonomous use of weapons.

[14] E.g. Noah Shachtman, The Baghdad Bomb Squad in: Wired Magazine (2005) (http://www.wired.com/wired/archive/13.11/bomb.html?pg=3&topic=bomb).

Secondly, the developments in engineering, sensor technology and especially computer systems and information technology, have made it possible to increasingly exploit the potential of unmanned systems. Even if the Revolution in Military Affairs (RMA) has not proven to be as effective as predicted, the concept of network-centric warfare did lay a foundation for the use of unmanned systems (and in this case especially for the use of UAS in surveillance and intelligence gathering).

Another aspect to be considered is the impact of unmanned systems on the strained budgets of the militaries throughout the world. It has been argued, that with unmanned systems, fewer soldiers will be needed to cover the growing areas of the current battlefields of counterinsurgency operations[15]. In addition, at least in the field of UAS, where unmanned systems can fulfill most of the roles of manned aircraft, they have proven to be generally cheaper in production and deployment than manned systems. On the other hand, it has also been noted, that the benefits of new possibilities like "persistent stare", result in more workload and require more personnel to maintain and operate these systems.[16]

Today's armed unmanned systems place an expensive machine between the soldier and his weapon. For small numbers of machines, this may not be much of an issue, but for large numbers of machines, this increases the cost of conducting warfare substantially.[17] The push is on to move from a "one operator, one machine" model of operations to a "one operator, many machines" model of operations in order to reduce the total cost of ownership by decreasing the cost of manpower needed,[18] as typically, the largest life-cycle cost item for a system is personnel.

One of the main aspects of change will be constituted by the impact of autonomous potential of military unmanned systems on warfare, something

[15] A Look at the Future Combat Systems (Brigade Combat Team) Program. An Interview With MG Charles A. Cartwright in: Army AL&T Magazine 2/2008.
[16] John Canning, A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles, NSWCDD/TR-0/36, 2005, p.13.
[17] John Canning, A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles, NSWCDD/TR-0/36, 2005, p.14.
[18] E.g. the development of a Multi-Robot Operator Control Unit for Unmanned Systems (http://www.spawar.navy.mil/robots/pubs/DefenseTechBriefs%20-%20MOCU%202008%2008%2001.pdf.

which, in its implementation, is yet difficult to predict[19]. The same applies to the role of autonomous robots in the human society as a whole, as Bill Gates has compared the present situation of the robotics industry with the situation of the computer industry in the 1970s.[20] Although, with the political agenda as it is, it can be considered as a certainty, that these systems will have a profound impact on the future of warfare and the role of the warfighter himself.[21]

Legal Aspects

First, let us stipulate that we are not talking about either ethical or legal aspects associated with any other area than with weaponization of robots. There are others that are looking at things such as safety of flight for UAS in the US National Airspace System, and associated legal concerns. Nor will we concern ourselves with issues such as the Collision-avoidance Regulations (COLREGS), known as the "rules of the road" for international sea-based navigation. We will not comment beyond weaponization aspects.

What are the legal aspects and challenges of the development and deployment of weaponized unmanned systems by the military? What is their impact on warfare and how could the use of military unmanned systems be regulated?

The first amended Protocol relating to the Protection of Victims of International Armed Conflicts from the 8th of June 1977 to the Geneva Convention from the 12th of August 1949 relative to the Protection of Civilian Persons in Time of War states under Article 36, that "in the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment

[19] E.g. : „Dramatic progress in supposrting technologies suggests that unprecendented, perhaps unimagined, degrees of autonomy can be introduced into current and future military systems. This could presage dramatic changes in military capability and force composition comparable to the introduction of ‚Net-Centricity'." Task Force (29.03.2010): Role of Autonomy in Department of Defense (DOD) Systems, The Under Secretary of Defense, Acquisition, Technology and Logistics: Memorandum for Chairman, Defense Science Board, http://www.acq.osd.mil/dsb/tors/TOR-2010-03-29-Autonomy_in_DoD_Systems.pdf.

[20] Bill Gates, Scientific American, 1/2007 (http://www.scientificamerican.com/article.cfm?id=a-robot-in-every-home).

[21] In his campaign, President Obama has identified unmanned systems as one of the five important military systems. Also the budget in this area has – unlike in many other areas of military spending – not been cut but increased. Peter W. Singer, Interview vom 5.8.2009 (http://www.irf.ac.at/index.php?option=com_content&task=view&id=293&Itemid=1).

would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law".[22]

On an international level, at the present time, there are no comprehensive treaties regarding the use and development of unmanned systems[23], though on a national level the use and development is regulated by the appropriate rules of law. In the United States, for example, the Armed Forces have to ensure the accordance of a new weapon system with international treaties, national law and with the humanitarian and customary international law. To ensure this, a new weapon system has to be approved in an evaluation process by the Judge Advocate General's Corps, the legal branch of the U.S. Armed Forces.[24]

In addition all branches of the Armed Forces have separate regulations, which specify the details of the evaluation process. A typical evaluation process would include the military necessity for the weapon; the ability of the weapon to distinguish lawful targets from protected persons and objects (i.e. discrimination); whether the damage caused by the weapon causes unnecessary suffering; treaties that may prohibit the acquisition and employment of the weapon, and domestic law. In addition the deployment and use of the weapon system would be governed by the current Rules of Engagement.[25]

It is the ability to discriminate between a lawful and unlawful target that drives most of the ethics concerns for armed robots, although the consideration for causing unnecessary suffering is not far behind. The latter is referred-to as a "collateral damage" issue, while the former is a "targeting" issue.[26]

---

[22] http://www.icrc.org/ihl.nsf/FULL/470?OpenDocument; It has to be noted, that this article refers to the use and development of weapons, but not their possession, as the protocol solely regulates international armed conflict. See: International Committee of the Red Cross, Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949, Geneva 1987, 1471. (http://www.icrc.org/ihl.nsf/COM/470-750046?OpenDocument); regarding peacekeeping and International Humanitarian Law see e.g.: Ray Murphy, United Nations Military Operations and International Humanitarian Law: What Rules Apply to Peacekeepers? In: Criminal Law Forum, Volume 14, Number 2 / Juni 2003, p. 153-194.

[23] Except for the the Missile Technology Control Regime (originated 1987), an informal and voluntary association of countries (34 in 2009)which share the goals of non-proliferation of unmanned delivery systems capable of delivering weapons of mass destruction.

[24] The necessity for this evaluation process is laid down in the Department of Defence Instruction 5000.1, E.1.15. (http://www.dtic.mil/whs/directives/corres/pdf/500001p.pdf).

[25] John S. Canning, Legal vs. Policy Issues for Armed Unmanned Systems, 2008: http://www.unsysinst.org/forum/download.php?id=51).

[26] Concerning the issue of „targeted killing" see Armin Krishnan, Killer Robots. Legailty and Ethicality of Autonomous Weapons, Farnham/Burlington 2009, p. 100-103.

These issues are considered separately during the “legal weapons review,” prior to full-scale production and use, and for its actual use on the battlefield. It is noted though that any “legal weapon” could be used in an illegal manner. The use of weapons on the battlefield is therefore addressed by the “Rules Of Engagement”.

The complexity and various dimensions of legal regulations concerning the use of weapon systems can be observed in the discussion of the use of weaponized UAVs by the United States in Pakistan. This topic, discussed intensely by the international community[27], has also been addressed by the Subcommittee on National Security and Foreign Affairs of the House of Representatives in two prominent hearings[28].

Robots and Humans – Changes in Warfare

Robots have no life to lose. There, in a nutshell, is the primary change in conducting warfare by using robots. Humans, however, are still mortal, and can be killed. Robots also know no suffering. This, too, is a primary change in conducting warfare by using robots. Robots can be damaged or destroyed, however. If damaged, they can be fixed. If destroyed, they can be replaced. If a human is killed, he (or she) is gone forever. If they are injured, it could be with irreparable damage such as losing a limb, and the quality of their remaining lives reduced as a result.

One of the less expected effects of the use of unmanned ground vehicles (UGVs) was the emotional link that human operators began to establish to the systems they teleoperated. This emotional bond between robots and humans has also shown the potential to endanger soldiers on the battlefield. There have been reports, that soldiers are taking excessive risks to retrieve unmanned systems under enemy fire to save them from

[27] See e.g.: Nils Melzer, Targetted Killing in International Law, Oxford/ New York 2008. and the Report of the Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Philip Alston,
http://www2.ohchr.org/english/bodies/hrcouncil/docs/14session/A.HRC.14.24.pdf.
[28] „Rise of the Drones: Unmanned Systems and the Future of War“:
http://www.oversight.house.gov/index.php?option=com_jcalpro&Itemid=19&extmode=view&extid=136 and „The Rise of the Drones II: Examining the Legality of Unmanned Targeting“:
http://www.oversight.house.gov/index.php?option=com_content&view=article&id=4903:hearing-on-the-rise-of-the-drones-ii-examining-the-legality-of-unmanned-targeting&catid=72:hearings&Itemid=30.

destruction.[29] This coincides with naming repair-shops for unmanned systems "robot hospitals"[30], the practice of operators to name and relate to their equipment similar as they would do with pets. [31] Recent studies suggest that with advanced artificial intelligence and robotics this phenomenon will be something that the human society will have to reckon with in all aspects of human-robot interaction.[32]

Another aspect normally not associated with ethical challenges of unmanned systems, is the change of the self-image of the warfighter and the role of the soldier operating unmanned vehicles through long distances. While living in the U.S., UAS operators fly their missions in Iraq and Afghanistan and return to their homes afterwards just as with a normal day at office. It has been argued, that this can be psychologically problematic for the UAS operators not only because of the dual experience of being at home and being at war at the same time but also because due to the kind of deployment they also experience a change in camaraderie. UAS Operators are said to experience combat stress on similar levels as soldiers deployed in Iraq but lack the possibility to share these experiences with other members of their unit and therefore do not as a unit have a rest and recovery period to cope with these experiences.[33] However, recent reports from the USAF indicate, that though it is yet not fully clear how these factors will influence the psyche and also the relationships of soldiers experiencing this in a way paradox variant of warfare, the impact might be a lot less substantial than generally assumed.[34]

---

[29] Peter W. Singer, Interview vom 5.8.2009 (http://www.irf.ac.at/index.php?option=com_content&task=view&id=293&Itemid=1).

[30] http://www.army-guide.com/eng/article/article_1050.html.

[31] E.g. the packbot named "Scooby-Doo" (http://news.cnet.com/2300-11386_3-10000731-6.html?tag=mncol). There have also been accounts that soldiers did not want a damaged robot to be merely replaced but they wanted this individual robot repaired. Peter W. Singer, Interview vom 5.8.2009 (http://www.irf.ac.at/index.php?option=com_content&task=view&id=293&Itemid=1). Peter W. Singer also reports an incident, where a Commander, after a UGVs was destroyed, writes a condolence letter to the manufacturer. Peter W. Singer, Wired for War. The Robotics Revolution and Conflict in the Twenty-first Century, New York 2009, 20-21.

[32] Fumihide Tanaka, Aaron Cicourel, Javier R. Movellan, Socialization between toddlers and robots at an early childhood education center, 2007, http://www.pnas.org/content/104/46/17954.full.

[33] Peter W. Singer, Interview vom 5.8.2009 (http://www.irf.ac.at/index.php?option=com_content&task=view&id=293&Itemid=1).

[34] AUVSI Unmanned Systems North America 2009, Panel: Ethics in Armed Unmanned Systems in Combat, Washington DC, 12.8.2009.

Two Ways of approaching the Ethical Challenge

The Ethical Governor

Dr. Ronald C. Arkin, from the Georgia Institute of Technology, has proposed the concept of what amounts to an ethical governor for armed unmanned systems.[35] Basically, this is an AI "ethics module" that would dispassionately process the existing Rules Of Engagement and make more ethical decisions regarding engagements than a human soldier could. An autonomous, armed machine so-equipped would then proceed to use lethal force against an enemy target, including the possible direct targeting of human enemy combatants, while at the same time avoiding the targeting and killing of non-combatants, or the engaging of other illegal targets. While potentially a more ethical approach to warfare than what exists today, there are two issues with this approach: (1) the bug-free development of the ethics module itself; and (2) the fact that this would have a machine autonomously targeting and killing people.

Regarding the bug-free development of the ethics module:

There is an entire industry today built around the concept of "software maintenance." Basically, this is the fixing of software problems that become apparent after an item has been delivered to the field for use. Most professional software developers would state that the probability of delivering a completely bug-free product, in something as complex as an ethics module, the first time around would have to be near zero – even with extensive testing beforehand. The unanswered question is "How long would it be before all the bugs are worked-out?" There may be no way of answering this question since how would you know if you had actually eliminated the last bug?

Regarding having a machine that can autonomously target and kill people:

Based on conversations with lawyers from the U.S. Navy's JAG Office in the Pentagon, and with the U.S. Office of the Secretary of Defense's Office of General Counsel[36], it is unlikely that such a system would be allowed to pass a legal weapons review, simply because of the fact that it would be

[35] http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf.
[36] John Canning, „You've Just Been Disarmed. Have a Nice Day!" in: IEEE p.15.

targeting a human. The issue is, particularly on today's battlefield, how do you tell an insurgent from an innocent civilian ("target discrimination")? They are both dressed the same and look alike. This is a tough problem for our human troops to handle today. It won't be any easier for a machine.

The Moral User

Peter Asaro recently has proposed an approach for tele-operated systems which centers on the ethical decision-making of the human operators. Asaro argues that Arkin's, and similar approaches, do not sufficiently take into account that the basis for ethical decision-making in warfare, Law of Armed Combat, Rules of Engagement and Just War Theory, are not always a set of clearcut rules but do include a hodgepodge of laws, rules, heuristics and principles subject to interpretation and value judgments.[37]

Therefore, drawing upon User-Centered Design, he brings forward his idea of "modeling the moral user", which would involve three elements. First, using the methods of cognitive psychology, the representations, decision rules and perceptual and emotional requirements for effective ethical decision-making should be sought to be understood. Second, drawing upon recent work in experimental philosophy, we should explore the nature of moral intuition, value comparisons and judgments and using experimental economics, we should also engage the nature of risk assessment and probability estimation. He also points out, that it might be necessary to evaluate the significance of rational thought in ethical decision making. Third, it would be necessary for the society to decide which ethical standards it wants to promote and to which extent it will be able to enforce these standards on the soldiers through the technology.[38]

Contrary to arguments, which see psychological stress mainly as a cause for unethical behavior, Asaro points out, that it might be necessary for operators of unmanned systems to experience these factors in order to make effective ethical decisions and to feel empathetic and sympathetic emotions.[39] Without prejudging any questions about the nature of morality – can an artificial intelligence or unmanned system gain a level of moral agency or not – the question if we decide to imagine unmanned systems as rule-based entities or if

[37] Peter Asaro, Modeling the Moral User: Designing Ethical Interfaces for Tele-Operation, in: IEEE Technology and Society 28/Spring 2009, p. 22.
[38] Asaro, IEEE p.23.
[39] Asaro, IEEE, p.24.

we strive to implement an emotional component, might very well become a crucial point for future developments in this field.

Another one of the key questions Asaro identifies is, that in the aim to effectively capture the range of moral reasoning, it might be necessary to consider that there can very well be a range of individual and cultural variations in ethical reasoning as well as different values and standards of moral reasoning.[40] Following the idea that warfare is a cultural practice and that of cultural and individual morals, Asaro continues to ask which ethical standards we should chose to implement in the design of unmanned systems and if the implementation of an ethical software system would in fact make the person operating it more ethical.[41]

Though Asaro mainly concentrates on systems at hand, which are teleoperated systems, there seems no inconsistency to widen the scope on to autonomous unmanned systems. However this may be, if we decide to accept, that it is a widely shared current ethical standard of warfare to expose other people to as little negative influence as possible but necessary to achieve a task, then averting the needless loss of life during warfare seems not only a sensible goal but leads us to an approach where we might find that removing the lethal component from armed conflict might be a way to solve – at least for the moment – the most prominent question concerning autonomous armed unmanned systems, that is, shall it be possible for a machine to act with the potential consequence of humans losing their life?

Managing the Ethical Challenge of Autonomous Use of Lethal Force – "You have been Disarmed"

Another approach to the autonomous use of force has been put forward by John Canning, following extensive discussions with representatives of the US Navy's JAG Office. It was noted that this JAG Office was going to require that weapons-bearing unmanned systems would be required to maintain a "man-in-the-loop" for target discrimination and weapons control, if they were designed to target people. It was noted, however, that if they were designed to target either the "bow" or the "arrow," but not the human

[40] Asaro, IEEE, p.23.
[41] Interview with Peter Asaro, 8.9.2009 (http://www.irf.ac.at/index.php?option=com_content&task=view&id=295&Itemid=22).

"archer," then there was the possibility for the autonomous use of weapons[42]. Pulling this thread, Canning discovered many weapon systems that had already been designed and fielded, based on this concept. Several examples: AEGIS weapon systems on US Navy ships when set to the AUTO-SPECIAL mode of operation; CAPTOR mine systems that would target enemy submarines, but not surface ships; the US Army's PATRIOT missile system in a mode similar to AEGIS' AUTO-SPECIAL mode.

In contrast, it was shown that anti-personnel landmines have been outlawed because they can't discriminate between a soldier and a child, but anti-tank landmines are still legal to use because they target "things" – not "people."[43]

Canning has taken this one step further by pointing-out that the weapon used by a robot does not have to be a traditional gun or missile, where there may be a substantial likelihood of collateral damage, but something else might be used instead. He is fond of saying that his "dream machine" is one that marches up to an enemy combatant on the battlefield; physically takes the rifle out of his hands; saws the rifle in half with a diamond-tipped saw; hands the two halves back to the enemy combatant; and then tells him to "Have a nice day!"[44]

The question is then one of "Is the enemy carrying his bow, such as a rifle or pistol, or is he riding it, such as a tank or warship?" Non-lethal weapons, such as Active Denial, might be used to separate an enemy combatant from his "bow" if he is carrying it, but if he is riding his bow, it is not necessary to achieve a "platform kill" in which a ship is totally sunk (drowning the crew), or a tank is obliterated (killing the crew). It may be enough to achieve either a "mobility kill," where you disable either the motor or the steering mechanism on a ship, or a "mission kill," where you might poke a hole through a tank's main gun barrel, thereby rendering it useless. However, even if a crew is killed or injured, they still do constitute a legitimate target under international humanitarian law, so in this case, certain, limited, amount of human collateral damage may be acceptable.

---

[42] John Canning, „You've Just Been Disarmed. Have a Nice Day!" in: IEEE Technology and Society 28/Spring 2009, p.12-15.

[43] Also see Patrick Hew, Autonomous Situation Awareness. Implications for Future Warfighting in: Australian Defence Force Journal, Issue 174, 2007, pp77-78 and pp 83-84.
The Western Militaries' Blind Spot in Robot-Enabled Warfare, in print.

[44] John Canning, „You've Just Been Disarmed. Have a Nice Day!" in: IEEE Technology and Society 28/Spring 2009, p.12-15.

Conclusion

As we have just shown, ethical considerations for robots have been around from the inception of the term “robot.” For most of the intervening time, “popular” ethics for robots were defined by Isaac Asimov’s science fictional works, but the near-at-hand development of real armed, autonomous military robots is forcing us to seriously consider the ethics of these machines in more pragmatic terms. Driven heavily by legal concerns for target discrimination, we are channeled into autonomously targeting either the “bow,” or the “arrow,” but not the human “archer,” thereby bringing up the possibility of disarming a foe, as opposed to killing him. This is a fundamental paradigm shift from the way mankind conducts warfare today. We would argue that this also marks a fundamental improvement to the ethics of conducting war. While this is an ethical challenge, we would argue it is one we cannot afford to ignore.

Disclaimer

The views or opinions contributed by Mr. Canning, and expressed in this document are those of Mr. Canning and do not represent the official position or policy of the United States Navy.

The views or opinions contributed by Mr. Dabringer, and expressed in this document are those of Mr. Dabringer and do not represent the official position or policy of the Austrian Military Chaplaincy.

Colin Allen:
Morality and Artificial Intelligence

How and why did you get interested in the field of machine morality?

The question of how to make a machine behave ethically was on a list compiled by one of artificial intelligence's luminaries of topics where philosophers could help. It seemed like an interesting challenge and I had just been invited to write an article for an artificial intelligence journal, so I decided to see where I could take it.

Artificial Intelligence, Machine learning and genetic programming, just to name a few branches, are highly complex fields of research. Coming as you did from a meta-science, how did you approach this challenge from an ethical perspective?

Well, let me start by saying I am not an ethicist! I'm a philosopher of science and philosopher of mind who to that point had mostly worked on issues in animal cognition, but I had also taken quite a few post-graduate courses in computer science, specializing in artificial intelligence. So, the first thing I did was to talk to an ethicist colleague of mine, Gary Varner, about my ideas for the article and he agreed to be a co-author. My approach was initially to ask the same technical questions about whether ethical theories such as Kant's or Bentham's could in fact be computed. Later, in the book with Wendell Wallach, this became what we called the "top down" approach.

Your book "Moral Machines" discusses the field of machines as moral agents. Should we define morality as purely human quality or should we use a concept of different qualities of morality? Also from a practical perspective: what concept of morality should we use while discussing the issues right at hand?

Wendell and I wrote the book with a very practical question in mind: How, as a matter of fact, would one improve the kinds of machines that are already being built so that they could be better, morally speaking? As such, we didn't want to prejudge any questions about the nature of morality, who or what has it, etc. We recognized that philosophers tend to gravitate towards the hard cases where there is much disagreement, because this is where theories get tested against intuitions. But despite this, there's a surprising amount of agreement about practical ethics. Whether

you're a utilitarian or Kantian, Christian or Buddhist, you can agree that stabbing the stranger sitting next to you on the train is morally bad, or, more subtly, that anyone to whom we cause a harm has a prima facie moral claim against us. Of course, there's lots of room for disagreement about what constitutes a harm, and when it is acceptable to cause a harm, but our basic premise was that most machines, robots and software bots, that are currently making harmful decisions don't even have the means to take those harms into account when making these decisions.

You have used the term "artificial moral agents", why and how would you differentiate natural from artificial moral agents?

Like artificial anything, we want to acknowledge that deliberately engineered products will not be the same as those that have grown organically. Artificial sweeteners aren't the same as sugars, and artificial intelligence only resembles biological intelligence. Whether artificial moral agents ever become as capable as biological moral agents is a question for science fiction and futurism. I should also acknowledge that for some ethical theorists, the central problem of moral agency is the conflict between selfish inclination and moral duty, but this assumes a form of psychology that may not apply to artificial agents. Nevertheless, for the time being we know that any artificial system we place in an ethically charged decision making situation will have strengths and limitations. Many of those limitations stem from our not really understanding, either at a scientific or humanistic level, what goes into making us moral agents. (Lots of theories, no consensus.) So in part the project of building artificial moral agents is partly a project of self- evaluation. If we don't flag what we're doing with the term "artificial" there's a risk of losing sight of our own role in shaping these systems.

Are there beneficial aspects of looking at morality from the perspective of the artificial intelligence theory?

One of the interesting things, I think, that comes out of the attempt to think in computational terms about morality or ethics is a richer conception of the space in which ethical behavior operates. Rather than seeing these as opposite poles, I'm more inclined to see them as separate axes or dimensions of the decision space. The time- and information-bounded nature of most decision making makes embodied dispositions an essential part of moral agency. There simply isn't enough time in the world to compute all of the consequences, actual or logical, of an action, even if one had perfect information. So, moral agents must

be disposed to react in ways that are morally acceptable.

These bottom up reactivities are also, however, subject to top-down evaluation, and, here emotions like pride, regret, or shame can serve to strengthen or weaken dispositions, but so can a reasoned determination to live up to an abstract principle. Given the abstract nature of most top-down principles, however, it is hardly surprising that they sometimes conflict with each other and with our dispositionally-formed intuitions. The result is that any moral principle could be overridden in a specific situation. As socially-enculturated human beings, it is natural for us to want to come up with some higher principle to adjudicate these conflicts, but in the absence of such a principle, what one has is a decision space in which duties, consequences, and dispositions are all relevant dimensions, but none is paramount. Moral agency involves a hybrid of bottom up and top down processes, often operating over different time scales. "Shoot first, ask questions later" is the wrong slogan because we can ask some questions first, but our ability to do so is often limited and we must return to the questions in retrospect, hoping to calibrate the shooting response better next time we are in a similar situation.

We are a long way from being able to build hybrid architectures for artificial moral agents to have such sophistication. But a chief goal of the book is to start a discussion about whether providing machines with just part of the bottom up or top down capacities for moral decision making would be better than having machines that are ethically insensitive to such considerations. What information does a battlefield robot or your bank's computer have to have in order to make decisions that a human moral agent would endorse? What reasoning capabilities would it need to be able to weigh collective outcomes against individual rights, either prospectively or retroactively?

Most people see robots and computers as predetermined machines without any ability to transcend into the sphere of decision making. How was your approach to this topic and how did people respond to your concept of artificial moral agents?

Whether predetermined or not, the fact is that machines are involved in all sorts of decisions, from approving credit card transactions to allocating resources in hospitals. They are even being deployed as automatic sentries on national borders. I'm not sure whether this means that they have "transcended into the sphere of decision making" but it does mean that without direct human oversight machines are selecting among options that have moral consequences. The metaphysical

questions about whether this is "really" decision making don't concern me as much as the practical questions about whether these machines can be made sophisticated enough to weigh the factors that are important for ethical decision making.

People react to the idea of artificial moral agents in several ways. Some assume that we are talking about human-level artificial intelligence and dismiss the topic as pure science fiction, and others assume we must be concerned with whether robots themselves deserve rights. For me, however, it is important to avoid science fiction and stay focused on what is likely to happen in the next decade or so. A different kind of worry comes from those who say that by using the word "agents" for machines we are contributing to the abdication of human responsibility for the consequences of our own technologies. I recognize the seriousness of the concern, but I think it's also likely that by referring to artificial moral agents we set up a kind of dissonance that might help users recognize that they should be wary of overestimating the capacities of these machines.

So what you are saying is, that right now we should focus more on the practical ethical challenges at hand which arise from the use of these systems (e.g. the Future Attribute Screening Technology (FAST) – Hostile Intent Detection of the Department of Homeland Security[1] than to engage in speculation on full moral agency of machines. Do you think that your book could be something like a whistleblower by starting this discussion?

It was certainly our intention to help start such a discussion. And it's interesting that we seem to be in the middle of a small explosion of interest in the topic. Just after our book came out, Peter Singer's more journalistic Wired for War came out to significant press coverage, and now Ron Arkin's Governing Lethal Behavior in Autonomous Robots has just been released, the first book to provide an actual design specification for robots capable of exercising lethal force. While these other books focus on military applications, we think it's important to recognize that the issues go far beyond the battlefield.

In your book you have put forward two dimensions for artificial moral agents: ethical sensitivity and autonomy. On this framework you differentiate between operational and functional morality as well as finally full moral agency. How can we understand these moralities and where on this framework are robots now (and where can they probably be finally)?

There are not intended to be hard and fast distinctions, but operational

morality is intended to include cases where the decisions about what is a morally acceptable behavior are largely in the hands of the designers and programmers, whereas functional morality implies some built-in capacities for moral reasoning or decision making. Operational morality generally applies to machines that operate in relatively closed environments with relatively few options for action.

Under these circumstances, designers may be able to anticipate the situations the machine will encounter and pre-specify the morally preferred actions in those circumstances. In more open environments where machines have greater autonomy, they must be designed to detect ethically relevant features of the situation, and select among options accordingly. We use the term "functional morality" primarily to acknowledge that these capacities may fall short of the full moral agency of human beings, although I would like to maintain that it's an open question whether there are any limits to what machines can do. At the current time, machine autonomy is increasing, meaning that machines are operating in more open environments without human oversight and with more options available to them. But aside from a few A.I. projects that are described in chapters 9 and 10 of the book, there is relatively little work on giving machines the kind of ethical sensitivity that, in combination with autonomy, would be necessary for functional morality.

Why do you think it is like that? It seems obvious that there is a need for research on this matter.

I don't think it is a deliberate omission, but a sign of how new the field is. Engineers tend to prefer well-defined problems, and as I've already mentioned, philosophers like controversial topics. For this and other reasons it's actually quite a challenge to bring the two cultures together. But it is coming. In addition to our book and the others that have recently appeared, a scholarly collection of essays edited by the computer scientist-philosopher husband-wife team of Michael and Susan Anderson is in the works. And a couple of graduate student projects that I'm aware of show that they are starting to pay attention are thinking creatively about how ethical capabilities might be important in a variety of online and real-world contexts.

What can robots with representations of emotions – like the projects KISMET and later on Nexi MDS – do for the development of artificial moral agents?

I think emotion-representing robots do two things for artificial moral agents. One is perhaps quite dangerous, in that it may cause people

to attribute more understanding of their own emotions to the machines than there really is. If Kismet or Nexi reacts to a person's sad face by looking sad, there is a risk that the person will assume more empathy than exists.

This is dangerous if the machine lacks the capacity to help the person properly deal with the situation that is causing the sadness. The other thing may be essential, however, since part of the ethical sensitivity required for functional morality involves being able to detect and react to the emotional responses of people interacting with the robot. All other things being equal, if a robot through its actions causes anger or sadness, then it needs to reevaluate that action. This is not to say that robots should always change their behavior whenever they detect a negative emotional response, or do whatever it takes to get a positive emotional response from the people it is interacting with. But such responses are crucial pieces of information in assessing the moral appropriateness of actions.

The KISMET Project has been very well documented and the emotional responses you refer to can be seen on videos on the webpage of the MIT Computer Science and Artificial Intelligence Laboratory[2]. What do you think about the use of robots in the entertainment industry? In some countries in Asia robots are being developed explicitly as "personal companions". What impact will that have on interpersonal relations of humans, especially children growing up with robotic pets?

The sex industry has driven a lot of technology development, from the earliest faxes through postcards to videotape recording and online video on demand. The more "respectable" face of robotic companions for the elderly and toys for children are just the tip of a very large iceberg. I think it's hard to say what kind of impact these technologies will have for human interpersonal relationships. It will probably bring benefits and costs, just as with the Internet itself. It's easy to find lots of people who lament the replacement of face-to-face interactions with Facebook, Twitter, and the like. But at the same time probably all of us can think of old friendships renewed, or remote relationships strengthened by the use of email and online social networking. I don't think robotic pets are inherently bad for children, although I am sure there are those who will complain that one doesn't have to be as imaginative with a robot as with a stuffed toy. I'm not so sure this is correct. With a robotic toy, a child may be able to imagine different possibilities, and a robotic pet will likely serve as a nexus of interactions in play with other children. And just as we are finding that highly interactive video

games can bring cognitive benefits to young[3] and old[4] alike, we may find that robotic companions do likewise. Of course there will be problems too, so we must remain vigilant without being fearful that change is always a bad thing.

Free will, understanding and consciousness are seen as crucial for moral decisions though they are often attributed exclusively to humans. You have argued that functional equivalence of behaviour is what really matters in the practical issues of designing artificial moral agents. What is your perspective on these three fields concerning artificial moral agents?

All of these are again looking towards more futuristic end of this discussion. People in A.I. have for over 50 years been saying that we'll have full human equivalency in 50 years. I don't know whether it will be 50 years or 100 years or never, because I don't think we know enough about human understanding, consciousness, and free will to know what's technologically feasible. My stance, though, is that it doesn't really matter. Military planners are already sponsoring the development of battlefield robots that will have greater autonomous capacities than the hundreds of remote-operated vehicles that are already in use. The military are sufficiently concerned about the ethical issues that they are funding research into the question of whether autonomous robots can be programmed to follow the Geneva conventions and other rules of war. These questions are pressing regardless of whether these machines are conscious or have free will. But if you want my futuristic speculation, then I'm a bit more pessimistic than those who are predicting a rapid take-off for machine intelligence in the next 25-30 years, but I would be very surprised if my grand- children or great-grandchildren aren't surrounded by robots that can do anything a person can do, physically or cognitively.

As you said military robots are a reality on the battlefields today and it seems clear that their number and roles will expand, probably faster than most of us think or would like them to. Do you think that the military is actually ready for the changes these semiautonomous systems bring to the army?

I'm encouraged by the fact that at least some people in the military understand the problem and they are willing to support research into solutions. Both the U.S. Navy and Army have funded projects looking at ethical behavior in robots. Of course, it's possible to be cynical and assume that they are simply trying to provide cover for more and more impersonal ways of killing people in war. But I think this underestimates the variety and

sophistication of military officers, many of whom do have deep moral concerns about modern warfare. Whether the military as a whole is ready for the changes is a different matter perhaps, because for someone on the front lines, sending a robot into a cave with authorization to kill anything that moves may seem like a pretty attractive idea. There will be missteps – there always have been – and I'm fairly sure that the military is not actually ready for all the changes that these systems will bring because some of those changes are unpredictable.

One of your other fields of study has been animal cognition. Have you found this helpful while developing your perspectives on artificial moral agents?

It's a good question because I started off really treating these as separate projects. However, thinking about the capacities of nonhuman animals, and the fact that it isn't really a dog-eat-dog world, leads to some ideas about the behavioral, cognitive, and evolutionary foundations of human morality. Various forms of pro-social (and proto-ethical) behavior are increasingly being reported by experimentalists and observers of natural behavior of animals. Of course, nonhuman animals aren't, as far as we know, reflective deliberators, but neither is all of basic human decency and kindness driven by explicit ethical reasoning. Animals give us some ideas about the possibilities for machines that aren't full moral agents.

So you are referring to studies like Benjamin Libet's through which the absolute predominance of reason in human decision making is questioned in favour of subconscious processes. It is easily comprehensible that these concepts will be seminal, though it seems to be harder to create a model of ethical behaviour by the means of animals, considering the complexity of the mind, than developing a simpler rule-based behaviour system. What do you think are the main areas where the development of artificial morality could benefit from the research in animal cognition? Or maybe one could even say, that concepts which stem from this field are crucial for a realistic approach to artificial morality?

One of the things we are learning from animals is that they can be quite sensitive to reciprocity of exchange in long term relationships. If one animal shares food with or grooms another, there doesn't have to be an immediate quid pro quo. Speaking only slightly anthropomorphically one could say that they build relationships of trust, and there is even evidence that early play bouts may provide a foundation for such trust. These foundations support generally "pro-social" behavior.

Humans are no different, in that we establish trust incrementally. However, what's remarkable about human society is that we frequently trust total strangers and it usually turns out all right. This is not a consciously reasoned decision and, as recent research in behavioral economics shows, may even involve acting against our immediate self interests. Artificial moral agents will also have to operate in the context of human society with its mixture of personal relationships based on medium to long term reciprocity and transactions with strangers that depend for their success on local social norms. Ethical robots have to be pro-social, but not foolishly so. Animal studies can do a lot to help us understand the evolution and development of pro-social behavior, and some of this will be transferable to our robot designs.

The purpose of the already mentioned NEXI MDS project at the MIT Personal Robots Group[5] is to support research and education goals in human-robot interaction, teaming, and social learning. Do you think projects like this which focus on the improvement of robots for interpersonal relations could benefit from the research in animal behaviour?

I recently attended a conference in Budapest on comparative social cognition that had both animal and robot researchers, so these are two communities that are already in dialogue. Particularly interesting, I think, is that we are finding a variety of social learning capabilities not just in the species most closely related to humans, the anthropoid apes, but in species that are much more distant from us. Especially interesting in this regard are dogs, who in many respects are even more human-like than chimpanzees in their capacity for social interaction and cooperation with us. By studying dogs, and which signals from us they attend to, we can learn a lot about how to design robots to use the same cues.

You have identified two main approaches to artificial moral agents, the top-down approach (one could say a rule-based approach) and the bottom-up approach (which is often seen in connection with genetic programming). How can these to approaches help in building artificial moral agents and where lie their strengths and weaknesses?

A strength of top-down approaches is that the ethical commitments are explicit in the rules. The rules can also be used to explain the decision that was taken. However, it is hard to write rules that are specific enough to be applied unambiguously in all circumstances.

Also, the rules may lead to what we have called a "computational black hole" meaning that it is really impossible to gather and process all

the information that would really be necessary to make a decision according to the rules. Bottom-up approaches, and here I'd include not just genetic algorithms but various kinds of learning techniques, have the strength of being able to adaptively respond and generalize to new situations based on limited information, but when systems become sufficiently complex they have the drawback that it is often unclear why a particular outcome occurred.

To overcome the restraints of both approaches you have suggested merging these two to a hybrid moral system. How can we imagine this?

I believe that we will need systems that continuously engage in a self-evaluative process. We describe it as a virtue-based approach because it has some things in common with Aristotle's ethics. Bottom-up processes form a kind of reactive layer that can be trained to have fundamentally sound responses to moral circumstances. A robot following an instruction by a human must not be completely opportunistic in the means it takes to carry out that instruction, running roughshod over the people for whom it is not directly working.

Rules alone can't capture what's needed. One can't say, for instance, "never borrow a tool without asking" or "never violate a direct order from a human being" for we want agents that are flexible enough to recognize that sometimes it is acceptable, and perhaps even obligatory, to do so. Such decisions are likely to require a lot of context-sensitivity, and for this, a bottom-up approach is best.

However, we will want these same systems to monitor and re-evaluate the outcomes in light of top-down principles. Sometimes one cannot know whether another's welfare is affected or rights violated until well after the fact, but a reflective moral agent, on learning of such an outcome, should endeavor to retrain its reactive processes, or reform its principles. But this is a very hard problem, and is perhaps where the project of artificial moral agents really does slide down the slope into science fiction. But by pointing out that there are reasons to think that neither a top-down or a bottom-up approach will alone be sufficient, we hope to have initiated a debate about how to develop machines that we can trust.

Would this monitor and evaluation system be something like the "ethical governor" which Ronald Arkin proposed in his project on "Governing Lethal Behaviour"?

Overall, there's considerable similarity between our hybrid approach and Arkin's "deliberative/reactive" architecture. However, because his "ethical governor" operates immediately prior to any action being

taken, actually what I have been describing is something closer to his "ethical adaptor" which is another component of his ethical control architecture, and which is responsible for updating the ethical constraints in the system if an after-the-fact evaluation shows that a rule violation occurred. A significant difference between our approach and Arkin's is that the rules themselves (e.g. the Geneva Conventions) are considered to be known and fixed, and not themselves subject to interpretation or revision. This approach is possible because he considers only the case of robots operating in a well-defined battlefield and engaging only with identifiable hostile forces. Arkin believes that in such circumstances, intelligent robots can actually behave more ethically than humans can. Humans get angry or scared and commit war crimes, and Arkin's view is that robots won't have these emotional reactions, although he recognizes that some sort of affective guidance is important.

Besides research and teaching you are also maintaining a blog on the theory and development of artificial moral agents and computational ethics[6], so I guess you will be working on these fields in the future? And which projects are you currently working on?

Right, I'll continue to keep an eye on machine morality issues, but I'm currently being reactive rather than pursuing any new lines of research in this area. My biggest current ongoing project is something completely different – with funding from the U.S. National Endowment for the Humanities we are developing software to help us build and maintain a complete representation of the discipline of philosophy, that we call the Indiana Philosophy Ontology, or InPhO for short[7]. I'm also continuing to work actively on topics in the philosophy of cognitive science, and I'm currently working on papers about the perceptual basis of symbolic reasoning and about the use of structural mathematical models in cognitive science, among other topics.

---

[1] http://www.dhs.gov/xres/programs/gc_1218480185 439.shtm.
[2] e.g. http://www.ai.mit.edu/projects/sociable/movies/affective-intent-narrative.mov.
[3] e.g. http://discovermagazine.com/2005/jul/brain-on-video-games.
[4] e.g. http://www.sciencedaily.com/releases/2008/12/081211081442.htm.
[5] http://robotic.media.mit.edu/projects/robots/mds/overview/overview.html.
[6] http://moralmachines.blogspot.com.
[7] http://inpho.cogs.indiana.edu.

George Bekey: Robots and Ethics

How and why did you get interested in the field of autonomous robots and specifically in military robots?

My interest in robotics developed as a synthesis of a number of technologies I had studied. My PhD thesis was concerned with mathematical models of human operators in control systems, e.g., a pilot controlling an aircraft. The goal was to develop a mathematical representation of the way in which a pilot (or other human operator of a complex system) generates an output command, such as movement of the control stick on the aircraft) in response to changes in the visual input. This work led to increasing interest in human-machine systems. Shortly after completing my graduate studies I developed a hybrid analog-digital computer at a Los Angeles aerospace company. The goal of this project was simulation of the flight of an intercontinental ballistic missile, where the flight control system was represented on the analog portion of the system, and the highly precise generation of the vehicle trajectory was done on the digital computer. These early experiences gave me a great deal of insight into military technology, while at the same time improving my knowledge and skills in computers and control systems. When I joined the University of Southern California in 1962 I continued to work in all these areas. When industrial robots became prominent in the late 1970s it became clear to me that here was a research field which included all my previous experience: human-machine systems, control theory and computers. Further, we hired a young faculty member from Stanford University who had some experience in robots. He urged me to write a proposal to the National Science Foundation to obtain funding for a robot manipulator. I did this and obtained funding for a PUMA industrial robot. From then on, in the 1980s and 90s my students and I worked in robotics, with an increasing interest in mobile robots of various kinds.

You asked specifically about my interest in military robots. As I indicated above, I started working with military systems in the 1960's, but largely left that area for some time. However, when I started looking for funding for robotics research, I found that a large portion of it came from the U.S. Defense Department. While most of the support for my research

came from the US National Science Foundation, during the 1990s I received several large contracts and grants from the Defense Department for work in robotics. While I was pleased to have the funding so that I could support my laboratory and my Ph.D. students, I became increasingly uncomfortable with work on military robots. For many years I had been concerned about the ethical use of technology. This led to participation in a Committee on Robot Ethics of the Robotics and Automation Society, one of main societies forming the professional core of the Institute of Electrical and Electronics Engineers (IEEE), a major international professional organization in the field of electrical engineering.

To summarize: my interest in robotics arose as a way of integrating my various research interests. Military robotics was a major source of research funding, but I was increasingly disturbed by the way robots were being used. Please note that this does not imply a direct criticism of the U.S. military establishment, since I consider myself to be a patriotic person and I believe that countries need a defense structure (since we have not yet learned to solve all disputes among nations by peaceful means). Rather, it represents a desire to contribute to the ethical use of robots, both in military and peacetime applications.

In the recent discussion of military unmanned systems or military robots, it has been argued that especially for future international legislation concerning this matter it would also be necessary to find a universal definition of what constitutes a "robot". How would you define a robot? Should we define robots opposed to intelligent ammunitions and other automated weapon systems or would a broader definition be more useful?

It is interesting that in many of the current discussions of military robots we do not define what we mean by "robot". In my own work I have defined a robot as: *"A machine that senses, thinks and acts"*. This definition implies that a robot:

- Is not a living organism,
- is a physical system situated in the real world, and is not only software residing on a computer,
- uses sensors to receive information from the world
- processes this information using its own computing resources (but these may be special purpose chips, artificial neural networks or other hardware which enable it to make decisions and approximate other aspects of human cognitive functions), and
- it uses actuators to produce some effect upon the world

With this very broad definition it is clear that automated weapon systems *are* robots, to the extent that

they are able to sense the world, process the sensed information and then perform appropriate actions (such as navigation, obstacle avoidance, target recognition, etc). Note that a system may be a robot in some of its functions and not others. Thus, a Predator aircraft is a robot as far as its takeoff, navigation, landing and stability properties are concerned; but it is not a robot with respect to its use as a weapon if the decision to fire and the release of a missile is done under human control. Clearly, if and when the decision to fire is removed from human control and given to the machine, then it would be a "robot" in these actions as well.

It should also be noted that the use of the word *"thinks"* is purposely vague, but it intentionally allows actions ranging from simple YES/NO binary decisions to complex cognitive functions that emulate human intelligence.

I do not believe that it would be useful to separate "intelligent ammunition" and "automated weapon systems" from robots in general. Clearly, there are (and will be more) military robots, household robots, eldercare robots, agricultural robots, and so on. Military robots are robots used for military purposes.

In your academic career you have published more than 200 papers and several books on robotics and over the years you have witnessed euphoria and disillusionment in this field. How would you assess the future of robots in the human society?

I believe that robots are now where personal computers were in the 1980s: they are increasingly evident and will be ubiquitous within the next 10 to 20 years. We already see robots in the home doing vacuuming, grass cutting and swimming pool cleaning. We see autonomous vehicles appearing, both in civilian and military contexts. I anticipate that robots will be so integrated into society that we will no longer think of them as separate systems, any more than we consider an automatic coffee maker a robot. However, there are major challenges to the further development of the field, such as: (1) the need to improve the ways in which robots and humans communicate and interact with each other, which is known as human-robot interaction or HRI, and (2) the need for robots to develop at least a rudimentary form of consciousness or self-awareness to allow for intellectual interaction with humans, and (3) the need to insure that robots behave ethically, whether in health care or home assistance or military assignments. Nevertheless, I believe that the so-called "service robots" which provide assistance in the home or the workplace will continue to proliferate. We have seen great successes in entertainment

robots and in various cleaning robots (for carpets, kitchen floors, swimming pools, roof gutters, etc). On the other hand, there have been some attempted introductions that did not succeed. I can think of two important ones: an automobile fueling robot and a window cleaning robot. Some 10 years ago one of the US oil companies developed a robot for filling the gasoline tank of automobiles. A bar code on the windshield provided information on the location of the filler cap. The hose moved automatically to the cap, filled the tank, and returned to its resting position; the amount of the charge was deducted from the balance of the customer's credit card. After some months of testing, the experiment was abandoned, but I still think it is a good idea. Also, several years ago one of the Fraunhofer Institutes in Germany developed a remarkable robot for cleaning windows in high rise buildings. The machine climbed up on the building using suction cups; it washed the windows as it moved and recycled the dirty water, so there was no spillage on to the sidewalk below. It was able to climb over the aluminium separators between window panes. After some significant publicity, it disappeared from public view, but it may still be available in Germany. These two systems are examples of the robotic innovations which will have a major impact on society, by using machines to replace manual labor. Clearly, such robots will also create major social problems, since all the displaced workers will need to be retrained for jobs requiring higher skills. While there may be objections to such displacements from labor groups, I believe they are inevitable. Another similar area of application lies in agriculture, where current experiments make it clear that robots can perform harvesting, crop spraying and even planting of seedlings; again, low-skilled workers would need training for new jobs.

It has been argued, that for a decision to be ethical, mere rational thought is not sufficient but emotion does play a large role. Apart from the most optimistic prognoses, Artificial Intelligence and therefore robots, will not obtain the full potential of the human mind in the foreseeable future, if at all. However, it is clear, that machines and their programming will become much more sophisticated. Colin Allen and Wendell Wallach have put forward, that machines eventually will obtain a "functional morality", possessing the capacity to assess and respond to moral challenges. How do you think will society respond, if confronted with machines with such a potential?

I basically agree with Allen and Wallach, but let me make a couple of comments. First, your question indicates that emotion plays an

important role in decision making. There is a great deal of research on robot emotions (at Waseda University in Japan and other institutions), i.e., on providing robots with the ability to understand certain emotional responses from their human co-workers and conversely, to exhibit some form of "functional emotion". This implies, for example, that a robot could display functional anger by agitated movements, by raising the pitch of its voice and by refusing to obey certain commands. Similarly a robot could appear sad or happy. Critics have said that these are not "real" emotions, but only mechanical simulations, and that is why I have called them "functional emotions". Clearly, a robot does not have an endocrine system which secretes substances into he bloodstream responsible for "emotional" acts. It does not have a human-like brain, where emotional responses may arise in the amygdala or elsewhere. (An excellent discussion of this matter can be found in the book by Jean-Marc Fellous and Michael A. Arbib, "Who Needs Emotions?: The Brain Meets the Robot"). Hence, I believe that a *functional morality* is certainly possible. However, it should be noted that ethical dilemmas may not be easier for a robot than for a human. Consider, for example, a hypothetical situation in which an intelligent robot has been programmed to obey the "Rules of War", and the "Rules of Engagement" of a particular conflict, which include the mandate to avoid civilian casualties to the greatest extent possible. The robot is instructed by a commanding officer to destroy a house in a given location because it has been learned that a number of dangerous enemy soldiers are housed there. The robot approaches the house and with its ability to see through the walls, interpret sounds, etc. it determines that there are numerous children in the house, in addition to the presumed dangerous persons. It now faces an ethical conflict: Should it obey its commander and destroy the house, or should it disobey since destruction would mean killing innocent children? With contemporary computers, the most likely result of such conflicting instructions will be that the robot's computer will lock up, and the robot will freeze in place.

In response to your direct question about society's response to the existence of robots equipped with such functional morality: I believe that people will not only accept it in robots, but will come to expect it. We tend to expect the best even of our non-robotic machines like our cars: "This car has never let me down in the past…", or we kick and curse our machines as if they intentionally disobey our requests. It would not surprise me if functional machine morality in robots could become a standard for judging human (biological) morality, but I cannot foresee the consequences for society.

Before completing your Ph.D. in engineering you have also studied world religions and you are also teaching courses which cover Hinduism, Buddhism, Taoism, Judaism, Islam, Zoroastrianism and other frameworks at the California Polytechnic State University. Have these experiences and the richness of human thought influenced your perspective on robots and ethics in robotics?

Of course. I believe that studying robots can also teach us a great deal about ourselves and our relationships with other human beings and the physical world. Robots are not yet conscious creatures, but as computing speed increases and we learn more both about the human brain and artificial intelligence, there will be increasingly interesting and complex interactions between humans and robots. These interactions will include the whole range of ethical and other philosophical issues which confront humans in society. My background in world religions has led me to study the ways in which different societies seek to find meaning in their lives, both individually and collectively. I believe that increasingly complex robots will find places in society where interactions with humans will go beyond mere completion of assigned tasks, but will lead to emotional issues involving anger, attachment, jealousy, envy, admiration, and of course, issues of right and wrong, e.g., ethics. I believe that it is possible, in fact likely, that future robots will behave in ways that we may consider "good", e.g., if they help humans or other robots in difficulty by being altruistic; or in ways we may consider "bad", such as taking advantage of others for their own gain. These and other forms of behavior are one of the major concerns of religion. However, religion is also concerned with the spiritual development of human beings; robots are unlikely to be concerned with such matters.

Though for the time being the question of "robot rights" seems far fetched, do you think that in the future this will be an issue human society will have to deal with?

Yes, but I believe that Kurzweil's prediction that robots will demand equal rights before the law by 2019, (and that by 2029 they will claim to be conscious) are somewhat exaggerated. Granted that Kurzweil is indeed a genius and that many of his technical predictions have come true, I think that the question of robot rights involves a number of non-technical social institutions, like law, justice, education and government as well as the cultural and historical background of a particular society. As you know, human beings become emotionally attached to inorganic objects, including automobiles and toys. Clearly, as robots acquire more human-like

qualities (appearance, voice, mannerisms, etc.), human attachments to them will grow. Children become so attached to toys that they attribute human qualities to them and may become emotionally disturbed if the toys are lost or damaged. There are stories that US soldiers in Iraq had become so attached to their Packbots that they become emotionally disturbed if their robot is damaged or destroyed, and they insist on some ceremony to mark its demise. Hence, I believe that indeed robots will acquire some rights, and that society will have to learn to deal with such issues. The more "conscious" and "intelligent" and "human-like" the robots become, the greater will be our attachment to them, and hence our desire to award them some rights normally reserved for humans. Please note that I believe such "rights" may be granted spontaneously by people, and may become tradition and law. I think it is much less likely that the robots will *demand* rights before the law, since this implies a high degree of consciousness which is not likely in the near future.

Your paper (together with Patrick Lin and Keith Abney) "Autonomous Military Robotics: Risk, Ethics and Design"[1] is the first systematically laid out study on this topic which has become known to the general public and is being cited by newspapers all over the world. How did you get involved in this project and did you expect such a resonance?

Actually, Ronald Arkin's work (which has now been published in book form) preceded ours, as did some of the papers of Noel Sharkey and others, although our project had significantly more emphasis on ethics and philosophical issues. As you know from my background, I have been interested in the broader implications of technology from my graduate student days. Several years ago I saw a news item about Patrick Lin who had recently joined Cal Poly's Philosophy Department, describing his interest in ethical issues in nanotechnology and robotics. I contacted him, we wrote a proposal on robot ethics, which was funded by the Office of Naval Research under an umbrella grant to the University after a careful evaluation with many competing proposals. Patrick, Keith and I get along very well, since we have different but complementary backgrounds. We have now submitted a major proposal to the National Science Foundation to study ethical issues involving robots in health care. This study, if it is approved and funded, will be done jointly with Prof. Maja Mataric at USC, and will involve actual robots in her laboratory, used in rehabilitation projects with patients. I am increasingly concerned with the lack of attention to ethical issues

involving robots in health care, and this study will begin to address some of them.

I am glad that you believe there is broad interest in our study. From my point of view, engineers and scientists as a whole are concerned with solving technical problems and answering scientific questions, and they tend to ignore broader social issues. And yet, it is clear that all technology has dual aspects, and may be used for good or evil. This is one of the lessons of ancient philosophies like Taoism, where "good" and "bad" are seen as inseparable aspects of the same reality. In the West we tend to be surprised when something developed for a good purpose is later used to produce harm. Certainly this was the case with Alfred Nobel's invention, and it is true of robotics as well.

You have pointed out, that robotic technology is used increasingly in health care, yet there is no widespread discussion of its ethical impact and consequences. Where do you see the main challenges of the proliferation of robotic technology in the different aspects of human society?

This is a very interesting question. As I indicated in my response to the previous question, one of my continuing concerns is that engineers who design and build robots are not concerned with possible ethical consequences of their inventions. In the past, to insure that no harm resulted from new systems, we incorporated "fail-safe" features. Of course, such design changes may come about only after some damage or destruction has occurred. With industrial robots, fences, enclosures and other protective systems were incorporated only after a robot in Japan malfunctioned and killed a worker. As robots are increasingly integrated in society, both in industry and in the home, the possibility of harmful malfunctions will also increase. Again, I suspect that many of the design features to protect people will not come about until some serious damage is done. There is a common belief in the US that new traffic control signals are not installed at intersections until a child is killed there by an automobile. Now, let us extrapolate this danger to a time in the future, say 20 or 30 or 40 years hence, when robots have been supplied with computers and software systems which enable them to display a level of intelligence and varieties of behavior approaching that of human beings. Clearly, we will not be able to predict all possible unethical actions of such robots. After all, Microsoft is not able to predict all possible malfunctions of a new operating system before it is released, but the consequences of such malfunctions in robots working in close proximity to humans are

much more serious. Consider some questions: Could a robot misinterpret physical punishment of a child as a violent act it must prevent, and thus cause harm to the parent? Or, would be constrained by some new version of Asimov's laws and not be able to defend its owner against a violent intruder, if it is programmed never to injure a human being? If the electricity fails in a home, what would a household robot do to protect its own power supply (and hence, its very existence? Will there be conflicts between robots designed for different functions if they interpret commands in a different way?

So, my answer to your question is that as robots proliferate in society, the potential ethical conflicts will also proliferate, and that we will be ill-prepared to handle them. There will be after-the-fact patches and modifications to the robots, both in hardware and software, since we will be unable to foresee all the possible problems from their deployment. Certainly every new generation of robot designers will attempt to incorporate lessons learned from earlier systems, but like in other systems, they will be constrained by such issues as cost, legal restrictions, tradition, and competition, to say nothing of the difficulty of implementing ethical constraints in hardware and software. We are moving into uncharted waters, so that we cannot predict the main challenges resulting from the introduction of new robots.

---

[1] http://ethics.calpoly.edu/ ONR_report.pdf.

Noel Sharkey:
Moral and Legal Aspects of Military Robots

How and why did you get interested in the field of robots, especially in military robots and their ethical challenges?

I have been working and conducting research and moving around the fields of Psychology, Cognitive Science, Artificial Intelligence, Engineering, Philosophy, Computer Science and Robotics for about 30 years. I am probably best known in the academic world for my work on neural network learning. A big motivation for me has been questions about the nature of mind that started when I was a teenager – I still haven't found the answers. Robotics became a favourite because it is so rich in challenges in a great variety of areas from sensing and control to construction and everything in between.

My background is not in ethics. I have had a private interest in ethical issues such as the treatment of animals, torture and mistreatment of humans, human rights, social justice and equality, and universal rights for children as long as I can remember and always like to dabble in philosophy but not professionally. I have no pretentions to being a moral philosopher and don't even have a coherent moral theory (yet). So it has all been a very sharp learning curve.

Most of my research now gets classed as applied ethics and I would describe myself as an ethical mongrel – a dash of virtue ethics with a bit of duty ethics, a drop of the deontological, and a healthy helping of consequentialism. I have a sense of what I think is fair and just and loot and plunder from the great ethical thinkers of the past. I am not ashamed to admit that I still have an incredible amount to learn.

I came into the area of robot ethics and the ethics of emerging technologies through the backdoor. I gained a high public profile in the UK through involvement in popular BBC TV programmes about robots and also from some major museum robotics projects – doing science in the public eye. This gave me great access to the public and led to a passion for public engagement and to a Research Council fellowship (EPSRC) with a remit to both encourage more young people into science and engineering and to engage with the public about issues of concern within my expertise.

Engagement means not just talking *at* the public but taking their point of view seriously and feeding it back to the appropriate bodies and policy makers and using the media to affect change on their behalf. I am very committed to the idea that senior academics[1] have a responsibility to the public. What I have found so attractive about public dialogue is that I most often learn more from them than they do from me.

My discussions with the public from around the world began to become more about the ethical issues in the application of technology and journalists were beginning to ask me about military robots. (What may seem surprising to some people is that personal conversations with journalists can provide a lot of solid information.) So I began to read all of the US military plans for robots and unmanned systems that I could get hold of.

From my knowledge of the limitations of AI and robotics, this set extreme alarm bells ringing. I was quite shocked by what I read – particularly the push toward autonomous systems applying lethal force. I felt a very strong urge to give priority to reading and writing and letting the public know about the dangers of this area. So I immersed myself in a study of military issues, the laws of war, Just War theory and the Geneva conventions and the various protocols as well as the legal aspects.

This opened the debate considerably for me and led to more focussed discussions and talks with a great number of people including the military themselves. I have been researching and writing both newspaper and journal articles about the issues ever since (as well as about a number of other ethical issues).

Although an increasingly number of people is beginning to express doubts, you are one of the people in the field, who for quite some time have been openly critical about the use of autonomous military systems with lethal potential. How do you see your role in the discussion of unmanned military systems?

I like to see myself as an unelected representative speaking on behalf of the public, to express their concerns and to inform them and policy makers about the issues involved.

I take opportunities to highlight the problems as they arise. Thinking about it now, I guess that there have been five major components to my role:

(i) providing a sanity check on the limitations of what the technology can do and is unlikely to be able to do soon;
(ii) keeping the issues in the public eye through the media and keeping a dialogue flowing;

(iii) discussing the ethical issues with the military;
(iv) bringing the issues to the attention of policy makers and trying to get international discussion going;
(v) keeping abreast of new developments both in the military and in other areas that might be useful to the military; keeping up to date with military plans, calls for proposals and new deployment and updating the public.

A lot of my time is taken up with these activities.

Unmanned military systems, though yet not fully autonomous, are a reality on the battlefields of today. What are the main ethical and legal challenges concerning the present use these of military systems?

There are many ethical issues and challenges facing us with the use of unmanned systems (autonomous or even man in the loop) that I have written about that are too lengthy to repeat here. The most pressing concern is the protection of the lives of innocents regardless of nationality, religious affiliation or ideology. Allowing robots to make decisions about who to kill would fall foul of ethical precepts of a Just War under jus in bello.

In particular armed autonomous robots are against the spirit of the law set down in the Geneva convention under the Principle of Distinction and the Principle of Proportionality. These are two of the cornerstone of Just War Theory.

The principle of distinction is there to protect civilians, wounded soldiers, the sick, the mentally ill, and captives. The law, simply put, is that we must discriminate between combatants and non-combatants and do everything in our power to protect the latter. In a nutshell the ethical problem is that no autonomous robots or artificial intelligence systems have the necessary sensing and reasoning capabilities to discriminate between combatants and innocents. We do not even have a clear definition anywhere in the laws of war as to what a civilian is. The 1949 Geneva Convention requires the use of common sense while the 1977 Protocol 1 essentially defines a civilian in the negative sense as someone who is not a combatant.

There is also the Principle of Proportionality which holds that civilian casualties are often unavoidable in warfare and that the number of civilian deaths should be proportional to the military advantage gained. But there is no objective measure available for a computational system to calculate such proportionality. It is down to a commander's militarily informed opinions and experience. I have written about the big problems of proportionality calculations for humans never mind machines.

Yes, humans do make errors and can behave unethically, but they can be held accountable. Who is to be held responsible for the lethal mishaps of a robot? Certainly not the machine itself. There is a long causal chain associated with robots: the manufacturer, the programmer, the designer, the department of defence, the generals or admirals in charge of the operation, the operators, and so on.

There are a number of ill specified dimensions in the Laws of War about the protection of innocents that are muddied incredibly by insurgent warfare. In history, state actors have often behaved reciprocally – you bomb our civilians and we will bomb yours. This is morally reprehensible but gets worse when we consider non-state actors. Who are their civilians? It is like asking who are the civilians of any arbitrary group such the railway workers or the bakers.

I have recently been thinking through the idea of a proportionality calculation based on a variant of the philosopher John Rawls' "original position", which was for a thought experiment about the principles of justice in a free and fair society. Rawls' notion is that representatives of citizens are placed behind a "veil of ignorance", that deprives them of information about the individuating characteristics of the citizens they represent. This lack of information forces them to be objective about the fairness of the social contract they are attempting to agree upon. Crudely put, it is a little like you cutting a pie knowing that I will have first choice of portion.

My "veil of ignorance for proportionality judgments" would similarly deprive the decision maker of information of the nationality, religion and ideology of the innocents that are likely to be killed. To take an extreme example, a baby in my country has as much right to protection as a baby in a country where insurgents are fighting. Through the veil of ignorance, the baby would expect a better chance of survival.

Ok, so the baby example is a bit emotive, but there is another example I can use taken from a drone strike of a village last year that I have written about elsewhere. The aim of the strike was to kill an al-Qaeda leader; a number of children were among the dead. In a newspaper article, senior US military were reported to say that they knew there was a high risk of killing the children, but the leader was such a "high value" target, it was worthwhile. (Subsequent DNA analysis of the corpses showed that the target had not been present in the village.)

To turn this into a concrete 'veil of ignorance' example, imagine that the commander in charge of the strike had just been informed that a

party of US school children may be visiting the village. Would the calculation of military advantage change?

This is a preview of an idea that I am working on for a paper and it needs more thought and discussion.

It seems unlikely to win the "hearts and minds" of people with military robots. Do you think that – for certain roles – unmanned military systems do have an eligibility in armed conflicts?

First, I think that you are absolutely right about the hearts and minds issue. I have been heartened recently by reports from the new head of the armed forces in Afghanistan, Lt General Stanley McCrystal. He seems to have really grasped the idea that killing civilians means creating many more insurgents and is actually fulfilling their goals for them. He sees that for every innocent killed, a number of their family members will take up arms. I won't take up time with his position here, but it is well worth checking out. It is a pragmatic rather than an ethical approach, but it highly correlates with the ethical and may have more impact.

I have no ethical issues against the use of unmanned systems for protecting soldiers in their normal functioning. Improvised explosive devices on roadsides kill very many soldiers and even the relatively crude robots deployed for disrupting these are of great benefit. I would much prefer to see some of the large budgets that are going into armed predators and reapers being used to develop better explosives detection – detection of explosive at a distance.

There are precedents for weapon systems, which have been banned from the battlefields, either because they lack the ability to discriminate or they cause unnecessary suffering. Could these international treaties act as guidance for how to cope with the questions surrounding the use of unmanned military systems?

Yes, these treaties are useful in setting out guidance. They are not binding and countries can give notice to no longer be signatories. Also not everyone signs up to them. For example China, Russia and the US were not among the 150 countries banning cluster munitions. However, although the US also did not sign up for the landmine treaty, they behave as if they did. These treaties are useful in setting out moral standards.

A similar treaty for unmanned systems or even armed unmanned systems would be much more difficult – at the very least there would be definitional problems. For example is a cruise missile an unmanned system? There are often academic debates about what is considered to

be a robot. I have my own ideas of course, but we will need a consensus. I would very much like to see, at the very least, some serious international debate about the possibility of setting up an unmanned systems arms control treaty. Proliferation now seems inevitable given the military advantages that have recently been showcased.

You could argue that unmanned systems are already covered under the Geneva Convention and the various treaties etc., and this is true in a general sense. But I think that very serious consideration needs to be given specifically to the detailed implications of these new weapons and how they will impact on civilians as they are developed further.

Some argue that robot weapons are just the same as other distance weapons and are just a later stage in the evolution started by the slingshot. I think that robots could be a new species of weapon. As they develop further they could become stand-ins for soldiers or pilots at ever greater distances. Unlike missiles or other projectiles, robots can carry multi-weapon systems into the theatre of operations and act flexibly once in place.

I am currently working on the idea of setting up a Committee for Robot Arms Control and would welcome any supporters of robot arms control reading this to get in touch.

Do you think that concepts to integrate ethical decision making capacities in automated systems, like for example Ronald C. Arkin's "Ethical Governor", will in the end result in systems that can be used in compliance with the laws of armed conflict and/or ethical considerations?

Ron's intentions are good and he has very important things to say. His motivation is based on his concerns about the ethical behaviour of soldiers in battle. He was shocked, like many of us, by the US Surgeon General's report of a survey of US troops in Iraq. He also, like me believes that autonomous armed robots seem to be inevitable. However, I have serious misgivings about his enterprise.

He says that robots don't get angry and will not seek revenge. I agree, but they will also not feel sympathy, empathy, compassion, remorse or guilt. I believe that these are needed for the kinds of moral judgements required in fighting a just war – particularly urban insurgent warfare.

One of the main problems that I see for the ethical governor is the discrimination problem. There is absolutely no point, apart from research purposes, in having a system of rules about ethical behaviour if the input does not tell them the right information to operate with.

We have a side bet running about the timescale for solving the discrimination problem. Ron believes we will have the discrimination technology in operation within the next twenty-five years and I think that he is being overly optimistic. Whichever of us is wrong will buy the other a pint of beer.

Another problem that I have with systems like this in general, is that they are too black and white – too absolute about the rules. The ethical governor is a deontological system. In war we often need consequentialist ethics (with clear moral underpinnings) – there are very many circumstances in war where behaving on the basis of the consequences of an action is more important than blind rule following. The principle of proportionality is intrinsically a consequential problem for a start.

In relation to this last point is that the Geneva Convention and all its associated bit are not written with computer programming in mind. To turn it into "if then rules", will require considerable interpretation.

Soldiers need to do a lot of reasoning about moral appropriateness (even if they are absolutist, they need reasoning to plug into their moral judgements). There are heart-warming reports of troops in the current Middle East conflict responding appropriately in a variety of situations such as letting insurgents pass with a coffin and taking off their helmets as a mark of respect.

It is not just a case of a conditional rule like "if combatant then fire". My worry is that there are a very large, possibly infinite set of exceptions that we could not predict in advance to programme into a computer. I know that current AI systems do not have the required reasoning abilities and I am not sure when or if they will ever have them.

The final problem that I have with such systems (and have to stop myself rambling on forever) is they may be used to push the development of armed autonomous systems with the promise of "don't worry, everything will be OK soon". With a potential (or apparent) technological solution in sight, it may allay political opposition to deployment of autonomous killers.

If armed autonomous robot systems are inevitable, work like this will be needed. In my view the ethical governor will raise more problems than it will solve and that is the only way to make progress. However, a preferable choice for me would be to have the money spent on better ethical training of the troops and more effective monitoring strategies as well as greater accountability.

Concerning not only military applications but all aspects of human

society, from care for the elderly to entertainment, where do you see robots and artificial intelligence in the foreseeable future?

I have written quite a lot about the areas of military, policing, eldercare and companionship, robot nannies and medical robotics, but there are many more ethical issues in other areas of robotics and emerging technologies – these are just the areas that I have thought about in detail.

There is a lot of cash to be made in robotics and they are becoming cheaper to manufacture all the time, and so we could see them entering human society in fairly large numbers soon. It is hard to predict their scope and tasks as there are many creative entrepreneurs and developers.

I am not expecting any great leaps in AI or the ability of robots to think for themselves but I am expecting a lot of very clever applications. Many of these will be welcome and perhaps take away the drudgery of some of our duller work although I don't think they will cause unemployment any more than the computer did.

All trends suggest to me that robots could enter our lives in many ways that we cannot predict – much like the internet and the web did. I do think, though, that there will be many ethical tradeoffs to deal with over the benefits and disadvantages of using robots and I suspect that there will be human rights issues as well. With large commercial interests in play, I guess that the biggest worry is that we will be confronted by very many novel applications before we have had time to think them through.

The keyword "human enhancement". Which kind of new challenges are you expecting in this field?

Wow! This is a very big question in the disguise of a small one. This is not my true area of expertise but I have looked into some of the issues for the UK think tank 2020HealthOrg. Our report will be released as a green paper in the House of Commons this year. At present it is difficult to sort the facts from the hopes, the fantasy and there are large commercial interests at stake.

I am going to be short and general about this one.

One person's enhancement can be another person's alleviation of a serious disability. For that reason, if for none other, there is great hope for brain and nervous system implants, new drugs and drug delivery implants. There is some great work, for example, in overcoming tremors in Parkinson's disease.

These applications have their own ethical issues. One specific to the UK, for example, is about whether the tax payer should foot the bill. The application for illness is not really classed as enhancement, but it is not always easy to draw the line. For example, what if we can enhance the ability to learn, and we use it to bring people with learning difficulties towards the norm. That in itself will, by definition, change the norm and so more people will need to be enhanced and so on.

One of the important ethical concerns is about deceit – the secret use of enhancement to gain advantage (think Olympic doping). For example a device (or drug) may be used to temporarily enhance intelligence to do better on examination, entrance tests or to deceive a potential employer. Let us be clear that legislation is unlikely to stop this practice any more than it stops the use of illegal drugs at present.

Another issue that concerns me is the inequity that could be created between the wealthy and the poor. The wealthy have big enough advantages as it is with the education system. At least now, those with strong analytical skills from a poor background can still work their way into top jobs. Expensive enhancement could see an end to that.

I will finish with one bit of speculation. A big issue that people of the future might have to face is "disenhancement". If we have the technology to enhance people cognitively and physically, we could turn it around to do the opposite. We have all heard of psychiatric drugs being used on political dissidents in the former Soviet Union. Political landscapes can change rapidly as well as treatment of criminals and what constitutes a crime (political or otherwise). We could end up with some very powerful tools to constrain people's thoughts and actions.

I have no doubt that we will be hearing a lot more about the ethical issues associated with implants over the coming years.

[1] I say "senior academics" because it is not a well rewarded career move although that is beginning to change in the UK.

Armin Krishnan: Ethical and Legal Challenges

How and why did you get interested in the field of military robots?

I got interested in military robots more by accident than by design. I was originally specialized in political philosophy and I later became interested in the privatization of warfare, a tendency which seems to fundamentally weaken the institution of the modern nation state, as it is built on the idea of a monopolization of legitimate force within a territory and the suppression of illegitimate violence deployed beyond its borders. Of course, I came across Peter Singer's excellent book on Private Military Firms, which meant for me that I needed to find a slightly different research problem. After looking for some time intensively for a good and original angle, I ended up researching the transformation of defense and national security industries in terms of shifting from a manufacturing based business concept to a services based business concept. The introduction of high-tech weapons, sensors, and communications meant for the armed forces a greater reliance on contractors for a great variety of tasks, most of them, however, related to maintaining and operating technology and not combat. This is not surprising, as mercenaries have always been a marginal phenomenon in military history, apart from some brief and exceptional periods where they prospered and where they could influence the outcome of major wars.

Anyway, when I was doing my research on military privatization and technology I figured that automation is one of biggest trends in the defense sector. Following the invasion in Afghanistan in late 2001 there has been a substantial increase in the use of military robots by the US military. Many defense projects started in the late 1990s, especially in the aerospace field, are relying on automation and robotics. They are aimed at developing systems that are either completely unmanned or are so automated that they require fewer crew members to operate a platform or system. I knew that there had been outlandish efforts by DARPA of building a robot army in the 1980s and that very little came out of it. This was the very stuff of the 1984 *Terminator* movie, which also highlighted public fears that machines could take over, or at least take away our jobs. So four or five years ago I was

observing a growth in the field of military robotics, but I was still very sceptical about the so-called *Revolution in Military Affairs* and military robots. These weapons and systems seemed only able to contribute very little to the military challenges at hand, namely dealing with internal conflicts characterized by guerrilla warfare and terrorism. On the other hand, I realized that it sometimes does not matter whether a particular weapon or technology is effective with regard to dealing with *present* challenges. The lure of new technology is so great that concerns about usefulness can be ignored and that a new weapon or technology will eventually find its own purpose and application. Automation and robotics has proved to be feasible and useful in many other societal contexts and industries. The armed forces cannot be artificially kept at a lower technological level and there are clearly military applications of robotics. I realized that it was only a matter of time before the military will take full advantage of new technologies such as robotics, no matter what. The next logical step was to consider the implications of having military robots fighting our wars. While precision weapons have helped to remove the human operator as far from danger as possible, wars fought by robots would actually mean that no human operators would need to be put at risk at all. This is indeed a very interesting problem from an ethical perspective: what is the justification for using force and for killing other people, who we may regard as our enemies, if this could be done without putting any lives at risk and without sacrifice? Would this be a much more humane way of waging war, or its ultimate perversion? This question kept me thinking for a while and encouraged me to write a book on the topic of the legality and ethicality of autonomous weapons. Unfortunately, I still have not yet found the ultimate answer to this question. Maybe the answer will just lie in what society ultimately decides to do with a technology that is so powerful that it may deprive us of purpose and meaning in the long run, as more and more societal functions are getting automated.

In your recent book "Killer Robots: The Legality and Ethicality of Autonomous Weapons" you explore the ethical and legal challenges of the use of unmanned systems by the military. What would be your main findings?

The legal and ethical issues involved are very complex. I found that the existing legal and moral framework for war as defined by the laws of armed conflict and Just War Theory is utterly unprepared for dealing with many aspects of robotic warfare. I think it would be difficult to argue that robotic or autonomous weapons are already outlawed by international law. What

does international law actually require? It requires that noncombatants are protected and that force is used proportionately and only directed against legitimate targets. Current autonomous weapons are not capable of generally distinguishing between legitimate and illegitimate targets, but does this mean that the technology could not be used discriminatively at all, or that the technology will not improve to an extent that it is as good or even better in deciding which targets to attack than a human? Obviously not. How flawless would the technology be required to work, anyway?

Should we demand a hundred percent accuracy in targeting decisions, which would be absurd only looking at the most recent Western interventions in Kosovo, Afghanistan and Iraq, where large numbers of civilians died as a result of bad human decisions and flawed conventional weapons that are perfectly legal. Could not weapons that are more precise and intelligent than present ones represent a progress in terms of humanizing war? I don't think that there is at the moment any serious legal barrier for armed forces to introduce robotic weapons, even weapons that are highly automated and capable of making own targeting decisions. It would depend on the particular case when they are used to determine whether this particular use violated international law, or not. The development and possession of autonomous weapons is clearly not in principle illegal and more than 40 states are developing such weapons, indicating some confidence that legal issues and concerns could be resolved in some way. More interesting are ethical questions that go beyond the formal legality. For sure, legality is important, but it is not everything. Many things or behaviors that are legal are certainly not ethical.

So one could ask, if autonomous weapons can be legal would it also be ethical to use them in war, even if they were better at making targeting decisions than humans? While the legal debate on military robotics focuses mostly on existing or likely future technological capabilities, the ethical debate should focus on a very different issue, namely the question of fairness and ethical appropriateness. I am aware that "fairness" is not a requirement of the laws of armed conflict and it may seem odd to bring up that point at all. Political and military decision-makers who are primarily concerned about protecting the lives of soldiers they are responsible for clearly do not want a fair fight. This is a completely different matter for the soldiers who are tasked with fighting wars and who have to take lives when necessary. Unless somebody is a psychopath, killing without risk is

psychologically very difficult. Teleoperators of the armed *Predator* UAVs actually seem to suffer from higher levels of stress than jet pilots who fly combat missions. Remote controlling or rather supervising robotic weapons is not a job well suited for humans or a job soldiers would particularly like to do. So why not just leave tactical targeting decisions to an automated system (provided it is reliable enough) and avoid this psychological problem? This brings the problem of emotional disengagement from what is happening on the battlefield and the problem of moral responsibility, which I think is not the same as legal responsibility. Autonomous weapons are devices rather than tools. They are placed on the battlefield and do whatever they are supposed to do (if we are lucky). The soldiers who deploy these weapons are reduced to the role of managers of violence, who will find it difficult to ascribe individual moral responsibility to what these devices do on the battlefield. Even if the devices function perfectly and only kill combatants and only attack legitimate targets, we will not feel ethically very comfortable if the result is a one-sided massacre. Any attack by autonomous weapons that results in death could look like a massacre and could be ethically difficult to justify, even if the target somehow deserved it. No doubt, it will be ethically very challenging to find acceptable roles and missions for military robots, especially for the more autonomous ones. In the worst case, warfare could indeed develop into something in which humans only figure as targets and victims and not as fighters and deciders. In the best case, military robotics could limit violence and fewer people will have to suffer from war and its consequences. In the long term, the use of robots and robotic devices by the military and society will most likely force us to rethink our relationship with the technology we use to achieve our ends. Robots are not ordinary tools, but they have the potential for exhibiting genuine agency and intelligence. At some point soon, society will need to consider the question of what are ethically acceptable uses of robots. Though "robot rights" still look like a fantasy, soldiers and other people working with robots are already responding emotionally to these machines. They bond with them and they sometimes attribute to the robots the ability to suffer. There could be surprising ethical implications and consequences for military uses of robots.

Do you think that using automated weapon systems under the premise of e.g. John Canning's concept (targeting the weapon systems used and not the soldier using it) or concepts like "mobility kill" or "mission kill" (where the primary goal is

to deny the enemy his mission, not to kill him) are ethically practicable ways to reduce the application of lethal force in armed conflicts?

John Canning was not a hundred percent happy with how I represented his argument in my book, so I will try to be more careful in my answer. First of all, I fully agree with John Canning that less than lethal weapons are preferable to lethal weapons and that weapons that target "things" are preferable to weapons that target humans. If it is possible to successfully carry out a military mission without using lethal force, then it should be done in this way.

In any case it is a very good idea to restrict the firepower that autonomous weapons would be allowed to control. The less firepower they control, the less damage they can cause when they malfunction or when they make bad targeting decisions. In an ideal case the weapon would only disarm or temporarily disable human enemies. If we could decide military conflicts in this manner, it would be certainly a great progress in terms of humanizing war. I have no problem with this ideal. Unfortunately, it will probably take a long time before we get anywhere close to this vision. Nonlethal weapons have matured over the last two decades, but they are still not yet considered to be generally a reasonable alternative to lethal weapons in most situations. In conflict zones soldiers still prefer life ammunition to rubber bullets or TASERS since real bullets guarantee an effect and nonlethal weapons don't guarantee to stop an attacker. Pairing nonlethal weapons with robots offers a good comprise, as no lives would be at stake in case nonlethal weapons prove ineffective. On the other hand, it would mean to allow a robot targeting humans in general. It is not very likely that robots will be able to distinguish between a human who is a threat and a human who isn't. It is hard enough for a computer or robot to recognize a human shape – recognizing a human and that this human carries a weapon and is a threat is much more difficult. This means that many innocent civilians, who deserve not to be targeted at all, are likely to be targeted by such a robot. The effects of the nonlethal weapon would need to be very mild in order to make the general targeting of civilians permissible. There are still serious concerns about the long term health effects of the *Active Denial System*, for example.

To restrict autonomous weapons to targeting "things" would offer some way out of the legal dilemma of targeting innocent civilians, which is obviously illegal. If an autonomous weapon can reliably identify a tank or a fighter jet, then I would see no legal problem to allow the weapon to attack targets that are clearly

military. Then again it would depend on the specific situation and the overall likelihood that innocents could be hurt. Destroying military targets requires much more firepower than targeting individuals or civilian objects. More firepower always means greater risk of collateral damage. An ideal scenario for the use of such autonomous weapons would be their use against an armored column approaching through uninhabited terrain. That was a likely scenario for a Soviet attack in the 1980s, but it is a very unlikely scenario in today's world. The adversaries encountered by Western armed forces deployed in Iraq or in Afghanistan tend to use civilian trucks and cars, even horses, rather than tanks or fighter jets. A weapon designed to autonomously attack military "things" is not going to be of much use in such situations. Finally, John Canning proposed a "dial-a-autonomy" function that would allow the weapon to call for help from a human operator in case lethal force is needed. This is some sort of compromise for the dilemma of giving the robot lethal weapons and the ability to target humans with nonlethal weapons and of taking advantage of automation without violating international law. I do not know whether this approach will work in practice, but one can always be hopeful. Most likely weapons of a high autonomy will only be useful in high-intensity conflicts and they will have to control substantial firepower in order to be effective against military targets. Using autonomous weapons amongst civilians, even if they control only nonlethal weapons, does not seem right to me.

In your book you also put the focus on the historical developments of automated weapons. Where do you see the new dimension in modern unmanned systems as opposed to for example intelligent ammunitions like the cruise missile or older teleoperated weapon systems like the "Goliath" tracked mine during the Second World War.

The differences between remotely controlled or purely automated systems and current teleoperated systems like *Predator* are huge. The initial challenge in the development of robotics was to make automatons mechanically work. Automatons were already built in Ancient times, were considerably improved by the genius of Leonardo da Vinci, and were eventually perfected in the late 18th century. Automatons are extremely limited in what they can do and there were not many useful applications for them. Most of the time they were just used as toys or for entertainment. In terms of military application there was the development of the explosive "mine" that could trigger itself, which is nothing but a simple automaton. The torpedo and the "aerial torpedo" developed in the First World

War are also simple automatons that were launched in a certain direction with the hope of destroying something valuable. In principle, the German V1 and V2 do not differ that much from earlier and more primitive automated weapons. With the discovery of electricity and the invention of radio it became possible to remote control weapons, which is an improvement over purely automated weapons in so far as the human element in the weapons system could make the remote controlled weapon more versatile and more intelligent. For sure, remote controlled weapons were no great success during the Second World War and they were therefore largely overlooked by military historians.

A main problem was that the operator had to be in proximity to the weapon and that it was very easy to make the weapon ineffective by cutting the communications link between operator and weapon. Now we have TV control, satellite links and wireless networks that allow an operator to have sufficient situational awareness without any need of being close to the remotely controlled weapon. This works very well, for the moment at least, and this means that many armed forces are interested in acquiring teleoperated systems like *Predator* in greater numbers. The US operates already almost 200 of them. The UK operates two of the heavily armed *Reaper* version of the *Predator* and has several similar types under development. The German Bundeswehr is determined to acquire armed UAVs and currently considers buying the *Predator*. Most of the more modern armed forces around the world are in the stage of introducing such weapons and, as pointed out before, the US already operates substantial numbers of them. The new dimension of *Predator* opposed to the *V1* or *Goliath* is that it combines the strengths of human intelligence with an effective way of operating the weapon without any need of having the operator in close proximity. Technologically speaking the *Predator* is not a major breakthrough, but militarily its success clearly indicates that there are roles in which “robotic” systems can be highly effective and even can exceed the performance of manned systems. The military was never very enthusiastic about using automated and remote controlled system, apart from mine warfare, mainly because it seemed like a very ineffective and costly way for attacking the enemy. Soldiers and manned platforms just perform much better.

This conventional wisdom is now changing. The really big step would be the development of truly autonomous weapons that can make intelligent decisions by themselves and that do not require an operator in order to carry out

their missions. Technology is clearly moving in that direction. For some roles, such as battlespace surveillance, an operator is no longer necessary. A different matter is of course the use of lethal force. Computers are not yet intelligent enough that we could feel confident about sending an armed robot over the hill and hope that the robot will fight effectively on its own while obeying the conventions of war. Certainly, there is a lot of progress in artificial intelligence research, but it will take a long time before autonomous robots can be really useful and effective under the political, legal and ethical constraints under which modern armed forces have to operate. Again introducing autonomous weapons on a larger scale would require a record of success for autonomous weapons that proves the technology works and can be useful. Some cautious steps are taken in that direction by introducing armed sentry robots, which guard borders and other closed off areas. South Korea, for example, has introduced the Samsung Techwin SGR-1 stationary sentry robot, which can operate autonomously and controls lethal weapons. There are many similar systems that are field tested and these will establish a record of performance. If they perform well enough, armed forces and police organizations will be tempted to use them in offensive roles or within cities. If that happened, it would have to be considered a major revolution or discontinuity in the history of warfare and some might argue even in the history of mankind, as Manuel DaLanda has claimed.

Do you think that there is a need for international legislation concerning the development and deployment of unmanned systems? And how could a legal framework of regulations for unmanned systems look like?

The first reflex to a new kind of weapon is to simply outlaw it. The possible consequences of robotic warfare could be similarly serious as those caused by the invention of the nuclear bomb. At that time (especially in the 1940s and 1950s) many scientists and philosophers lobbied for the abolition of nuclear weapons. As it turned out, the emerging nuclear powers were not prepared to do so. The world came several times close to total nuclear war, but we have eventually managed to live with nuclear weapons and there is reasonable hope that their numbers could be reduced to such an extent that nuclear war, if it should happen, would at least no longer threaten the survival of mankind. There are lots of lessons that can be learned from the history of nuclear weapons with respect to the rise of robotic warfare, which might have similar, if not greater repercussions for warfare.

I don't think it is possible to effectively outlaw autonomous weapons completely. The promises of this technology are too great to be ignored by those nations capable of developing and using this technology. Like nuclear weapons autonomous weapons might only indirectly affect the practice of war. Nations might decide to come to rely on robotic weapons for their defense. Many nations will stop having traditional air forces because they are expensive and the roles of manned aircraft can be taken over by land based systems and unmanned systems. I would expect the roles of unmanned systems to be first and foremost defensive. One reason for this is that the technology is not available to make them smart enough for many offensive tasks. The other reason is that genuinely offensive roles for autonomous weapons may not be ethically acceptable. A big question will be how autonomous should robotic systems be allowed to become and how to measure or define this autonomy. Many existing weapons can be turned into robots and their autonomy could be substantially increased by some software update. It might not be as difficult for armed forces to transition to a force structure that incorporates many robotic and automated systems. So it is quite likely that the numbers of unmanned systems will continue to grow and that they will replace lots of soldiers or take over many jobs that still require humans.

At the same time, armed conflicts that are limited internal conflicts will continue to be fought primarily by humans. They will likely remain small scale and low tech. Interstate conflict, should it still occur, will continue to become ever more high-tech and potentially more destructive. Hopefully, politics will become more skilled to avoid these conflicts. All of this has big consequences for the chances of regulating autonomous weapons and for the approaches that could be used. I think it would be most important to restrict autonomous weapons to purely defensive roles. They should only be used in situations and in circumstances when they are not likely to harm innocent civilians. As mentioned before, this makes them unsuitable for low-intensity conflicts. The second most important thing would be to restrict the proliferation of autonomous weapons. At the very least the technology should not become available to authoritarian regimes, which might use it against their own populations, and to non-state actors such as terrorists or private military companies. Finally, efforts should be made to prevent the creation of superintelligent computers that control weapons or other important functions of society and to prevent "doomsday systems" that can automatically retaliate against any attack. These are still

very hypothetical dangers, but it is probably not too soon to put regulatory measures in place, or at least not too soon for having a public and political debate on these dangers.

Nonproliferation of robotic technology to nonstate actors or authoritarian regimes, which I think definitively an essential goal, might be possible for dedicated military systems but seems to be something which might not be easily achieved in general, as already can be seen by the use of unmanned systems by the Hamas. In addition the spread of robot technology in the society in nonmilitary settings will certainly make components widely commercially available. How do you see the international community countering this threat?

Using a UAV for reconnaissance is not something really groundbreaking for Hamas, which is a large paramilitary organization with the necessary resources and political connections. Terrorists could have used remote-controlled model aircraft for terrorist attacks already more than thirty years ago. Apparently the Red Army Fraction wanted to kill the Bavarian politician Franz-Josef Strauß in 1977 with a model aircraft loaded with explosives. This is not a new idea. For sure the technology will become more widely available and maybe future terrorists will become more technically skilled. If somebody really wanted to use model aircraft in that way or to build a simple UAV that is controlled by a GPS signal, it can clearly be done. It is hard to say why terrorists have not used such technology before. Robotic terrorism is still a hypothetical threat rather than a real threat. Once terrorists start using robotic devices for attacks it will certainly be possible to put effective countermeasures in place such as radio jammers. There is a danger that some of the commercial robotic devices that are already on the market or will be on the market soon could be converted into robotic weapons. Again that is possible, but terrorists would need to figure out effective ways of using such devices.

Generally speaking, terrorists tend to be very conservative in their methods and as long as their current methods and tactics "work" they have little reason to use new tactics that require more technical skills and more difficult logistics, unless those new tactics would be much more effective. I don't think that would be already the case. At the same time, it would make sense for governments to require manufacturers of robotic devices to limit the autonomy and uses of these devices, so that they could not be converted easily into weapons. I think from a technical point of view that would be relatively easy to do. National legislation would suffice and it would probably not require

international agreements. To tackle the proliferation of military robotics technology to authoritarian regimes will be much more challenging. Cruise missile technology has proliferated quickly in the 1990s and more than 25 countries can build them. Countries like Russia, Ukraine, China, and Iran have proliferated cruise missile technology and there is little the West can do about it, as cruise missiles are not sufficiently covered by the Missile Technology Control Regime. What would be needed is something like a military robotics control regime and hopefully enough countries would sign up for it.

A lot of people see the problem of discrimination and proportionality as the most pressing challenges concerning the deployment of unmanned systems. Which are the issues you think need to be tackled right now in the field of law of armed combat?

I think most pressing would be to define autonomous weapons under international law and agree on permissible roles and functions for these weapons. What is a military robot or an "autonomous weapon" and under which circumstances should the armed forces be allowed to use them? It will be very difficult to get any international consensus on a definition, as there are different opinions on what a "robot" is or what constitutes "autonomy". At the same time, for any kind of international arms control treaty to work it has to be possible to monitor compliance to the treaty. Otherwise the treaty becomes irrelevant. For example, the Biological and Toxin Weapons Convention of 1972 outlawed biological weapons and any offensive biological weapons research, but included no possibility of monitoring compliance through on-site inspections. As a result, the Soviet Union violated the treaty on massive scale. If we want to constrain the uses and numbers of military robots effectively we really need a definition that allows determining whether or not a nation is in compliance with these rules. If we say teleoperated systems like *Predator* are legal, while autonomous weapons that can select and attack targets by themselves would be illegal, there is a major problem with regard to arms control verification. Arms controllers would most likely need to look very closely at the weapons systems, including the source code for its control system, in order to determine the actual autonomy of the weapon. A weapon like *Predator* could theoretically be transformed from a teleoperated system to an autonomous system through a software upgrade. This might not result in any visible change on the outside. The problem is that no nation would be likely to give arms controllers access to secret military technology. So how can we monitor

compliance? One possibility would be to set upper limits for all military robots of a certain size no matter whether they would be teleoperated or autonomous. This might be the most promising way to go about restricting military robots. Then again, it really depends on how one defines military robots. Under many definitions of robots a cruise missile would be considered a robot, especially as they could be equipped with a target recognition system and AI that allows the missile to select targets by itself. So there is a big question how inclusive or exclusive a definition of "military robot" should be. If it is too inclusive there will never be an international consensus, as nations will find it difficult to agree on limiting or abolishing weapons they already have. If the definition is too exclusive, it will be very easy for nations to circumvent any treaty by developing robotic weapons that would not fall under this definition and would thus be exempted from an arms control treaty.

Another way to go about arms control would be to avoid any broad definition of "military robot" or "autonomous weapon" and just address different types of robotic weapons in a whole series of different arms control agreements. For example, a treaty on armed unmanned aerial vehicles of a certain size, another treaty on armed unmanned land vehicles of a certain size, and so on. This will be even more difficult or at least time consuming to negotiate, as different armed forces will have very different requirements and priorities with regard to acquiring and utilizing each of these unmanned systems categories. Once a workable approach is found in terms of definitions and classifications, it would be crucial to constrain the role of military robots to primarily defensive roles such as guard duty in closed off areas. Offensive robotic weapons such as *Predator* or cruise missiles that are currently teleoperated or programmed to attack a certain area/target, but that have the potential of becoming completely autonomous relatively soon, should be clearly limited in numbers, no matter whether or not they already have to be considered autonomous. At the moment, this is not urgent as there are technological constraints with respect to the overall number of teleoperated systems that can be operated at a given time. In the medium to long-term these constraints could be overcome and it would be important to have an arms control treaty on upper limits for the numbers of offensive unmanned systems that the major military powers would be allowed to have.

Apart from the Missile Technology Control Regime, there seem to be no clear international regulations concerning the use of unmanned systems. What is the relevance of

customary international law, like the Martens Clause, in this case?

Some academics take the position that "autonomous weapons" are already illegal under international law, even if they are not explicitly prohibited, as they go against the spirit of the conventions of war. For example, David Isenberg claims that there has to be a human in the loop in order for military robots to comply with customary international law. In other words, teleoperated weapons are OK, but autonomous weapons are illegal. This looks like a reasonable position to have, but again the devil is in the detail. What does it actually mean that a human is "in the loop" and how do we determine that a human was in the loop post facto?

I already mentioned this problem with respect to arms control. It is also a problem for monitoring the compliance to the *jus in bello*. As the number of unmanned systems grows, the ratio between teleoperators and unmanned systems will change with fewer and fewer humans operating more and more robots at a time. This means most of the time these unmanned systems will make decisions by themselves and humans will only intervene when there are problems. So one can claim that humans remain in the loop, but in reality the role of humans would be reduced to that of supervision and management. Besides there is a military tradition of using self-triggering mines and autonomous weapons have many similarities with mines. Although anti-personnel land mines are outlawed, other types of mines such as sea mines or anti-vehicle mines are not outlawed. I think it is difficult to argue that autonomous weapons should be considered illegal weapons under customary international law. Nations have used remote-controlled and automated weapons before in war and that was never considered to be a war crime in itself.

The bigger issue than the question of the legality of the weapons themselves is their usage in specific circumstances. If a military robot is used for deliberately attacking civilians, it would be clearly a violation of the customs of war. In this case it does not matter that the weapon used was a robot rather than an assault rifle in the hands of a soldier. Using robots for violating human rights and the conventions of war does not change anything with regard to illegality of such practices. At the same time, using an autonomous weapon to attack targets that are not protected by the customs of war does not seem to be in itself to be illegal or run counter the conventions of war. Autonomous weapons would only be illegal if they were completely and inherently incapable of complying with the customs of war. Even then the decision about

the legality of autonomous weapons would be primarily a political decision rather than a legal decision. For example, nuclear weapons are clearly weapons that are not discriminative and that are disproportionate in their effects. They should be considered illegal under customary international law, but we are still far away from outlawing nuclear weapons. The established nuclear powers are still determined to keep sizeable arsenals and some states still seek to acquire them. One could argue that nuclear weapons are just the only exception from the rule because of their tremendous destructive capability that makes them ideal weapons for deterrence. Furthermore, despite the fact that nuclear weapons are not explicitly outlawed there is a big taboo on their use. Indeed, nuclear weapons have never been used since the Second World War. It is possible that in the long run autonomous weapons could go down a very similar path.

The technologically most advanced states are developing autonomous weapons in order to deter potential adversaries. But it is possible that a taboo against their actual usage in war might develop. In military conflicts where the stakes remain relatively low such as in internal wars a convention could develop not to use weapons with a high autonomy, while keeping autonomous weapons ready for possible high-intensity conflicts against major military powers, which have fortunately become far less likely. This is of course just speculation.

Another aspect which has come up in the discussion of automated weapon systems is the locus of responsibility. Who is to be held responsible for whatever actions the weapons systems takes? This may not be a big issue for teleoperated systems but gets more significant the more humans are distanced from "the loop".

Are we talking about legal or moral responsibility? I think there is a difference. The legal responsibility for the use of an autonomous weapon would still need to be defined. Armed forces would need to come up with clear regulations that define autonomous weapons and that restrict their usage. Furthermore, there would need to be clear safety standards for the design of autonomous weapons. The manufacturer would also have to specify the exact limitations of the weapon. The legal responsibility could then be shared between a military commander, who made the decision to deploy an autonomous weapon on the battlefield and the manufacturer, which built the weapon. If something goes wrong one could check whether a commander adhered to the regulations when deploying the system and whether the system itself functioned in the way guaranteed by the

manufacturer. Of course, the technology in autonomous weapons is very complex and it will be technically challenging to make these weapons function in a very predictable fashion, which would be the key to any safety standard. If an autonomous weapon was not sufficiently reliable and predictable, it would be grossly negligent of a government to allow the deployment of such weapons in the first place. With respect to moral responsibility the matter is much more complicated. It would be difficult for individuals to accept any responsibility for actions that do not originate from themselves. There is a big danger that soldiers get morally "disengaged" and that they no longer feel guilty about the loss of life in war once robots decide whom to kill. As a result, more people could end up getting killed, which is a moral problem even if the people killed are perfectly legal targets under international law. The technology could affect our ability to feel compassion for our enemies. Killing has always been psychologically very difficult for the great majority of people and it would be better if it stayed that way. One way to tackle the problem would be to give the robot itself a conscience. However, what is currently discussed as a robot conscience is little more than a system of rules. These rules may work well from an ethical perspective, or they may not work well. In any case such a robot conscience is no substitute for human compassion and ability to feel guilty about wrongdoings. We should be careful with taking that aspect of war away. In particular, there is the argument that bombers carrying nuclear weapons should continue to be manned, as humans will always be very reluctant to pull the trigger and will only do so in extreme circumstances. For a robot pulling the trigger is no problem, as it is just an algorithm that decides and as the robot will always remain ignorant of the moral consequences of that decision.

In addition to the common questions concerning autonomous unmanned systems and discrimination and proportionality you have also emphasized the problem of targeted killing. Indeed, the first weaponized UAVs have been used in exactly this type of operation, e.g. the killing of Abu Ali al-Harithi in Yemen in November 2002. How would you evaluate these operations from a legal perspective?

There are two aspects to targeted killings of terrorists. The first aspect is that lethal military force is used against civilians in circumstances that cannot be defined legally as a military conflict or war. This is in any case legally problematic no matter how targeted killings are carried out. In the past Special Forces have been used for targeted killings of terrorists. So the *Predator* strikes are in this respect not something

new. For example, there has been some debate on the legality of the use of ambushes by the British SAS aimed at killing IRA terrorists. If there was an immediate threat posed by a terrorist and if there were no other ways of arresting the terrorist or of otherwise neutralising the threat, it is legitimate and legal to use lethal force against them. The police are allowed to use lethal force in such circumstances and the military should be allowed to do the same in these circumstances. At the same time, one could question in the specific cases whether lethal action was really necessary. Was there really no way to apprehend certain terrorists and to put them to justice? I seriously doubt that was always the case when lethal action was used against terrorists.

This brings us to the second aspect of the question. I am concerned about using robotic weapons against terrorists mainly because it makes it so easy for the armed forces and intelligence services to kill particular individuals, who may be guilty of serious crimes or not. “Terrorist” is in itself a highly politicised term that has often been applied to any oppositionists and dissenters out of political convenience. Besides it is always difficult to evaluate the threat posed by an individual, who may be a “member” of a terrorist organization or may have contacts to “terrorists”. If we define terrorism as war requiring a military response and if we use robotic weapons to kill terrorists rather than apprehend them, we could see the emergence of a new type of warfare based on assassination of key individuals. Something like that has been tried out during the Vietnam War by the CIA and it was called *Phoenix Program*. The aim was to identify the Vietcong political infrastructure and take it out through arrest or lethal force. In this context 20,000 South Vietnamese were killed. Robotic warfare could take such an approach to a completely new level, especially, if such assassinations could be carried out covertly, for example through weaponized microrobots or highly precise lasers. This would be an extremely worrying future scenario and the West should stop using targeted killings as an approach to counterterrorism.

Where do you see the main challenges concerning unmanned systems in the foreseeable future?

I think the main challenges will be ethical and not technological or political. Technology advances at such a rapid pace that it is difficult to keep up with the many developments in the technology fields that are relevant for military robotics. It is extremely difficult to predict what will be possible in ten or 20 years from now. There will always be surprises in terms of breakthroughs that did not happen and breakthroughs that happened. The best

prediction is that technological progress will not stop and that many technological systems in place today will be replaced by much more capable ones in the future. Looking at what has been achieved in the area of military robotics in the last ten years alone gives a lot of confidence for saying that the military robots of the future will be much more capable than today's. Politics is much slower in responding to rapid technological progress and national armed forces have always tried to resist changes. Breaking with traditions and embracing something as revolutionary as robotics will take many years. On the other hand, military robotics is a revolution that has been already 30 years in the making. Sooner or later politics will push for this revolution to happen. Societies will get used to automation and they will get used to the idea of autonomous weapons. If one considers the speed with which modern societies got accustomed to mobile phones and the Internet, they will surely become similarly quickly accustomed to robotic devices in their everyday lives. It will take some time for the general public to accept the emerging practice of robotic warfare, but it will happen. A completely different matter is the ethical side of military robotics. There are no easy answers and it is not even likely that we will find them any time soon. The problem is that technology and politics will most likely outpace the development of an ethic for robotic warfare or for automation in general. For me that is a big concern. I would hope that more public and academic debate will result in practical ethical solutions to the very complex ethical problem of robotic warfare.

Peter W. Singer: The Future of War

How and why did you get interested in the field of military robots?

I have always been interested in changes in warfare. It is my sense that this field of new technologies might be one of the biggest changes not just in our lifetime, but over the last several years – millennia even we could argue. I first got into it in a sense drawn by two things: First, I have always loved science fiction as a young boy, and robots of course populate that. Second, I was struck by how I kept seeing more and more of these things from science fiction that I had grown up with – robots – popping up in the experience of my friends in the military itself. I recall, for example, talking to a friend in the US Air Force who was fighting in the war in Iraq, but he never left the US. That means he was part of operations using these drones, and it was just very different from the way we understood war.

The same thing you would notice more and more mention of these robotics in civilian industry and in civilian life. For example, I own a robot vacuum cleaner. And, yet the people who study war, who talk about war, were not talking about it, and it was striking at me. I remember going to a conference in Washington DC about what was revolutionary in war today. It had all of the top experts, the well known people, as well as leaders in the military, and yet the word *robot* was never said once. This just did not fit with what was happening there, it did not fit the experience of my friend in the Air Force and it did not fit the raw numbers how we are using these systems more and more.

That is what set me off on this journey to write the book "Wired for War," really to capture just what was happening in this incredible moment in time, who are the people who use these systems in all sorts of different ways, and what are their perspectives on it. But I also wanted to capture the deeper questions. As we start to use more and more robots in war, what would that present to us in areas of ethics, law, public policy? Do they make it more or less likely to go to war, what is their impact on our democracies? So, that was really what I was trying to do – to capture this moment in time.

In your books "Children at War", "Corporate Warriors" and "Wired for War" you have tackled crucial issues in a substantial way. How do

you see the role of the media in these issues and their influence on the general public but also on politics? How do you see your role and can books like yours help to provide a differentiated approach?

What has been striking about each one of those books that I have written is that at the time I started them, that issue was not much in the media; in fact it was not much studied in the research community. I, for example, remember starting out on my journey dealing with "corporate warriors" – private military firms – and I was actually told by a professor, that I had at Harvard, that I should quit graduate school and go to Hollywood and become a screen writer, instead, for thinking to write on such a fiction as private companies operating in war. And, of course, today, this is a multibillion dollar industry; there are more than a hundred thousand of these private military contractors serving in Iraq and another seventy thousand of them serving in Afghanistan.

This, I think, is one of the challenges for those of us in the research role, but it carries over to the media side, which is often reactive, often ex-post, and does not report on a trend that is becoming important until after something bad happens. You can use that same example of the private military industry that I looked at in "Corporate Warriors" and that much of the media reporting of it really does not take off until the 2007-period, most particularly after the shootings involving employees of Blackwater in Nisour Square in Baghdad[1]. We already had well over a hundred thousand of these contractors on the ground, and yet the media was not truly covering it. In fact, there was a study that was done of news stories coming out of Iraq. It found that less than one percent of all these news stories mentioned private military contractors. Now, let us put that in the context: More than half of the soldiers on the ground were private military contractors and yet only one percent was mentioned in news stories. I think this again points to the issue of how the media often is chasing after the news rather than trying to take a step back and figure out what is really happening today. There is also a change in the media, of course, right now, which is that it has become often aimed at servicing the public in a way that is profitable. By that, I mean that it is often not trying to report the news, but rather report the news in a way that will make the public feel good about itself. We see that with the way news networks have become aligned with one partisan political position or the other, the "Fox News Effect," for example, but you see its opposite on the opposite side of the coin. So people turn to media to see news stories that validate their preexisting understandings of the world

around them. That is unfortunate because it does not equip us well to deal with changing circumstances in the world.

Now for myself, for my research obviously, I am drawn to these things that are changing and so I see the role of my books as a way to create a resource book for the media and the public, a book to turn to when these issues emerge in importance. That means whenever that topic comes to the fore, that I have already done the book that lays out the issues, explains the dynamics, and presents some of the questions that people need to wrestle with. I tried to do that on the private military side and on the child soldiers issue. This was also my approach for "Wired for War", given that we have something emerging of great importance, namely the growing use of these robotics, the growing use of them in war. Let us capture that moment, figure out what are some of the key dynamics, meet the various players and also look at the implications of this on various areas that we care about. Then, hopefully, when people start to wrestle with these dilemmas, I have fleshed out a fact-based study to turn to, something that is written in a way that is very accessible.

I think that another challenge of those of us in research is that we often intentionally disconnect ourselves from the public, from the media. We only engage in discourse with each other and the result is that often public policy, as well as often the media, is not all that well informed. It is not just them to blame, but it is us, because we are often speaking only to ourselves. You can see this, for example, in the debates in academic journals, which have become so esoteric at time that I do not even like to read them anymore, although I actually *do* theory and research. I think that presents another challenge to those of us in the field: How to take what we are working on and apply it to real world problems in a way that real world people can understand?

Before we get in medias res of military robots themselves, I would like to ask for your assessment of the impact the new government under President Obama will have on unmanned systems regarding budget, strategy and related fields?

I am obviously biased on this; I was a big supporter of President Obama. In fact I coordinated his defence policy team during the campaign, so take what I am saying here with a grain of salt. There are a couple of indicators to show that we are going to see greater and greater use and purchasing of these systems under the administration of President Obama. The first indicator is that in the defence policy statements that he made during the

campaign itself, he only identified a very limited set of military systems, that he pushed for greater research and investment and understanding of. I believe there were just five of these, and unmanned systems were one of those five. So, out of the entire realm of all the various military weapons and systems, the fact that he said here are the five that I think are important, and that unmanned systems are one of those five is a pretty good indicator. The next indicator is the defence department budget itself: The new one is coming in. The budget itself for the overall US-military is relatively flat and some people predict that in coming years it will decline. However, within that budget, there is one area that is growing and that is unmanned systems. For example, on the aerial side they are retiring several jet fighters such as the F-16. They are retiring them earlier than planned and purchasing more unmanned systems to replace them. The idea is to use the Predator and Reaper drones as a replacement for 250 manned jetfighters.

This is not something though that is just limited to President Obama. You saw this growth take off during the period of President Bush: For example, when we went into Iraq we had just a handful of these drones in the US-Military inventory, and by the end of 2008 we had more than 7,000. Now, under the new budget, we are going to continue to add to that. The point here is, this is not a system, this is not a technology, that you can describe as partisan, as one President being involved in and another not being. This is a sea-change in war itself. These are systems that are being used in greater and greater numbers and they are not going away regardless who the president is; it is a global technology shift. And the parallels that people make to this in history are very instructive ones. Bill Gates, the founder of Microsoft, for example, described that robotics are right now where the computer was in 1980. It is poised for a breakup and for a takeoff to the extent that very soon we will not call them robots any more. The same way we have computers all around us, but we do not call them computers. In my car, for example, there are more than a hundred computers, but I do not call it a "computer car"; I have a computer in my kitchen, but I call it a "microwave-oven." The point is, if that is a parallel, we would not describe the computer as being democrat or republican; it was a new technology and the same thing is happening with robotics today and their use in war.

The technization of the military (unmanned systems, surveillance, precision ammunition), models like the Future Combat Systems (the soldier as one system of systems)

and new concepts of using private military contractors have changed the role and the (self-)image of the army and the soldiers in the last decade. How is your perspective on this fundamental change?

This is a fantastic question. It cuts to one of the biggest issues, the biggest changing dynamics at play of warfare today and maybe even overall history. Think about our image of the warrior. If we imagine a warrior, if we imagine a soldier, there is a certain image that comes into our mind. It is most likely a man. They are most likely wearing a uniform. If they are wearing a uniform, it means they are probably part of the military. If they are part of the military, of course they are serving for that nation. And what motivates that service? Patriotism. Why is that military sent into war? Because of politics, because it is linked to the nation state.

That is our image, our understanding, our assumption of the warrior. And yet compare it to what is actually taking place. It is not just men, it is of course women, but it is also children (more than 10% of the combatants in the world are under the age of 18; many as young as 5 years old), and it is also increasingly not human. The US Military for example has 7,000 drones in the air and another 12,000 unmanned ground vehicles. The organisations that they fight in are not just militaries. In fact, look at the experiences of the US Military and NATO in places like Afghanistan and Iraq. Who are they fighting against? They are fighting against warlords, terrorists, insurgents, drug cartels. Look at who is fighting on their behalf: the "coalition of the willing", that President Bush supposedly built to fight in Iraq, actually had far more private military contractors than they had troops from others state allies. So if we are being honest, we simply had not a "coalition of the willing"[2], but a "coalition of the billing", the rise of this private military industry, which does not seem to be going away.

Then you look at the motivations: a soldier serves, he is motivated by patriotism. He goes to war because of politics and national interest. But there are other motivations at play now for other actors. So a contractor, for example, does not serve; he works, he carries out a contract. The motivations for why someone might go to war can be anything from their personal profit for a contractor; it might be for religious reasons, if we look at some of the various radical groups out there; it might be because they were forced into it, such as young child soldiers. And, of course, the motivations for the organisation itself are very different. Name me one war right now that is just about politics, where national interest is the sole driver in terms of the political level. Wars are driven by anything from politics to

religion, to economics at the organisational level. But also at the micro-level, they are driven by ethnicity, society etc. There is not this clear-cut assumption that we have of war, and I think this is one of the changes of the 21st century, understanding that it is much more complex out there than our assumptions.

Would you think there is a need for additional national and international legislation on the deployment and development of military robots? And is there a plausible possibility of international treaties regarding this matter?

I think there is very much a need for a look at the legal issues as well as the ethical issues that surround this entire new technology. And, again, think of the parallels that people make to this revolution. Some people describe that it is akin to the rise of the computer, other people note that it is just about parallel to when automobiles were first introduced; they make the parallel that it is about 1908. Some other people say, 'You know it is equivalent to the invention of the atomic bomb and that it is something that can both change warfare but maybe we might later on determine that we ought not to have built it'. That is what a lot of the scientists that I interviewed for the book discussed. The point here is this: each of these parallels are ones where we realize that we do need to create a regulatory environment around it, a sense of accountability around it, a debate about what are the laws, what are the ethics, what is the right and wrong that surrounds this. And this is true of any new weapon and almost any new technology, as they create new questions to figure out. And these questions, these legal questions, can have a huge impact.

I'll give you an example from history, a parallel, that I think of. Before World War One, there were a number of technologies that just seemed like science fiction; in fact they were only talked about in science fiction, for example the airplane, the tank, the submarine. In 1914, Arthur Conan Doyle, who was the creator of Sherlock Holmes, wrote a short story about the use of submarines to blockade Great Britain[3]. It was a science fiction story. The British Admiralty, the British Royal Navy actually went public to mock Arthur Conan Doyle's vision of the idea of using this new technology in war this way. They mocked it not because of operational reasons, but because of legal reasons. They said that no nation would use submarines to blockade civilian shipping, and if any submarine did, its officer would be shot by his own nation for committing this kind of crime. Well ,of course, just a couple of months later, World War One begins and the German Navy starts a submarine blockade of

Great Britain, just along the lines that Arthur Conan Doyle had predicted using this new technology. Now what is interesting is not that it just happened, but also it was a debate and a dispute over the legality of this, i.e. how to use these new technologies in this way. That is actually what helped draw the United States into that war. There was a dispute over the right and wrong of attacking civilian shipping using this new technology, the submarine. And the dispute over it is part of why the United States entered the war, because it took a very different view than of course Germany had during this period. That debate was also part of the US becoming a global superpower. So, my point is that these questions of right and wrong can have a huge impact.

Now when it comes to robotics in war, there are all sorts of different legal questions that we have got to wrestle with: Who should be allowed to build them? What are the parameters in terms of what you can put on them? How autonomous can they be? Can they be armed or not? Who can utilize them; are they just something which should be just limited to the state? Which states? Are they something that can be utilized by non-state actors, and which non-state actors? Are we comfortable with, for example, private military companies using them; are we comfortable with non-state actors like the Hezbollah having them? – Well, you know what, too late, they already have them. Another example: Can they be utilized by governments for other functions, such as policing? – Well, guess what, too late, they are already starting to be utilized in these roles; you have police departments in places like Los Angeles or Vancouver in Canada that have been exploring drones for their use. How about individuals, should they be allowed to have armed robots? Is that my 2nd amendment constitutional right as an American[4]?

My point is this: It may sound like very silly science fiction, but these questions are very real ones that we have to flesh out. Unfortunately, these questions of right and wrong, this ideal of legislation, of legality, really is not being wrestled with all that much. You certainly cannot find any legislation about it at the national level. The closest you come is in Japan, where there are safety limitations on certain commercial robots, and the reason for it had nothing to do with war. It was that at a robotics convention, where companies were showing their latest systems, the organizer got worried about a robot running someone over, and that was the point of it. It was a sort of personal safety thing that had to do with liability.

You have a similar problem at the international level. One of the things

I talk about in the book is a meeting with folks at the International Red Cross, which is an organization, that has done so much for international law, basically the sort of godparents of international law itself. And yet when it comes to robotics, when it comes to unmanned systems, they say, 'You know what, there is so much bad going on in the world today, we cannot waste time with something like that.' It is a valid answer from one perspective; there are a lot of bad things going on in the world, be it the genocide in Darfur to human rights problems around the world, you name it. And so why would you want to waste time – so to speak – on this new technology. But the problem is that you could have said the same exact thing about that submarine back in 1914 or you also could have said the same thing about that crazy invention of using radioactive materials to create a bomb. There were so many bad things happening during World War Two; why should people wrestle with the right and wrong of this new weapon? The point of this, and this is what concerns me, is that our track record is usually waiting for the bad thing to happen first and that is also for those who deal with the law side of both the national and the international level. So, we did not start to wrestle with the implications of atomic bombs until it was, in a sense, to late. And then we have 40 years of arms control movement trying to roll that back, and we are still not there yet. It is the same thing I worry a little bit about the robotics side: Unless we start a dialog about it, we are going to play catch-up for the long term.

Military robots are a reality on the modern battlefield, and many nations beside the United States have begun ambitious projects in military robotics. Do you see the danger of a new arms race?

This revolution – this robotics revolution – is not merely an American revolution; and this is one of the, perhaps, biggest misunderstandings among those from other countries, particularly from Europe, who wrestle with these issues. They often look at the American use of this and say, "Gosh, that's the Americans again using their toys, using their technology" And, then you also see an angle of coverage on the drones' strikes into Pakistan for example saying this is just prototypically American. It is just fundamentally wrong. And by that I mean that there are 43 other countries working on using military robotics today. They range from large countries like the United Kingdom, France, Germany, Russia, and China to smaller countries like Pakistan, Iran, and Belarus. This is not a revolution that is going to be limited to anyone nation.

It is not going to be limited to just states themselves. Again, non-state

actors of a whole variety of different types have utilized these unmanned systems. It is everything from Hezbollah, which flew drones against Israel during its recent war, to one of the groups in the book, a group of college kids in Pennsylvania. They negotiated with a private military company for the rental of a set of military grade drones, that they wanted to deploy to Sudan. These were college kids starting to use that advanced military system.

This globalization leads to what I view as almost a flattening of the realm of war and the technologies that are used in it. By that I mean we are seeing warfare go the same way that software has gone. It is going "open source." The most advanced technologies are not just limited to the big boys. All actors can now buy them, build them, use them. The same way it is played out for software. And that is happening in warfare as well.

Now, a concern for states is of course how do they keep up with this trend and how do they limit it and does it lead to just a quickening and the potential risk of an arms race. It is also, I think, a concern for some of the western states, in particular for the US in this trend and that they are ahead right now but that is not always the case. There is a lesson in both technology and war: there is no such thing as a permanent first mover advantage. Think about this in technology: It was companies like IBM, Commodore, Wang that were the early movers in the computer realm. And, yet, they are not the dominant players anymore. It is now companies like, for example, Microsoft or Google or Apple. So being first did not mean that you came out on top in the end.

The same thing has happened in war. For example, it was the British who invented the tank. It was the Germans who figured out how to use the tank better. And the question for the US and its partners in Western Europe is, where does the state of their manufacturing today as well as the state of their science and mathematics and engineering training in their schools have them headed? That is, where does the current trajectory of these important underliers have them headed in this revolution? Or another way of phrasing it is: What does it mean to be using more and more soldiers whose hardware is increasingly built in China and whose software is increasingly being written in India? Where does that have you headed? So it is not just a concept of an arms race, but in fact will some of the players in that race find it sustainable for themselves?

For your book you have spoken with many soldiers. How is your estimate on the influence of military robots on the soldiers using them

(individualization and anthropomorphism of robots is something which comes to mind but also the psychological stress of UAV remote operators)?

For that book I made a journey of meeting with everyone, from people who design robots to the science fiction authors who influenced them; from the soldiers who use them on the ground and fly them from afar to the generals who command them; from the insurgents that they fight to the news journalist who cover them; add to this the ethicists and human rights lawyers, who wrestle with the right and wrong of it. These are all the type of people that I interviewed for the book. One of the most important findings and one of the things that was fascinating to me, is that all of the ripple effects of this new technology, all the things that are important about robots' impact on our real world do not come back to the machine, but come back to human psychology. It is all about us and how we view and understand the world around us and how these technologies help reshape that – that is the important part of the discussion.

I think we can see this, for example, on the soldiers themselves. We are seeing this going lot of different directions. One is of course the distancing effect, the change of what it means to be fighting from afar, fighting by remote. It has taken that phrase "going to war" and given it an entirely new fundamental meaning. For the last 5,000 years, when we described somebody as going to war – whether we are talking about the ancient Greeks going to war against Troy or my grandfather going to war against the Japanese in the Pacific during World War Two –, we were at a most fundamental level talking about going to a place where there was such danger that that soldiers might never come home again, that they might never see their family again. That is what going to war has meant for the last 5,000 years… until now.

One of the people I remember meeting with was a US Air Force Predator drone pilot, who fought against insurgents in Iraq but never left Nevada. He talked about what it was like to go to war in this case, where he described how he would wake up in the morning, drive into work, for twelve hours he would be putting missiles on targets, killing enemy combatants, and then at the end of the day, he would get back in the car and he would drive home. And 20 minutes after he had been at war, he would be at his dinner table talking to his kids about their school work.

And so we have this entire new experience of war of being at home and simultaneously at war. And that is creating some psychological

challenges for those who fight from afar. They found for example that many of these remote warriors were suffering from levels of combat stress equal or in some cases even greater than some of the units physically in Iraq and Afghanistan. It is very early, we are still learning about this, and as one military doctor put it, 'We have 5,000 years of understanding normal combat stress, but we only have a couple of years understanding this entire new model.' But there is a couple of drivers that we believe: One is that the human mind is not set up for this sort of dual experience of being at war and being at home and going from killing someone to then having your wife be upset at you because you were late for your son's football practice. People are having that experience right now. Another is the grinding nature of the remote work. These units may be fighting from afar, but they are doing it day after day after day and, in fact, doing it for years, and they do not get weekends off, they do not get holidays off, because war, of course, does not play that way. And so they do not deploy in and out the way that soldiers have traditionally done. Therefore, it can be quite grinding.

The other aspect that people point to is the change in camaraderie: It is tradition that soldiers who deployed together and have experienced the battle together then have also gone through the sort of psychological management of those stressors together. Air Force officers for example talk about flying out on mission, but then after the mission is done going to "beer call." It is basically that they sit down, the squadron, they have a beer and they get out all the emotions they just had to go through, for example from losing one of their buddies. In the remote warrior work, you do not have a "battle buddy" as they put it. You are sitting behind a computer screen, you are experiencing these aspects of war, but you are never sharing it; and then you clock out and you go home. And so the unit is never together, never has that rest and recovery period.

The final part of it is that while you are fighting remotely in many ways they are seeing more of war than recent generations have. For example a bomber pilot will fly in, they will drop the bomb and they will fly away. Drone pilots will do the same, remotely, but unlike that man bomber pilot, they will see the target up close beforehand using the high-powered video cameras. They will see that target for minutes, in some cases for hours, in some cases for days, as they watch it develop out. And then they will drop the bomb and they will see the effects of it afterwards. That means that war may be happening at a distance, but it is very much in their face. Then of course again, they go home

and they are talking to their kids 20 minutes later.

This stressor can also be one that plays out for their fellow troops. I remember talking to an US Air Force NCO. He described how dramatic it was when they were operating an unarmed drone that was flying above a set of US soldiers that were killed in a battle. And they could only fly above them and watch as these soldiers were killed in front of them. You can imagine just how dramatic that is, that sense of helplessness, and then to walk outside the control command centre, where you can go to the grocery store. America is at peace, but you have just seen people die. You have just seen fellow soldiers die. And so this is one of the remarkable challenges.

It is interesting though, we are seeing other connections, other bonds being built, though in strange new ways. For example, while we are seeing this disconnect from soldiers fighting from afar and the new experiences they are having, other soldiers are bonding with their robots themselves. One of the stories that opens the book is about a US military unit that has their robot killed – it is blown up by a roadside bomb. It literally sends the unit into a deep moral spiral, and the commander of the unit writes a condolence letter back to the manufacturer, the same way he would have written a condolence letter to someone's mother back in the day.

There is another case in the book about a soldier who brings in his damaged robot to the robot hospital – again they call it the robot hospital even though it is just a repair yard, a garage. And he is crying as he carries this robot in, and the repairmen look at him and they say, 'We can't fix it, it's completely blown up but don't worry we can get you another robot.' And he says, 'I don't want another robot, I want this one. I want Scooby Doo back.' It sounds silly, it sounds absurd, but the thing is, he took this to heart, he bonded with this robot because that robot had saved his life countless times, again and again. And so why would he not start to bond with it?

We have seen other cases of course naming them, giving them ranks, taking risks for the robots in a way that they really should not, when we pull back and think about it. There was one incident where a robot was stuck and a soldier in Iraq ran out 50 meters under heavy machinegun fire to rescue his robot. The whole point of us using robots in war is to limit risks, and yet here he was taking far greater risk to rescue it.

It may actually turn again on our psychology and even our brain physiology. One of the interesting things is that they did a study of

human brains. They linked them up to a monitor and they found – there is a part of the brain called the mirror neuron – that the mirror neuron fires when you see something that you believe is alive. So every time you see a dog, an insect, a fellow person, that part of your brain, that mirror neuron, that nerve cell fires. What was interesting in the study is, when they showed these people robots and things that they knew were machines, they knew they were not alive, the mirror neurons in their brains still fired. And so it may just be that we cannot help ourselves, we cannot help but attach our human psychology to these mechanical creations.

What impact will military robots have on the future of warfare itself? And what will we have to expect in unconventional/ asymmetric warfare and terrorism?

These technologies, these systems *are* the future of war. That is the growth curve of their usage is the same growth curve that we saw with the use of gunpowder, the use of machineguns, the introduction of airplanes and tanks, where they were used in small instances often not all that effective at the start and then we began to use them more and more in lots of different ways and they began to globalize. And soon something that was once seen as abnormal was now the new normal. And that is taking place with robotics today. We may think of them as just science fiction but they are battlefield reality. And we are only seeing their use grow and grow. For example the US military has gone again from a handful of these drones in the air to 7,000 in the air, from zero on the ground to 12,000 on the ground all in just the last five years. But this is just the start. One US Air Force three-star general I met said, we very soon will be using "tens of thousands" of robots. Again, it will not just be the US Air Force, it is all of the various militaries out there. You have got 43 other countries building and using these systems and everybody wants more of them. It is the future of war, like it or not.

It is also the future of war for non-state actors. As I discussed earlier, it has that flattening effect of allowing more and more players to use high technologies. So, it is not like in the past where the tools of war were limited just to states, just to governments, just to militaries; this is not the Napoleonic age anymore; now, all the different players can use it. The implications of that for terrorism are of concern, because it means that small groups and individuals will have the lethality of the state. I think we can see this on another impact: it widens the scope of those who can play in the realm of terrorism. That means it is not just that Al-Qaeda 2.0 or the next generation version of the Una-

bomber[5] or a Timothy McVeigh[6] is going to be more lethal with greater distance. For instance there was a group of model plane hobbyists who flew a drone from the United States to Europe – well, one person's hobby can be another person's terrorist operation.

In fact, a recent government report said that the next generation of IEDs – the next generation of these improvised explosive devices that have been so deadly in Iraq and in Afghanistan – are going to be aerial ones, small drones that carry these explosives. But it is not just again their greater lethality; it is the fact that more and more people can play in these roles. You no longer have to be suicidal to have the impact of a suicide bomber. You can utilize the robot to carry out missions, to take risk that previously you had to be suicidal to do. And one of the people I interviewed was a scientist for the US military's DARPA institution (our advanced research lab) and his quote was this: "If you give me 50,000 Dollars and I wanted to, I could shut down New York City right now using robotics." That is a remarkable illustration of the technology itself, but also of the world that we are entering, and then finally how so much of whether it is a good or an evil again depends on us. It is not the technology that is the most important part; it is his willingness or not to utilize that technology that way.

Consequently, when I pull back and think about these technologies, I often go to how I close the book, which is this question: We have built these incredible technologies, we have built these incredible systems that can do remarkable things. They are truly cutting edge. And yet, what does it say about us; that is, we are building technologies that both scientists as well as science fiction authors believe may even be an entirely new species, but we are only doing it to make ourselves more lethal, to give us greater capability to kill each other. Therefore, the ultimate question is this: Is it our machines that are wired for war or is it us?

---

[1] On September 16, 2007, Blackwater guards shot and killed 17 Iraqi civilians in Nisour Square, Baghdad. The incident occurred while Blackwater personnel were escorting a convoy of U.S. State Department vehicles. The next day, Blackwater's license to operate in Iraq was revoked.

[2] The term "coalition of the willing" has been used by George W. Bush to refer to the countries who supported the 2003 invasion of Iraq.

[3] Arthur Conan Doyle, Danger! Being the Log of Captain John Sirius in: The Strand Magazine, July 1914.

[4] The Second Amendment to the United States Constitution is the part of the United States Bill of Rights that protects a right to keep and bear arms.

[5] Theodore John Kaczynski, known as the Unabomber, carried out a campaign of mail bombings in the United States from 1978 to 1995, killing three people and injuring 23.

Kaczynski is serving a life sentence without the possibility of parole.

[6] Timothy James McVeigh was responsible for the bombing of the Alfred P. Murrah Building in Oklahoma City on April 19, 1995. The bombing killed 168 people, and was the deadliest act of terrorism within the United States prior to September 11, 2001. He was sentenced to death and executed on June 11, 2001.

Robert Sparrow:
The Ethical Challenges of Military Robots

How and why did you become interested in the field of military ethics and, in particular, the field of military robots?

I've been interested in military ethics ever since I first started studying philosophy at the age of 17. I've always thought that questions of political philosophy are the most urgent philosophical questions because they relate to the way we should live alongside each other. Questions of military ethics – or at least of Just War theory – have been some of the most controversial political questions in Australia, given the Australian government's tendency to follow the United States into various wars around the globe despite the absence, in most cases, of any direct threat to Australia. So I have always been interested in Just War theory insofar as it provided me with the tools to think about the justification of these wars.

I became interested in the ethics of military robotics via a more roundabout route. I originally started writing about ethical issues to do with (hypothetical) artificial intelligences as an exercise in applying some novel arguments in moral psychology. Similarly, I wrote a paper about the ethics of manufacturing robot pets such as Sony's Aibo in order to explore some issues in virtue ethics and the ethics of representation. However, in the course of writing about robot pets I began reading up on contemporary robotics and became aware of just how much robotics research was funded by the military. So I wrote my paper, "Killer Robots", partly – like the earlier papers – as a way of investigating the relationship between moral responsibility and embodiment, but also because I thought there was a real danger that the development of military robots might blur the responsibility for killing to the point where no one could be held responsible for particular deaths. Since then, of course, with the development and apparent success of Global Hawk and Predator, robotic weapons have really taken off (pardon the pun!) so that issues that even 10 years ago looked like science fiction are now urgent policy questions. Consequently, my current research is much more focused on responding to what we know about how these weapons are used today.

The United States' Army's Future Combat System is probably the

most ambitious project for fielding a hybrid force of soldiers and unmanned systems to date. From a general perspective, what are your thoughts on the development and deployment of unmanned systems by the military?

In a way, I think the current enthusiasm for military robotics is a reflection of the success of anti-war movements in making it more difficult for governments to sustain public support for war once soldiers start coming home in body bags. I suspect that governments and generals look at unmanned systems and see the possibility of being able to conduct wars abroad over long periods without needing to worry about losing political support at home. So the desire to send robots to fight is a perverse consequence of the triumph of humanist values. The extent to which this development has occurred at the cost of concern for the lives of the citizens of the countries in which these wars are fought is an indication of the limited nature of that triumph.

At the same time, of course, it's entirely appropriate and indeed admirable that the people in charge of weapons research and procurement should be concerned to preserve the lives of the men and women that governments send into combat. Unmanned systems clearly have a valuable role to play in this regard and it would be a mistake to downplay this. It is difficult to see how there could be anything wrong with the use of robots to neutralise IEDs or clear minefields, for instance.

I also think there is a certain "gee whiz" around robot weapons that is responsible for much of the enthusiasm for them at the moment. Certainly, it's easier to get the public excited about a military robot than about human beings fulfilling similar roles. And I suspect this is even true *within* some parts of the military-industrial complex. Defence ministers want to be able to claim that their country has the most "advanced" weapons, even where the new weapons don't perform that differently from the old. Spending money on military equipment puts more money in the pockets of the corporations that provide campaign funding than does spending money on personnel, which works to the advantage of the robots. It's also worth remembering that there is often an enormous gap between what arms manufacturers claim a system will be capable of when it is commissioned and what they actually deliver. This is especially the case with robots. The PowerPoint presentations and promotional videos in which the systems function flawlessly are often a far cry from the reality of how they work in chaotic environments. However, it is surprising how influential the PowerPoint presentations seem to

be when it comes to determining which systems are funded.

Finally, even if systems do function reliably, it is possible they will be much less useful than their designers intend. One suspects that, in the not-too-distant future, there will be a re-evaluation of the usefulness of military robots, with people realising they are a good solution only in a very limited range of circumstances. To a person with a hammer, everything looks like a nail, so when militaries possess unmanned systems they will tend to want to use them. Yet there is more to war than blowing people up. It's pretty clear that the Predator is precisely the *wrong* weapon to use to try to "win" the war in Afghanistan, for instance. Insofar as anyone has any idea about what it would mean to win this war, it would involve winning the "hearts and minds" of Afghanis to the West's cause and creating conditions that might allow Afghanis to govern themselves and to live free of poverty and fear. No amount of destroying "high-value targets" from 16,000 feet will accomplish this. Indeed, it seems probable that the civilian casualties associated with Predator strikes radically decrease popular support in Afghanistan for Western goals there. As David Kilcullen and Andrew Mcdonald Exum pointed out in a recent *New York Times* opinion piece, missile strikes from Predator are a tactic substituting for a strategy. There are features of unmanned systems that encourage this – the "gee whiz" nature of what they can do and the fact that they don't place warfighters' lives in jeopardy.

What would you say are currently the most important ethical issues regarding the deployment and development of military robots?

Last time I counted, I had identified at least 23 distinct ethical issues to do with the use of robotic weapons – so we could talk about the ethics for a long time ... To my mind, the *most* important issue is the ethics of what Yale philosopher, Paul Kahn, has described as "riskless warfare". If you watch footage of UAVs in action it looks a lot like shooting fish in a barrel. The operators observe people in Iraq or Afghanistan, make a decision that they are the enemy, and then "boom" – they die. The operators are never in any danger, need no (physical) courage, and kill at the push of a button. It is hard *not* to wonder about the ethics of killing in these circumstances. What makes the particular men and women in the sights of the Predator legitimate targets and others not? Traditionally, one could say that enemy combatants were legitimate targets of our troops because they were a threat to them. Even enemy soldiers who were sleeping might wake up the next morning and set about

attacking you. Yet once you take all of our troops out of the firing line and replace them with robots remotely operated from thousands of kilometres away, then it is far from clear that enemy combatants pose any threat to our warfighters at all. Armed members of the Taliban might *want* to kill us but that may not distinguish them from their non-combatant supporters.

Kahn has suggested that when the enemy no longer poses any threat, we need to move from "war" to "policing", with the justification for targeting particular individuals shifting from the distinction between combatants and non-combatants to the question of whether particular individuals are involved in war crimes at the time. I'm not sure the notion of "threat" does all the work Kahn's argument requires, because, as the legitimacy of targeting sleeping combatants suggests, even in ordinary warfare the enemy is often only a hypothetical or counterfactual threat. Nonetheless, there does seem to be *something* different about the case in which the enemy has only the desire and not the capacity to threaten us and some of my current research is directed to trying to sort out just what the difference is.

After that, there are obvious concerns about whether unmanned systems might lower the threshold of conflict by encouraging governments to think that they can go to war without taking casualties, or by making accidental conflict more likely. There are also some interesting questions about what happens to military culture and the "warrior virtues" when warfighters no longer need to be brave or physically fit. Finally, there is an important and challenging set of issues that are likely to arise as more and more decision making responsibility about targeting and weapon release is handed over to the robots. At the moment, systems rely upon having human beings "in the loop" but this is unlikely to remain the case for too much longer; in the longer term, systems that can operate without a human controller will be more deadly and survivable than those that rely upon a link to a human controller. Eventually we will see an "arms race to autonomy" wherein control of the weapons will be handed over to on-board expert systems or artificial intelligences. A whole other set of ethical issues will arise at that point.

In passing, I might mention that one of the objections people raise most often about robot weapons – that they make it easier to kill, by allowing "killing at a distance" – seems to me to be rather weak. Crossbows allow people to kill at a distance and cruise missiles allow them to kill without ever laying eyes on their target. Operating a weapon by remote control doesn't seem to add

anything new to this. Indeed, one might think that the operators of UAVs will be *more* reluctant to kill than bombardiers or artillery gunners because they typically see what happens to the target when they attack it.

You mentioned earlier that it is hard to see anything wrong with the use of robots for tasks like mine clearing or IED disposal. In your 2008 article, "Building a Better WarBot. Ethical Issues in the Design of Unmanned Systems for Military Applications", you go further than that and suggest that it is not just ethical to use robots but ethically mandated to do so if possible. Are there other scenarios in which you think the use of robots is morally required? Also, in that paper, you point towards the often neglected effects the use of teleoperated robots has on their operators. Is this something which should be considered more in the discussion of ethical challenges of military robots?

There is some truth to the thought, "why send a person, when a robot can do it?" Commanders should be trying to protect the lives of those they command. Thus, if a robot can do the job instead of a human being, without generating other ethical issues, then, yes, it would be wrong not to use the robot.

Of course, there are two important caveats in what I've just said.

Firstly, the robot must be capable of succeeding in the mission – and, as I've said, I think there are fewer military applications where robots are serious competitors with human warfighters than people perhaps recognise.

Secondly, there must not be other countervailing ethical considerations that argue against the use of the robot. In particular, attacks on enemy targets by robotic systems must meet the tests of discrimination and proportionality within *jus in bello*. As long as there remains a human being "in the loop", making the decision about weapon release, this need not present any special difficulty, so the use of teleoperated weapons such as the Predator will often be ethically mandated if the alternative is to put a human being in danger in order to achieve the same tasks. Suppression of enemy air defences is another case, often mentioned in the literature, where it may be wrong not to use a robot. If fully autonomous weapons systems are involved, however, the balance of considerations is likely to change significantly. Except in very specific circumstances, such as counter-fire roles, wherein it is possible to delineate targets in such a way as to exclude the possibility of killing non-combatants, these weapons are unlikely to be capable of the necessary discrimination. Moreover, with both sorts of weapons there may be other ethical issues to

take into account, which might make it more ethical to send a human warfighter.

It is also worth keeping in mind that the ethics of using a weapon once it exists and the ethics of developing it may be very different. We may have good reasons *not* to develop weapons that it might be ethical to use – for instance, if the development of the weapon would make war more likely.

Regarding the operators, yes, I very much believe that people should be paying more attention to the effects that operating these weapons will have – indeed, are already having – on their operators and to the ethical issues arising from them. Remotely operating a weapon like the Predator places the operator in a unique position, both "in" and outside of the battlespace. Their point of view and capacity for military action may be in Afghanistan, while they themselves are in Nevada. After they fire their weapons, by pressing a few controls, they "see" the bloody results of their actions. Yet they have access to few of the informal mechanisms arising out of deployment in a foreign theatre that may help warfighters process the experiences they have been through. I have heard anecdotal reports from several sources that suggest the rates of post-traumatic stress disorder in the operators of the Predator are extremely high – and it certainly wouldn't surprise me if this was the case.

Gustav Däniker coined the term "miles protector" in 1992 after the Gulf War and summed up the new tasks of the future soldier in the slogan "protect, aid, rescue" ("Schützen, Helfen, Retten"). On the other hand there are arguments for the soldier to return to the role of the warfighter, often called the "core task" of soldiers. Do you think the shift from "war" to "policing" will have a significant impact on the self-image of soldiers and could you elaborate on your research in this matter?

I don't think increased use of robots will lead to a shift from "war" to "policing". Rather, I am arguing that the appropriate model to use in order to think about the justification of killing people who are no threat to you is "policing". Police need to take much more care to protect the lives of bystanders and have far fewer moral privileges in relation to killing than do soldiers during wartime. So, yes, *were* soldiers to start to take on this role, this would require a significant shift in their self-image. However, as I said, the argument I'm making about robots concerns the question of when, if ever, killing people via a robot is justified – not how often this is likely to happen. Unfortunately, I think it is much more likely that armed forces will use robots to kill people when they

shouldn't than it is that they will change the nature of the missions they engage in because they have these new tools.

Robots are of limited use in the sorts of peace-keeping and peace-enforcement missions that Däniker had in mind when he coined the phrases you mention. However, they do clearly have their place. Avoiding casualties may be especially important when governments cannot rely upon public support for taking them because the national interest is not at stake. Mine-clearing and bomb disposal are often an important way of winning the support of local populations – and robots can play a role here. The sort of surveillance that UAVs can provide is clearly a vital asset if one's goal is to prevent conflict and keep enemies apart. To the extent that armed UAVs can attack targets more precisely, with less risk of unintended deaths, they may also contribute to the success of peace enforcement missions. However, ultimately success in these sorts of deployments will depend upon talking with local people and on building trust and relationships on the ground. Robots have nothing to contribute to this goal and may even get in the way of achieving it – if, for instance, commanders' access to intelligence from UAVs prevents them from seeking human intelligence, or if the robots function in practice to isolate and alienate troops from the local population.

On the other hand, I do think the use of robotic weapons has the potential to radically unsettle the self-image of soldiers ... if not along the lines you suggest. For instance, there is no need for warfighters to be courageous – at least in the sense of possessing physical courage – if they will be operating weapons thousands of miles away; nor need they be especially fit or even able-bodied. There can be no argument that women should not take on "combat" roles operating robots, as physical strength is irrelevant in these roles, as is vulnerability to sexual assault (I'm not saying these were ever good arguments – just that it is especially obvious that they have absolutely no validity in this circumstance). It is hard to see how notions of "comradeship" apply when troops involved in the same battle – or even in the same unit – may be in completely different locations. It is not clear that one can really display mercy by means of a robot: one might refrain from slaughtering the enemy but this in itself is not sufficient to demonstrate the virtue of mercy. Indeed, there are whole sets of virtues and character traits currently associated with being a good "warrior" that may be completely unnecessary – or even impossible to cultivate – if one's role is operating a robot.

Of course, it has always only been a minority of those serving in the armed forces who needed to be brave, resolute, physically fit, etcetera, and we are a long way yet from being able to replace significant numbers of frontline troops with robots. Yet it is clear that there is a real tension between the dynamics driving the introduction of unmanned systems and the traditional function and self-image of soldiers. Eventually, I suspect, this will cause real problems for military organisations in terms of their internal cultures and capacity to recruit.

Since the St Petersburg Declaration of 1868 there have been various initiatives to restrict the use of weapons which cause unnecessary suffering. Do you think there is a need for additional international legislation to regulate the development and deployment of robots by the military? If so, what could be brought forward in favour of such legislation?

I definitely think we should be working towards an international framework for regulating the development and deployment of military robots – although perhaps not for the reason you suggest nor by the means you suggest.

I haven't seen any reason yet to believe that the use of robots will cause unnecessary suffering in the way that, for instance, nerve gas or dum dum bullets arguably do. Nor will robots necessarily kill any more people than the weapons and systems they will replace.

The reason to be worried about the development of more and more sophisticated robotic weapons is that these systems may significantly lower the threshold of conflict and increase the risk of accidental war. The fact that governments can attack targets at long distances with robotic weapons without risking casualties may mean that they are more likely to initiate military action, which will tend to generate more wars. I think we have already seen this effect in action with the use of the Predator in Pakistan and northern Africa. If robotic weapons begin to be deployed in roles with "strategic" implications – for instance, if the major powers start to place long-range and heavily armed uninhabited aerial vehicles or unmanned submersibles on permanent patrol just outside the airspace or territorial waters of their strategic rivals – then this will significantly decrease the threshold of conflict and increase the risk of accidental war. If fully autonomous weapons systems enter into widespread use then this will put a trigger for war into the hands of machines, which might also increase the risk of accidental war.

So, yes, there are very good reasons to want to regulate the development of these weapons. However,

for pragmatic reasons to do with the likelihood of reaching agreement, I think it might be better to approach this as a traditional case for arms control, with bilateral or regional agreements being a priority, perhaps with the ultimate goal of eventually extending these more widely. It is hard to see the United States or Israel, which have a clear lead in the race to develop robotic weapons, accepting restrictions on the systems until it is in their interests to do so. Yet if their strategic competitors become capable of deploying weapons that might pose a similar level of threat to them then they might be willing to consider arms control. Concerns about the threshold of conflict and risk of accidental war are familiar reasons to place limits on the number and nature of weapons that nations can field. As I argue in a recent paper, "Predators or Ploughshares?", in *IEEE Technology and Society Magazine*, a proper arms control regime for robotic weapons would need to govern: the range of these weapons; the number, yield, and range of the munitions they carry; their loiter time; and their capacity for "autonomous" action. If we could achieve one or more bilateral agreements along these lines it might then be possible to extend them to a more comprehensive set of restrictions on robotic weapons, perhaps even in the form of international law. I suspect we are a long way from that prospect at this point in time.

When it comes to the attribution of responsibility for the actions of military robots you have suggested an analogy between robots and child soldiers. Could you elaborate on this?

It is important to clarify that I was writing about cases in which it might be plausible to think that the robot "itself" made the decision to kill someone. There are actually three different scenarios we need to consider when thinking about the responsibility for killing when robots are involved.

The first is when the "robot" is a remote control or teleoperated device, as is the case with Predator and other UAVs today. In this case, it is really the human being that kills, using the device, and the responsibility rests with the person doing the killing.

The second is where the robot is not controlled by a human being but reacts to circumstances "automatically" as it were, as it would if it were controlled by clockwork or by a computer. In this case, the appropriate model upon which to conceptualise responsibility is the landmine. While there is a sense in which we might say that a landmine "chose" to explode at some particular moment, we don't think that there is any sense in which the moral responsibility for the death that results rests with the mine.

Instead, it rests with the person who placed the mine there, or who ordered it to be placed there, or who designed it, etcetera. This model remains appropriate to robots, even if the robot contains a very sophisticated onboard computer capable of reacting to its environment and tracking and attacking various targets, etcetera, – as long as there is no question that the robot is a machine lacking consciousness and volition. When computers are involved it may be difficult to identify which person or persons are responsible for the "actions" of the machine. However, it is clear both that the question of responsibility will be no different in kind to others that arise in war due to the role of large organisations and complex systems and that the appropriate solution will usually be to *assign* responsibility to some person.

A third scenario will arise if robots ever come to have sufficient capacity for autonomous action that we start to feel uncomfortable with holding human beings responsible for their actions. That is, if we ever reach the point where we want to say that the robot itself made the decision to kill someone. It's clear that none of the current generation of military robots come anywhere near to possessing this capacity – whether they ever will depends upon the progress of research into genuine artificial intelligence.

It was this third scenario that I was investigating in my article on "Killer Robots". I was interested in whether it will *ever* be possible to hold even genuine artificial intelligences *morally* responsible for what they do, given the difficulties involved in applying some of our other concepts, which are connected to responsibility, to machines – concepts such as suffering, remorse, or punishment. It seems as though there is a "gap" in the spectrum of degrees of autonomy and responsibility, wherein certain sorts of creatures – including, possibly, robots – may be sufficiently autonomous that we admit they are the origin of their actions, but not to the extent that we can hold them morally responsible for their actions. When we are dealing with entities that fall into this gap then we rightly feel uncomfortable with holding someone else responsible for their actions, yet it is hard to see what the alternative might be – unless it is to admit that no one is responsible. The latter option is not something we should accept when it comes to the ethics of war.

The use of child soldiers was the best model I could come up with to help think about this scenario. With child soldiers, you can't really hold them morally responsible for what they do, however, nor would it be fair to hold their commanding officer morally responsible for what they do, if he or she was ordered to send

them into battle. Even the person who conscripts them seems to be responsible for *that* rather than for what the children do in battle. One – though not necessarily the most important – of the reasons why using child soldiers in warfare is unethical, then, is that they may cause deaths for which no one may properly be held responsible. I think there is a similar danger if we ever reach the point where we would be willing to say that robots were really making the decision as to who should live or die ...

Though it is still disputed whether there will be ever something like a genuine artificial moral agent, it seems clear that artificial intelligence in military robots will continually improve and the roles of military robots will expand in future armed conflicts. So if robots gradually enter this third scenario – being sufficiently autonomous that they are the origin of their actions but not such that we can hold them morally responsible for their actions – how could this be integrated in the existing ethics of war? And is "keeping the human in the loop" – which the military always insist they will do, whenever these weapons are mentioned – a serious and plausible possibility?

The answers to your two questions are closely connected. Let me begin with your second question because it is, perhaps, slightly easier to answer and because the answer to this question has important implications for the answer to your first question.

We could *insist* upon keeping human beings in the loop wherever robots are used but this could only be sustained at a high cost to the utility of these systems – and for that reason I think it is unlikely to happen, despite what military sources say today. The communications infrastructure necessary to keep a human being in the loop is an obvious weak point in unmanned systems. In the longer term, the tempo of battle will become too fast for human beings to compete with robots. For both these reasons, the military is eventually likely to want to field systems that are capable of operating in "fully autonomous" mode: if an arms race to build robotic weapons should develop, then nations may have little choice *but* to field autonomous weapons. Moreover, there are some potential roles for unmanned systems, such as long-range anti-submarine warfare or "stealthed" precision air strikes, where it simply will not be possible to put a human being in the loop. Yet, again, these are applications that nations in pursuit of military supremacy – or even parity – can ill afford to ignore. It is therefore a politically expedient fiction, which the military are promulgating, to insist that there will always be a human in the loop. What's more, I

think the better military analysts know this!

The answer to your second question is therefore both "yes" and "no". Keeping human beings in the loop *is* plausible in the sense that we could do it and – I will argue in a minute – we may have good reasons to do it. However it is *not* a serious possibility in the sense that it is not likely to happen without a concerted effort being made to achieve it.

To turn now to your first question. As far as integrating autonomous weapons systems into the ethics of war goes, I believe this will be very difficult – as my comparison with child soldiers suggests. The obvious solution, which is, I believe, the one that militaries will eventually come to adopt, is to *assign* responsibility for the consequences of the use of autonomous weapons to the person who orders their use; we might think of this as insisting that the commander has "strict liability" for any deaths that result. However, the question then arises as to whether or not this is fair to the military officers involved? Commanders are currently held responsible for the activities of the troops they command but this responsibility is mitigated if it can be shown that individuals disobeyed their orders and the commander took all feasible steps to try to prevent this. Where this occurs, the moral responsibility for the troops' actions devolves to the troops themselves. It is this last step that will be impossible if it is machines that have "chosen" to kill without being ordered to do so, which is why we may need to insist upon the strict liability of the commander. However, this means there is a risk the commander will be held responsible for actions they could not have reasonably foreseen or prevented. I must admit I also worry about the other possibility – that no one will be held responsible.

If we do begin using autonomous weapons systems with something approaching genuine artificial intelligence in wartime, then we must insist that a human being be held responsible for the consequences of the operations of these weapons at all times – this will involve imposing strict liability. The alternative would be to reject the use of these systems and to insist upon keeping a human being in the loop. However, as I've said, there are many dynamics working against this outcome.

I should mention that another alternative that has received a significant amount of attention in the literature and the media recently – that we should "program ethics" into the weapon – is to my mind an obvious non-starter. Ron Arkin at Georgia Tech has recently published a book advocating this. However, with all respect to Ron, who

was extremely kind to me when I visited him at Georgia Tech, this is a project that could only seem plausible as long as we entertained a particularly narrow and mechanical view of ethics.

It *will* undoubtedly be possible to improve the capacity of robots to discriminate between different categories of targets. Moreover, there are, perhaps, some categories of targets that it will almost always be ethical to attack. John Canning, at the US Naval Surface Warfare Centre, is very keen on the idea that autonomous weapons systems might be programmed to attack only those holding weapons or even to attack only the weapon system, thereby disarming the enemy.

However, even if it is possible to build such systems there is a real possibility of deadly error. The proper application of the principles of discrimination and proportionality, which largely determine the ethics of using lethal force in wartime, is extremely context dependent. Even if the potential target is an enemy Main Battle Tank – which you'd normally think it would be okay to attack – whether or not this is ethical in any particular case will depend on context: whether the enemy has surrendered, or is so badly damaged as to no longer pose a threat, or has recently started towing a bus full of school children. More generally, assessments of when someone or something is a legitimate military target will often depend on judgements about the intentions of the enemy, which in turn need be informed by knowledge of history and politics. Robots don't have anywhere near the capacity to recognise the relevant circumstances, let alone come to the appropriate conclusions about them – and there is no sign that they are likely to have these for the foreseeable future. So even the idea that we could rely upon these systems to be capable of discrimination seems to me a fantasy.

When it comes to the idea that they could actually reason or behave ethically, we are even more firmly in the realm of science fiction. Acting ethically requires a sensitivity to the entire range of human experience. It simply isn't possible to "algorithmatise" this – or at least no philosopher in human history has been able to come up with a formula that will determine what is ethical. I would be very surprised if any engineer or computer scientist managed to do so!

You mentioned at the outset that your early research was about non-military robots. Before we finish, can we talk about that for a moment? Do you have any thoughts on the use of robots more generally, their impact on society, and their possible influence on interpersonal relations? I know that people are

talking about a future for robots in the entertainment and sex industries and that you have written about the ethics of using robots in aged care settings. Should we be looking forward to the development of robot pets and companions?

I think it's highly improbable that robots will have much influence on society or interpersonal relations for the foreseeable future – mostly because I think it is unlikely that robots will prove to be useful in our day-to-day lives anytime soon. Since the 1950s at least, people have been talking about how we would soon have robots living and working alongside us. I am still waiting for my robot butler!

There are some pretty straightforward reasons for the absence of any useful robots outside of very specific domains, although they are often ignored in media discussions of the topic. Humans are complex and unpredictable creatures, which makes us hard for robots to deal with. In order for robots to be able to perform useful roles around the home or in the community, they would need to be large, which means they will be heavy and therefore dangerous, and extremely sophisticated, which means they will be expensive and difficult to maintain. For all these reasons, robots and humans don't mix well and in domains where robots do play a significant role, such as manufacturing, this has been made possible by keeping robots and people apart.

Bizarrely, war turns out to be a relatively friendly environment for robots. Killing someone, by pointing and firing a weapon at them, is a much easier task for a robot than helping them is. War is also a domain in which it is plausible to think one might be able to reliably separate those humans we don't want to place at risk of injury from the robots that might injure them through the simple expedience of ordering the human beings to stay clear of the robots. This also has the virtue of protecting the robots. Budgets for "defence" spending being what they are, military robots can be very expensive and still profitable to sell and manufacture. "Domestic" robots would have to compete with underpaid human carers and servants, which makes it much tougher to make them commercially viable. There is, admittedly, more room for the development of more-and-more sophisticated robotic toys, including sex toys, but I think we are a long way from the point where these will start replacing relations between people or between people and their (real) pets.

None of this is to say that I don't think there are ethical issues associated with the attempt to design robots for these roles. Designing robots so that people mistake them

for sentient creatures involves deception, which may be problematic. Thinking it would be appropriate to place robots in caring roles in aged care settings – or even to use them to replace human workers, such as cleaners, who may be some of the few people that lonely older people have daily contact with – seems to me to involve a profound lack of empathy and respect for older people.

I *am* looking forward to seeing more robots. Robots are cool! I think the engineering challenges are fascinating, as is what we learn about the problems animals and other organisms have solved in order to live in the world. However, we should remember that engineers want to – and should be funded to – build robots because of the challenges involved and that often the things they are required to say nowadays to secure that funding involve them moving a long way outside of their expertise. As soon as people start talking about real-world applications for robots, the most important things to consider are facts about people, societies, politics, economics, etcetera. These are the things that will determine whether or how robots will enter society. Indeed, it has always been the case that when people appear to be talking about robots, what they are mostly talking about is human beings – our values, our hopes and fears, what we think are the most pressing problems we face, and what sort of world we want to live in. This is one of the reasons why I chuckle whenever I hear anyone talking about Asimov's "three laws of robotics" as though these were a serious resource to draw upon when thinking about how to build ethical robots. Asimov was writing about people, not robots! The robots were just devices to use to tell stories about what it meant to be human.

The fact that human beings build – and talk about – robots to satisfy and amuse other human beings means that the most important truths about robots are truths about human beings. When it comes to talking about the future of robotics, then, you would often do just as well – or even better – talking to a philosopher or other humanities scholars rather than to an engineer or roboticist.

References

Sparrow, R. 2004. "The Turing Triage Test." Ethics and Information Technology 6(4): 203-213.

Sparrow, R. 2002. "The March of the Robot Dogs." Ethics and Information Technology 4(4): 305-318.

Sparrow, R. 2007. "Killer Robots." Journal of Applied Philosophy 24(1): 62-77.

Kilcullen, David, and Andrew Mcdonald Exum. 2009. “Death from above, Outrage Down Below.” New York Times, May 17, WK13.

Kahn, Paul W. 2002. “The Paradox of Riskless Warfare.” Philosophy & Public Policy Quarterly 22(3): 2-8.

Sparrow, R. 2009. “Building a Better WarBot : Ethical issues in the design of unmanned systems for military applications”. Science and Engineering Ethics 15(2):169–187.

Daniker, Gustav. 1995. The Guardian Soldier. On the Nature and Use of Future Armed Forces. UNIDIR Research Paper No. 36. New York and Geneva: United Nations Institute for Disarmament Research.

St Petersburg Declaration 1868. Declaration Renouncing the Use, in Time of War, of Explosive Projectiles Under 400 Grammes Weight. Saint Petersburg, 29 November /11 December 1868. Available at http://www.icrc.org/IHL.NSF/FULL/130?OpenDocument.

Sparrow, R. 2009. “Predators or Plowshares? Arms Control of Robotic Weapons”. IEEE Technology and Society 28(1): 25-29.

Arkin, Ronald C. 2009. Governing Lethal Behavior in Autonomous Systems. Boca Raton, FL: Chapman and Hall Imprint, Taylor and Francis Group.

Canning, John S. 2009. “ You've Just Been Disarmed. Have a Nice Day!”. IEEE Technology and Society 28(1): 12-15.

Sparrow, R., and Sparrow, L. 2006. “In the hands of machines? The future of aged care.” Minds and Machines 16:141-161.

Peter Asaro:
Military Robots and Just War Theory

How and why did you get interested in the field of robots and especially military robots?

When I was writing my dissertation on the history of cybernetic brain models and their impact on philosophical theories of the mind, I became very interested in the materiality of computation and the embodiment of mind. From a technological perspective, materiality had a huge impact on the development of computers, and consequently on computational theories of mind, but this material history has been largely ignored, perhaps systematically to make computation seem more like pure mathematics.

During this time, I was asked to write a review of a book by Hans Moravec, about robots with human-level cognition, which made some pretty wild speculations based on the notion that cognition was a purely Platonic process that would someday escape its materiality. For instance, the idea that computational simulations might become just as good as real things if they were complicated enough, and contained enough detail and data. It seemed to me that this missed the role of material processes in cognition and computation.

This led me to start thinking about explicitly material forms of artificial cognition, more specifically robots as computers with obvious input-output relations to the material world. Pretty soon I was making a documentary film about social and emotional robotics, *Love Machine* (2001), which explored how important embodiment is to emotions like love and fear, and how roboticists were seeking to model these and what it would mean to build a robot that could love a person.

Because of that film, a few years later I was invited to write a paper on "Robot Ethics." In researching that paper, I came across Colin Allen and Wendell Wallach's work on artificial moral agents, I was struck again by a sense that embodiment and materiality were not getting the attention they deserved in this emerging field. It seemed to me that the goal of robot ethics should not be to work out problems in ethics using computers, but to actually figure out ethical rules and policies for how to keep real robots from doing real harm to real people. The most obvious place where such

harm might occur, and thus ethical considerations should arise, also turns out to be the area of robotics research that is receiving by far the most funding: military applications. The more research I did on the state-of-the-art of military robotics, the more I realized that this was a social and political issue of great importance, as well as one of philosophical interest. So I pursued it.

In the last couple of years, how did philosophy as a professional field adjust to the intensified development and deployment of artificial intelligence, robots in general and of unmanned systems by the military in particular? As a philosopher yourself, in your personal opinion, how should and how could philosophers contribute to the debates in this field?

I would say that as a professional field, I am a bit disappointed that philosophy has not had a better organized response to the rise of technology in general, and the intensified development and deployment of AI and robots in particular. While there are some good people working on important issues in these areas, there are only a handful of groups trying to organize conferences, workshops and publications at the intersection of philosophy and real-world computing and engineering. Especially compared to other subfields like medical ethics, bio-ethics, neuro-ethics, or even nano-ethics, where there seems to be more funding available, more organizations and institutes, and more influence on the actual policies in those areas. But information ethics has been getting traction, especially in the areas of information privacy and intellectual property, so perhaps robot ethics will start to catch up in the area of military robotics. It is still a small group of people working on this problem, and most of them seem to be on your interview list.

In my opinion, philosophers can make significant contributions to the debates on the use of military robotics. Philosophers are often accused of navel-gazing and irrelevance, whereas the development and use of lethal military robotics presents philosophically interesting problems with pressing real-world relevance. So this issue has the potential to make philosophy more relevant, but only if philosophers are willing to engage with the real-world complexity of the debate. And doing so can be fraught with its own moral and ethical issues – you have to consider if your own work could be used to justify and rationalize the development of some terrible new weapon. The theoretical work requires a great deal of intellectual integrity, and the policy work requires a great deal of moral sensitivity. I think these are the traits of a good philosopher.

A lot of people think about military robots and unmanned systems merely in technological categories. Why do you think it is necessary to broaden the approach and to stress ethical and philosophical aspects if machines are to be developed and used in military contexts?

Part of the reason that military robots snuck up on us so quickly, despite the warnings from science fiction, is that in many ways they are only small technological steps beyond military systems that we already know and accept in modern warfare. The initial strategy to call these systems into question is to argue that "autonomy" is a critical disjunction, a qualitative leap, in the evolution of military robots. But I do not think it is necessary to make that argument in order to question the morality of using robotics. In fact, my most recent article focuses on the morality of teleoperated robotics. Rather, I think we can look at the history of military strategy and technology, especially beginning in World War I and continuing through the Cold War and the Global War on Terror, and see how our generally accepted views of what is ethical in war have evolved along with new technologies. It is not a very flattering history, despite the fact that most officers, soldiers and engineers have made concerted efforts to make ethical choices along the way.

In my view, the critical ethical issues are systemic ones. We will not have more ethical wars just because we have more ethical soldiers, or more ethical robots. First of all, this is because there will always be a fundamental question of whether a war is just or not. The moral justification for developing and amassing military power will always depend upon the morality of the group of individuals who wield that power and how they choose to use it (Just War theorists call this *jus ad bellum*).

Second of all, warfare is a cultural practice. While it is cliché to say that warfare has been around as long as humans have (or even longer among other animals, perhaps), it is important to note that how wars are fought is built upon social, cultural and ethical norms that are very specific to a time and a culture. Over the last two centuries, warfare has become increasingly industrialized, subjected to scientific study, and made increasingly efficient. One result of those efforts is the incredibly sophisticated weapons systems that we now have. On the one hand, it is not necessary that efficiency should be the highest value – nations could have pursued honour, chivalry, valour, glory, or some other values as the highest, and then warfare would look different now. On the other hand, efficiency alone is not sufficient to win a war

or control a population because there is a huge socio-psychological element as well – which is why we have also seen militaries develop and deploy media and communication technologies, as well as rhetoric and propaganda, to shape people's perceptions and beliefs. Even if we believe Machiavelli when he advises his prince that it is better to be feared than loved, fear is still a psychological phenomenon, and even the most ruthless and technologically advanced tyranny could not maintain itself without sufficiently aligning the interests of the people with its own. There are numerous examples of great and mighty militaries that have successfully destroyed the military forces of their enemies, but ultimately failed to conquer a territory because they failed to win the "hearts and minds" of those who lived there. Which is just another way of saying that warfare is a cultural practice. Of course, there are also many examples of conquerors simply trying to eliminate the conquered peoples, and the efficiency of modern weapons makes genocide more technically feasible than it was historically. Robot armies could continue this trend to terrible new levels, allowing even smaller groups of people to dominate larger territories and populations, or commit genocides more quickly and with fewer human collaborators. Hannah Arendt argued that because of this, robot armies are potentially more insidious than atomic weapons.

If we want to take a step back from history, and the question of why we have come to a place where we are building lethal military robots, we can ask how we should build such robots, or whether we should build them at all, or what we should be building instead. So from a strategic point of view, the US might undermine support for terrorists more efficiently through aid programs to places where terrorism thrives due to poverty, than they would by putting those funds towards demonstrating their military superiority. We can also ask what values a nation is projecting when they commit such vast amounts of time and resources to fighting a war by remote-control, or with autonomous robots. Having received your questions just after the 40th anniversary of the Apollo 11 moon landing, I am reminded that despite its being a remarkable event in human history, it only occurred because of the specific history of the Space Race as a competition between the ideologies of the Cold War. In that case, the US scored a symbolic victory in technological achievement by landing a man on the moon, but it was also about projecting values of ingenuity, technological sophistication and teamwork. The US also spent a vast amount of mental and monetary resources in achieving that goal. In the case of military robotics,

I think it is a philosophical question to ask what values are being promoted and projected by these technologies, and if those are the values society ought to be pursuing above others. If we want to project technological prowess and pragmatic ingenuity, this could also be done through developing technologies, public works, aid, and environmental projects that ameliorated the underlying social, political and resource problems.

Contrary to most of the media coverage, the unmanned systems deployed by the military today are in general mostly tele-operated (though including some autonomous functions or potential) but not fully autonomous. In your last article for the IEEE Technology and Society magazine[1] you were specifically pointing out the importance of ethical considerations regarding these systems, which rely on human decision making and analyzed three different approaches to the design of these systems. Could you elaborate on that?

In that paper I was approaching the ethics of tele-operated lethal military robots as a problem in engineering ethics. That is, I wanted to ask what it would mean to actually design such a system "ethically." Mary Cummings, a former Navy combat pilot who now teaches interface design at MIT, has taken a similar approach. She calls her approach "value-centered design" and the idea is to have engineers brainstorm about potential ethical or safety issues, establish sets of values that should be design goals (like limiting civilian deaths), and then to actually evaluate and compare the alternative system designs according to those values. Another view proposed by Ron Arkin (actually for autonomous robots but it could be applied to tele-operated robots as well) is that of the "ethical governor." Basically, this is a system which follows a set of rules, like the Laws of Armed Conflict and Rules of Engagement, and stops the robot if it is about to commit a war crime or an atrocity. This approach assumes that you can develop a set of rules for the robot to follow which will guarantee it does nothing unethical on the battlefield.

The problem with both of these approaches is that they see values and ethical rules as black boxes. It is as if we can simply program all the ethical rules and make the robot follow them without considering the context in which ethical decisions are made. However, in real-world moral and ethical decision-making, humans *deliberate*. That is, they consider different perspectives and alternatives, and then decide what is right in a given situation. Am I really more ethical because my gun will not fire when I point it at innocent people, or am I just less likely to shoot them? I think that if we

really want to make robots (or any kind of technology) more ethical, we should enhance the ethical decision-making of the people that operate them. The paper then asks: What would it mean to build technologies that actually do that? I propose a "user-centered approach," which seeks to understand how people actually make ethical decisions, as an information-processing problem. What kind of information do people actually use to make these lethal decisions? What roles do emotion, empathy, and stress play? We really do not understand these things very well yet, but I think the answers might surprise us, and might also lead to the design of technological systems which actually make it harder for people to use them unethically because they are better informed and more aware of the moral implications of their use of the system.

So if I understand you correctly, instead of equipping the user with an artificial ethical governor, you would prefer to "equip" the user with ethical values and understanding and leave the actual decision-making in the human sphere. This would be similar to the "keep the human in the loop" approach, which has also been put forward by some people in the militaries. On the other hand, especially the amount of information to be processed in shorter and shorter time by the human operator/supervisor of military systems is likely to increase beyond the capacity of the human physique, which might offer an advantage to systems without human participation. Do you think that this user-centered approach (and similar matters) could be regulated by international legislation, for example a ban on all armed autonomous systems without human integration of decision-making?

The short answer is: Yes, we should seek an international ban on all autonomous lethal systems, and require all lethal systems to have significant human involvement in the use of lethal force. Just what "significant human involvement" might mean, and how to make that both technologically effective and politically acceptable to potential participants to a treaty is a matter for discussion. Sure, there are questions about how to implement and enforce such a treaty, but just having an international consensus that such systems are immoral and illegal would be a major step.

I think we should strive to keep the human in the loop both because this clarifies moral responsibility in war, and because humans are already very sophisticated ethical information processing systems. Information technologies are quite plastic and can be developed in a variety of ways depending on our goals and interests. What I am suggesting is that instead of trying to

formalize a set of rules for when it is OK for a robot to kill someone, and build that into a robot as a black-box module, that as an ethical engineer one might instead invest technological development resources into improving the lethal decision-making of humans.

I have heard various versions of the argument that there is too much information, or not enough time, for humans to make the necessary decisions involved, and so there is, or soon will be, a need to automate the process. For instance, those who supported the "Star Wars" Strategic Defense Initiative argued that human reaction times were not sufficient to react to a nuclear assault, and so the missile defense system and retaliation should be made fully automatic. But while our intuitions might be to accept this in a particular high-risk case, this is actually a misleading intuition. If that particular case is highly improbable, and there are many potential high-risk system malfunctions with having such an automated system, then the probability of catastrophe from malfunction could be much higher than from the case it is designed to defend against. I think we are better off keeping humans in the loop and accepting their potential fallibility, as opposed to turning our fate over to an automated system that may have potentially catastrophic failures.

The mistaken intuition comes from the fact that you can justify all sorts of things when the fate of the whole world (or all of humanity, or anything of infinite or absolute value) is at stake, even if the probabilities are vanishingly small compared to the risks you incur from the things you do to avoid it. There is much more to the debates about keeping humans in the nuclear loop, particularly in nuclear deterrence theory, and in training simulations where many people (not aware it is a simulation) do not "push the button" when ordered to. I bring up this example because the history of this kind of thinking continues to have a huge influence on military technology and policy well after the end of the Cold War. While in the case of nuclear war the decisions may result in the end of civilizations, in robotic war the decisions may only result in the end of tens or hundreds human lives at a time (unless you are worried about robots taking over). The stakes are smaller, but the issues are the same. The differences are that our intuitions get distorted at the extremes on the one hand, and on the other hand that because the decision to kill one person on a battlefield where so many already die so senselessly does not seem like much of a change, so we might be seduced into accepting autonomous lethal robots as just another technology of war. For robotic systems, our intuition might

be to accept autonomous lethal robots with some kind of built-in safety system, or even believe that they might be "better" than humans at some decision-making task. However, the real risks of building and deploying such systems, and their negative long-term effects on strategy and politics, are probably much higher than the safety gains in the hypothetical design cases, but we just do not have any easy way to measure and account for those systemic risks.

I rather like Arkin's concept of the ethical governor for robots, actually, and think it is compatible with keeping humans in the loop. My disagreement is with his argument that such a system can outperform a human in general (though for any well-defined, formalized and operationalized case you can probably program a computer to do better if you work at it long enough) because the real world will always present novel situations that are unlike the cases the robot is designed to deal with. The basic idea for the ethical governor is for it to anticipate the consequences of the robot's actions, and to override the planned actions of the robot whenever it detects that someone will be wrongly killed as a result. That could be used as a safety mechanism that prevents humans from making mistakes by providing a warning that requires an override. Moreover, when we look at the current situation, and see that humans do far better than robots when it comes to ethical decision making, why are we investing in improving robot performance, rather than in further improving human performance?

Besides, if we really want to automate ethical decision-making, then we need to understand ethical decision-making, not just in theory but empirically. And so I argue that the first step in user-centered design is to understand the ethical problems the user faces, the cognitive processes they employ to solve those problems, and to find out what kind of information is useful and relevant, so that we can design systems that improve the ethical decision-making of the people who operate these lethal systems. I call this "modelling the moral user." If part of the problem is that there is too much information, that just means that we need to use the technology to process, filter and organize that information into a form that is more useful to the user. If part of the problem is that users do not know how much to trust or rely upon certain pieces of information, then the system needs to make transparent how and when information was obtained and how reliable it is. These are questions that are important both philosophically, as matters of practical epistemology and ethics, and from an engineering perspective.

In the last couple of years unmanned systems were deployed and used by the US Armed Forces in considerable numbers, e.g. in Afghanistan and Iraq, and are becoming a more and more common sight in and above the operational areas. With the ongoing developments, the ethical and legal debate on the deployment of robots as military (weapon) systems has intensified. From your point of view, what should be the main considerations regarding the Law of Armed Conflict and Just War Theory?

There are several crucial areas of concern in the Pentagon's increased adoption of robotic technology. It is hard to say what the greatest concern is, but it is worth paying attention to how military robots are already contributing to new strategies.

We should be immediately concerned at the increasing use of armed UAVs within Pakistan over the past 12 months--a policy begun under President Bush and embraced by President Obama. This policy is born out of political expediency, as a military strategy for operations in a country which the US is not at war with, nor is there any declared war.

By stating that it is a matter of political expediency I mean that the fact that these robotic technologies exist provides a means for a kind of lethal US military presence in Pakistan which would not be possible otherwise, without either the overt consent of the Pakistani government, expand the official war zone of the Afghan war to include parts of Pakistan, an act of war by the US against Pakistan's sovereignty, or the US risking the loss of pilots or commandos in covert raids (who would not be entitled to the rights of prisoners of war under the Geneva Conventions because they would not be participating in a war). There is a lack of political will within Pakistan to allow the US military to operate freely against the Taliban within its borders (though it was recently revealed that Pakistan does allow the US to operate a UAV launching base within its borders), just as there is a lack of political will in the US to destabilize Pakistan and take responsibility for the consequences. The UAVs provide a means to conduct covert raids with reduced risks, and while these raids are publicly criticized by officials of the Pakistani government, the situation seems to be tolerated as a sort of compromise solution. Despite the recent news that a US drone has assassinated the head of the Taliban in Pakistan, I am skeptical that these UAV "decapitation" raids will make a significant impact on the military or political problems that Pakistan faces, and may do more harm than good in terms of the long-term stability of Pakistan. This is a bad precedent for international conflicts insofar as it

appears to have resulted in numerous unnecessary civilian casualties outside of a declared war zone, and moreover it seems to legitimate a grey area of covert war fought by robots (thus allowing robots to circumvent international and local laws against extra-judicial and targeted killings and kidnappings much in the way on-line casinos circumvent laws against gambling through the physical separation of an agent and their actions). It is not surprising that these missions are under the operational control of the CIA (rather than the military), and that the CIA actually outsources the arming and launching of the UAVs in Pakistan to non-governmental mercenary forces such as Blackwater/XE. So while proponents of lethal robots are invoking Just War Theory and arguing that they can design these robots to conform to its standards, we see that the most frequent use of lethal robots today, in Pakistan, falls completely outside the requirements of Just War Theory because there is no war, and the military is not even pulling the trigger precisely because it is illegal for them to do so.

However, it should be noted that in Afghanistan the civilian casualties have been far greater in airstrikes from conventional aircraft and from commando raids, than from UAVs. I believe this is probably due to the fact that the Predator UAVs are only armed with Hellfire missiles, which are fairly accurate and relatively small compared to the large guided bombs dropped by conventional aircraft (but are now carried by the recently deployed Reaper UAVs), and because there have been comparatively fewer armed UAV missions so far. Commando raids probably have higher civilian casualty rates in part because the commandos have a strong interest in self-preservation and are much more vulnerable than aircraft (manned or unmanned), and due to the particular circumstances in Afghanistan – where nearly every household keeps guns and often military assault rifles for home-defense, and the natural reaction to gunfire in the streets is to come out armed with the house-hold gun. When those circumstances are combined with Rules of Engagement that allow commandos to kill civilians presenting a threat by carrying guns, it is not surprising that many civilians who support, or at least have no interest in fighting against, the US forces wind up getting killed in such raids. So while on the one hand we might want to argue that UAVs could reduce civilian casualties in such raids, we could also ask the systemic question of whether such raids are an effective or ethical strategy at all or, as some have argued, are really a tactic posing as a strategy. The Dutch military forces in Afghanistan have developed a very different strategy based on a community-policing

model, rather than a surgical-strike model, though unfortunately it is not being used in all regions of the country.

Ultimately, the situations in both Afghanistan and Pakistan require political solutions, in which the military will play a role, but even the most sophisticated robotic technologies imaginable cannot improve the situation by military means alone. So I think it is also a philosophical question to ask whether military technologies are being used in ways that actually work against, or merely postpone, addressing and solving the underlying problems.

In the near term of the next decade, I think the primary concern will be the proliferation of these technologies to regional conflicts and non-government entities. UAVs are essentially remote-controlled airplanes, and the ability to obtain the basic technologies and arm them is within the grasp of many organizations, including terrorists and other non-state actors. This is also being coupled with a trend towards unconventional, asymmetric war, and organized violence and terrorism which we often call “war” but actually falls outside the purview of Just War Theory and international law. *Al Qaeda* may be waging a campaign of international violence with political aims, but they are not a nation fighting a war for political control of a geographic territory. President Bush decided to call it a war and to use the military to fight *Al Qaeda*, and that decision has created other problems with treating members of *Al Qaeda* as prisoners of war, and putting them on trial for crimes, *etc.* So even if we have an international treaty that bans nation-states from building autonomous lethal robots, we will still face a challenge in preventing individuals and non-state organizations from building them. Of course, an international ban would dissuade the major military technology developers by vastly shrinking the potential economic market for those systems, which would greatly slow their current pace of development. Everyone would still be better off with such a ban, even if some systems still get built illegally. It will be much easier for small terrorist groups to obtain these technologies once they have been developed and deployed by militaries all over the world, than for them to try to develop these technologies themselves.

In the coming years we need to be vigilant of the Pentagon's efforts to make various robotic systems increasingly autonomous. Even autonomous self-driving cargo trucks have the potential to harm civilians, but obviously it is the armed systems that should be watched most closely. The current paradigm of development is to have

a single soldier or pilot controlling multiple robotic systems simultaneously through videogame-like interfaces. While this reduces personnel requirements, it also leads to information overload, confusion, mistakes, and a technological "fog of war." This may actually increase the pressure to make robotic systems fully autonomous, with engineers arguing that robots will actually perform better than humans in high-stress lethal decision making.

In the long term we need to be very concerned about allowing robotic systems to make autonomous lethal decisions. While there are already systems like Phalanx and Patriot that do this in limited ways, they are often confused by real-world data. In two friendly-fire incidents in 2003, Patriot missile defense systems operating in an automatic mode mistook a British Tornado and an American F-18 as enemy missiles and shot them down. Of course, we can design clever control systems, and improved safeguards, and try to prevent such mistakes. But the world will always be more complex than engineers can anticipate, and this will be especially true when robots engage people face-to-face in counter-insurgency, urban warfare, and security and policing roles (domestic as well as military). To distinguish someone fearfully defending their family from someone who represents a genuinely organized military threat is incredibly complicated – it depends on social, cultural and linguistic understanding that is not easily formalized as a set of rules, and is well beyond our technological capabilities for the foreseeable future. We need to be vigilant that such systems are not put in service without protest, and we should act now to establish international treaties to ensure that such systems are not developed further.

Interpreting and applying the Laws of Armed Conflict (LOAC) and developing Rules of Engagement (ROE) involve legal, political and military considerations. Because they have the potential to overwhelm individual ethical choices, or the ethical designs of robots, these interpretive processes ought to be open to critical investigation and public discussion. Arkin is confident that we can build the LOAC and ROE into the robots, but I think there are some problems with this. First, robots will not be able to do the interpretive work necessary to apply the rules to real-world situations. So what is really being put into the robots is an interpretation already made by system designers, built upon numerous assumptions and engineering considerations which may not work out in the real world. Second, sometimes the ROE are vague, confusing, or even inconsistent, and humans do not always understand when or how they

should be applied, so I cannot see how robots could do better.

Apart from the practical concerns of the technologies currently being developed, we should also be concerned about the shift in the philosophy of warfare they represent. The trend is to remove soldiers from the battle. While this is certainly good for their safety, it comes at a cost to the safety of others – in particular civilians on both sides of the conflict. The psychological distance created by remote-control or automated warfare serves to diminish the moral weight given to lethal decisions. It also serves to turn soldiers into civilians in that they start fighting wars from computer terminals in air-conditioned rooms miles away from the battle. As such it lends credence to terrorists who would claim civilians as legitimate targets. If you look at the wars that the US has been involved in over the last century, you see that as the military technology advances, the overall ratio of civilians to soldiers killed has also increased. And that is despite the wide-spread use of so-called "smart" weapons in Iraq. So while we are making war safer for soldiers, we are not really making it safer for civilians. We should be very concerned about the tendency of new military technologies to shift the risks from soldiers to civilians, as this can actually undermine the possibility of a "just war" even as the new technologies are being called "smart" or "ethical."

Concerning the ability of discrimination, it has been brought forward, that on the one hand artificial intelligence and sophisticated sensors could be more capable in performing this task than any human. And on the other hand that it would not even be necessary for autonomous systems to excel in the distinction of combatants/non-combatants but it would be sufficient if they equalled their human counterparts. Regarding Just War Theory, is this a maintainable argument and how would you review these and similar approaches?

Discrimination is a crucial criterion for Just War Theory, and it has been argued that automated systems might perform better than humans at the discrimination task. I think the question is: If we accepted that automated systems could outperform humans, or if we were actually presented with evidence that some system could perform the discrimination task at or above human levels, is that a good argument for allowing them to make autonomous lethal decisions? The short answer is: No.

First, discrimination is necessary but not sufficient for ethical killing in war. The point of the discrimination criterion is that it is never acceptable to intentionally kill innocent civilians, or to kill people indiscriminately in war. This does not imply that it is always acceptable to kill

enemy combatants (except, it is argued, in "total war" though I do not accept that argument). The way it is usually construed, combatants have given up their right not to be killed by putting on a uniform. Even under this construal, it is immoral to unnecessarily kill enemy combatants. For instance, killing retreating soldiers, especially just before a clearly immanent final victory or surrender, is generally viewed as immoral, though it is legal under international law. According to a rights-based view of Just War Theory, it is necessary for enemy combatants to also present an actual threat in order to justify their being killed. This could be much more difficult for automated systems to determine, especially since enemy combatants might only pose a threat to the robot, and not to any humans – does that count as a sufficient threat to warrant killing them?

Second, the other major criterion for Just War Theory is proportionality – that the killing and violence committed is proportional to the injustice that it seeks to correct. Just War Theory allows the killing of just enough enemy soldiers in order to win the battle or the war. Proportionality also requires that the use of violence is calibrated to justice. For example, if you punch me in the arm I might be justified in punching you back, but not justified in killing you. Similarly, if one nation were to repeatedly violate the fishing rules in the territorial waters of another nation, this would not justify a full-scale invasion, or the bombing of the offending nation's capital city, though it might justify sinking an offending fishing vessel. In this sense, proportionality can be viewed as a retributive component of Just War Theory. Just War Theory also allows for the unintentional killing of innocent civilians, often called "collateral damage," through the doctrine of double-effect. But the potential risk of killing civilians and the potential strategic value of the intended target, for example when considering whether to bomb a military installation with a school next to it, must both be taken into account in determining whether the risks and costs are justified. I do not believe that an automated system could be built that could make these kinds of determinations in a satisfactory way, because they depend upon moral values and strategic understandings that cannot be formalized. Of course, there are utilitarians and decision theorists who will argue that the values of innocent human lives, and the values of strategic military targets can be objectively established and quantified, but the methods they use essentially treat humans as oracles of value judgements – usually individual preferences or market-established values derived from aggregates of unquestioned individual valuations – rather than actually

provide an algorithm for establishing these values independently of humans. So again, I would not trust any automated algorithm for establishing values in novel situations.

According to the criteria of Just War Theory, do you think there could be a substantial objection against a military operation because of unmanned systems/military robots being used in it, now or – thinking of the future potential of increasing autonomy of these systems – in a future conflict?

Since I think that merely meeting the discrimination criterion of Just War Theory is not sufficient for meeting the other criteria, and I doubt that any fully automated system will ever meet the proportionality criteria, I think there are grounds for arguing against the use of systems that make fully automated lethal decisions in general.

Of course, I think we can make a substantial case for international bans on autonomous lethal robots, or other things like space-based weapons, regardless of whether they violate Just War Theory in principle. International treaties and bans depend more upon the involved parties seeing it as being in their mutual interest to impose binding rules on how warfare is conducted. The fundamental weakness of Just War Theory, as Walzer presents it, is that it cannot really be used to argue definitively against any military technology, insofar as both sides consent to use the technology against each other. The Ottawa Treaty is a notable exception here, insofar as it bans anti-personnel landmines on the basis of their indiscriminate killing of civilians, even long after a war. Mostly that treaty succeeded because of international outrage over the killing and maiming of children by landmines, and the expense of cleaning up mine fields. Basically, politicians could look good and save money by banning a weapon with limited applications that does not really change the balance of military powers.

International treaties tend to be somewhat arbitrary in what they ban, from the perspective of Just War Theory. Blinding enemy combatants is a more proportional way to neutralize the threat they pose than killing them, yet blinding lasers are banned as “disproportionately harmful” weapons. Space-based weapons are not intrinsically unjust, but they represent a potential “tragedy of the commons” in that destroying just a few satellites could put enough debris in orbit to start a chain-reaction of collisions that would destroy most of the orbiting satellites and make it nearly impossible to launch any more into orbit in the future. So it really is in the long-term interest of all nations to ban space-based weapons. There is a

United Nations Committee On the Peaceful Uses of Outer Space (UNCOPUOS) in Vienna that has done some really good work forging international cooperation in space. They have been working for many years to convince the international community to ban space-based weapons, but it is curiously unfortunate that the US, which stands to loose the most strategically from space-based weapons because it has so many satellites in orbit, is the country that is blocking treaties to keep weapons out of space. Perhaps we could form a UN committee on the peaceful uses of robotics?

In your posing of the question, you seem to be asking about whether one could argue against the use of autonomous lethal systems in a particular military operation. The particular case is actually harder to argue than the general case. If military planners and strategists have chosen a specific target, and planned an operation, and plan on using autonomous lethal robots to execute the plan, then we might appear to have a case where these technologies seem acceptable. First of all, there is a significant amount of human decision-making already in the loop in such a case, especially in that there is a valid "target." Second, if it is the kind of mission where we would be deciding between firing a cruise missile to destroy a target, or sending autonomous lethal robots to destroy the same target, that case is much trickier. Taking the case in isolation, the robots might spare more innocent civilians than a missile, or might collect some valuable intelligence from the target before destroying it. Viewing it in a broader systemic context can change things, however, as there will be new options made possible by the technology. So while there could be cases where an autonomous robot might offer a better option than some technology we already have, there may also be other new technologies that provide even better options. And we can always invent a hypothetical scenario in which a particular technology is the best possible option. But again, I think we need to be careful about how we define and think about autonomy and the level of control of the "humans-in-the-loop." If the humans using this option are willing to take responsibility for the robots completely destroying the target (as would be the case if they used a missile instead), and are in fact held responsible if the target turns out to be a school full of children with no military value, then the fact that they used robots instead of a missile makes little difference. The problem we must avoid is when the humans are not held responsible because they relied on the robot having a safety mechanism that was supposed to prevent it from

killing children. Our frameworks for ethical decision-making do not take into account how technologies change the options we have. The easiest solution to the problem is to make such autonomous systems illegal under international law.

[1] Peter M. Asaro, Modeling the Moral User in: *IEEE Technology and Society*, 28, 2009, p.20-24.

Jürgen Altmann: Uninhabited Systems and Arms Control

How and why did you get interested in the field of military robots?

I have done physics-based research for disarmament for 25 years. One strand concerned automatic sensor systems for co-operative verification of disarmament and peace agreements. My second, more interdisciplinary, focus is on assessment of new military technologies under viewpoints of peace and international security, and possibilities of preventive arms control. In 2000-2001 the German Research Association Science, Disarmament and International Security (FONAS) did joint projects on preventive arms control. In that context I studied potential military uses of micro-systems technology (Altmann 2001).

Already in that research I looked into the problem of military robots, then mostly small and very small ones. When I investigated military applications of nanotechnology, a very broad field, uses in uninhabited vehicles with sizes from large to extremely small were investigated (Altmann 2006). Limitations for such vehicles figured high in my recommendations for preventive arms control. Aware of the increasing number of countries developing and producing uninhabited air vehicles, of the large efforts for uninhabited ground and water vehicles, and of the rising trend to equip uninhabited vehicles with weapons, we proposed a research project which was granted in 2009.

Currently you are directing the project on "Unmanned Armed Systems – Trends, Dangers and Preventive Arms Control". Could you elaborate on the focus of your research?

This project – funded by the German Foundation for Peace Research (DSF) for 1.5 years – has four goals:

1. Compile the status in research, development and deployment of uninhabited armed systems;

2. Describe the technical properties of uninhabited armed systems to be expected in the next twenty years with the approximate times of their introduction;

3. Assess the systems to be expected under criteria of preventive arms control;

4. Analyse limitation options and verification possibilities.

These goals (with main focus on uninhabited aerial vehicles, UAVs) will be pursued in interdisciplinary research with considerable scientific-technical content. The results are to be published in a monograph.

You are also one of the founding members of the International Committee for Robot Arms Control (ICRAC). What were your motivations to set up the Committee and what do you hope to achieve by it?

At present we are four scientists from various disciplines: robotics, philosophy, physics/peace research – all of them contributing in this volume (P Asaro, N. Sharkey, R. Sparrow and myself) (ICRAC 2009). We are worried by the accelerating trend to arm uninhabited military vehicles, by the high numbers of non-combatants killed in present US and UK remote-control attacks in Iraq, Afghanistan and Pakistan, and by the seriously discussed prospect that soon computers may decide, when and whom to kill. We see dangers for the laws of warfare – discrimination and proportionality demand assessment of a complex war situation which for the foreseeable future artificial-intelligence systems will likely not be able to make. When the US near-monopoly of armed UAVs will be broken, additional dangers can be foreseen: from the undermining of arms-control treaties via the destabilisation of the situation between potential adversaries to proliferation and to possible use by terrorists. Politically, the prospect of sending fewer human soldiers and using mostly uninhabited combat systems may raise the inclination to go to war for some states.

We hope to raise awareness of the dangers connected to armed uninabited vehicles in the public as well as with decision makers. The goal is to prevent an unconstrained global arms race. For this, the important arms-producing states need to agree on mutual limitations with adequate verification mechanisms. Based on our founding statement, we want to develop concrete proposals for such limitations and hope that some states will take the initiative. For presenting and discussing concepts we shall convene an international expert workshop on robot arms control in September 2010 in Berlin.

What are the recommendations of the Committee?

They are contained in its founding statement:

“Given the rapid pace of development of military robotics and the pressing dangers that these pose to peace and international security and to civilians in war, we call upon the international community to urgently commence a discussion

about an arms control regime to reduce the threat posed by these systems.

We propose that this discussion should consider the following:

- Their potential to lower the threshold of armed conflict;
- The prohibition of the development, deployment and use of armed autonomous unmanned systems; machines should not be allowed to make the decision to kill people;
- Limitations on the range and weapons carried by "man in the loop" unmanned systems and on their deployment in postures threatening to other states;
- A ban on arming unmanned systems with nuclear weapons;
- The prohibition of the development, deployment and use of robot space weapons."

The founding of the ICRAC did produce considerable media interest. What kind of responses did the Committee receive from the international community and fellow researchers?

From governments, not many up to now. But committee members are regularly being invited to present their arguments to conferences, including ones organised by the military or for the military. Among the few other researchers worldwide who have written on potential problems from armed uninhabited vehicles we feel general support. This includes robot ethicists. The vast community of robotics and artificial-intelligence researchers has mostly not yet really taken up the problem of killing robots. We hope that this will change with a new robot-ethics book which covers military uses in three chapters (Capurru/Nagenborg 2009), with our upcoming workshop and related publications.

Where do you see the main challenges for the international community regarding the use of armed unmanned systems by the military. What are the specific challenges of autonomous systems as compared to current telerobotic systems?

The main challenge is in deciding whether the present trend should continue and expand to many more countries and to many more types of armed uninhabited vehicles (in the air, on and under water, on the ground, also in outer space), or whether efforts should be taken to constrain this arms race and limit the dangers connected to it. Here not only governments, but non-governmental organisations and the general public should become active.

Autonomous systems obviously would open many new possibilities for war by accident (possibly escalating up to nuclear war) and for

violations of the international laws of warfare. On the general ethical issue of machines autonomously killing humans, see the other interviews in this volume. A human decision in each single weapon use should be the minimum requirement.

Do you think the Missile Technology Control Regime (MTCR) could play a part in the non-proliferation of UAV technologies?

Yes, it does so already – its limitations concern UAVs (including cruise missiles) capable of carrying a payload of 500 kg over 300 km range. For UAV systems with autonomous flight control/ navigation or beyond-visual-range remote control and aerosol-dispensing mechanisms, there is neither a payload nor a range threshold. These rules could be expanded beyond aerosol dispensing. However, one-sided export-control regimes such as the MTCR do not encompass all developer/ producer/ exporter countries, and they do not limit the armaments of the regime members themselves. Truly effective would be export controls embedded in comprehensive prohibitions valid for all relevant countries, that is, in arms control and disarmament treaties, as is the case with biological and chemical weapons. Limits on armed uninhabited vehicles will need to be more differentiated and pose some definitional issues, but with the understanding of states that such limits are in their enlightened national interest the detailed rules could be worked out. Some general ideas have been published by members of our Committee (Altmann 2009, Sparrow 2009).

Regarding international humanitarian law, would you think there is a need for additional legislation concerning the deployment of unmanned systems?

The biggest problem is posed by autonomous attack decisions. In principle, the requirements of discrimination and proportionality would suffice to rule this out for one to two decades because artificial intelligence will at least for this time not achieve the level of human reasoning – and this is the standard of international humanitarian law. However, it has to be feared that military reasons and political motives lead to autonomy in weapon use much earlier, thus an explicit legal requirement to have a human making each single weapon-release decision is required. For remotely controlled systems a self-destruct mechanism in case of communication failure should be mandatory. Further rules will probably be needed – this should be the subject of legal research. Legal research would also be helpful in finding out whether

video images as the sole real-time information are sufficient for compliance with the laws of armed conflict, and if specific rules are needed here.

In your work you have stressed the threats autonomous armed systems can pose to arms-control treaties and to international humanitarian law. What would be the most pressing problems at the moment?

Seen from today, with a detailed analysis still pending, armed uninhabited vehicles – autonomous or not – would undermine nuclear-reduction treaties (INF Treaty, New START successor) if they were used as new nuclear-weapon carriers. The Treaty on Conventional Armed Forces in Europe would be endangered by armed ground vehicles outside of the Treaty definitions (of tanks or armoured combat vehicles) or by disagreement about which armed UAVs count as combat aircraft or attack helicopters (for some more information see Altmann 2009).

Most pressing are the issues of international humanitarian law. Already now remote-control UAV attacks in Iraq, Afghanistan, Pakistan – directed from thousands of kilometres away, based only on images from a video camera – lead to many civilian deaths, so that compliance with the requirements of discrimination and of proportionality is doubtful. With armed UAVs the only action-possibility is to shoot; soldiers on site would have more possibilities to act – check identities, search for weapons, take people into custody.

Even more problems would be created by autonomous attack – delegation of the authority to select targets to computers. If such autonomous armed uninhabited vehicles were to be introduced before one or two decades, one can expect a marked increase in civilian casualties.

This could be prevented by a prohibition of autonomous attack. At least as important are efforts to reduce the likelihood of war in the first place – with respect to the issue at hand by preventive arms control for armed uninhabited vehicles, on a more general level by general limitations of weapons and armed forces, combined with political measures of reducing confrontation.

As you noted, the use of unmanned systems can affect the decision to go to war. Do you think, with the possibility to wage war without putting one's own troops at risk, one of the principles of just war theory – war being the last resort (ultima ratio) – might be challenged?

This is not my area of expertise, but the thought suggests itself.

Apart from questions regarding the right to go to war (ius ad bellum), there is also the question of military necessity of actions in an armed conflict. Without the "man in the loop", and even if it is ensured that the target is a legitimate one, do you think autonomous systems should or could ever be entrusted with decisions as how, when and even if to attack such a target?

In a purely scientific view one can argue that autonomous systems could only be entrusted with such decisions if and when they had proven that they can assess complex situations in war at a level comparable to the one of a capable human commander. The slow speed of robotics/ artificial-intelligence development during the last fifty years and the scepticism of credible roboticists about progress in the coming decades lead me to the conclusion that this requirement will likely not be fulfilled in the next one or two decades. This conclusion is corroborated by the time frame envisaged for realisation of the "ultimate goal of the RoboCup Initiative", namely a team of humanoid robot soccer players winning against the World-Cup winner, which is "mid-21st century". If at some future time robotic systems consistently demonstrated better performance than humans, then one could argue that international humanitarian law and the ethics of war would even demand replacing humans.

However, robots/ artificial intelligence at or beyond the human level would raise fundamental ethical questions much beyond war and could bring existential dangers. Consideration of the interests of humankind and the precautionary principle could well lead to a rational decision for a general prohibition of the development of such systems. Ensuring compliance with such wide-ranging rules – similar ones will probably also be required with some future developments in nanotechnology – may need a transformation of the international system: moving away from trying to provide security by national armed forces to a system with a democratically controlled supranational authority with a monopoly of legitimate violence. Otherwise perceived military necessities and military resistance against far-reaching inspection rights could prevent nations from agreeing on strong limits on research and development, even though highest human interests would demand them.

In the discussion of the NATO air strike in Afghanistan near Kunduz in September 2009, it has been brought forward that the use of UAVs might have helped to prevent the amount of civilian casualties. Do you think the limited use of UAVs might actually increase the battlefield awareness of soldiers and eventually could help to achieve

proportionality and target discrimination on a higher level?

In principle it could. Unfortunately not all details of that attack are available. From media accounts it seems that the commanding officer consciously decided to have the two stolen fuel trucks bombed together with all people surrounding them, despite several offers of the bomber pilots to first overfly the scene to scare people away. So in this case the use of armed UAVs would probably not have made a difference.

Generally, having a weapon at hand where a UAV is observing could serve for more precise targeting and for reaction to short-term changes on site. But this could in principle also be provided by piloted aircraft. Video observation from very far away brings the possibility of misjudgements as many incidences of killing the wrong persons in Afghanistan and Pakistan demonstrate. But pilots on board aircraft have limited sensory input, too.

A final problem is that the awareness is only guaranteed in a very asymmetric situation: when one side has UAVs available while the other does not. The "fog of war" would be much thicker if both sides possess (armed) UAVs, jam each other's communication links etc.

In the last years you also have worked on projects concerning non-lethal / less-lethal weapon systems (e.g. acoustic weapons, a millimetre-wave skin-heating weapon). Where do you see the potential and the challenges of these systems, especially if they are mounted on autonomous weapon platforms?

Acoustic weapons do not really exist. An existing long-distance loudspeaker system (the so-called Long Range Acoustic Device from the USA) can be turned to higher intensity which would result in permanent hearing damage if unprotected persons are exposed at distances below, say, 50 m for longer than a few seconds (Altmann 2008). This demonstrates the main problem with acoustic weapons in the audio range: The transition from annoying or producing ear pain to lasting damage is very fast. (Infrasound, on the other hand, has no relevant effect and is difficult to produce in high intensities.) So if real acoustic weapons were deployed on UAV and used to attack a crowd, mass incidence of permanent hearing damage would be the probable outcome.

Concerning millimetre-wave weapons for producing pain by skin heating, the existing U.S. Active Denial System (with 500 to 700 m range, tested but not yet deployed) is very big, requiring a medium truck (Altmann 2008). Research is underway to develop an even stronger system to be carried on aircraft – it

is doubtful if that would be used without pilots and operators on board. If so, the general problems of applying force over a distance, not being on the scene, would be aggravated. The same would hold if other "non-lethal" weapons were used from uninhabited (air, ground) vehicles, say, tasers or, more traditionally, water cannons.

With "non-lethal" weapons, much depends on the scenario of use (armed conflict? peace-keeping operation? crowd? few criminals?), on the context and the general culture (democratic control of security forces?) in the respective society. One can suspect that putting them on uninhabited vehicles can increase, rather than decrease, the level of violence.

References

Altmann, J. 2001. Military Uses of Microsystem Technologies – Dangers and Preventive Arms Control, Münster: agenda.

Altmann, J. 2006. Military Nanotechnology: Potential Applications and Preventive Arms Control, Abingdon/New York: Routledge.

Altmann, J. 2008. Millimetre Waves, Lasers, Acoustics For Non-Lethal Weapons? Physics Analyses and Inferences, Forschung DSF No. 16, Osnabrück: Deutsche Stiftung Friedensforschung, http://www.bundesstiftung-friedensforschung.de/pdf-docs/berichtaltmann2.pdf.

Altmann, J. 2009. Preventive Arms Control for Uninhabited Military Vehicles, in Capurro/Nagenborg 2009, http://e3.physik.tu-dortmund.de/P&D/Pubs/0909_Ethics_and_Robotics _Altmann.pdf.

Capurro, R., Nagenborg, M. (Eds) (2009). Ethics and Robotics, Heidelberg: AKA/IOS.

Sparrow, R. 2009. Predators or Plowshares? Arms Control of Robotic Weapons, IEEE Technology and Society, 28 (1): 25-29.

Gianmarco Veruggio/ Fiorella Operto: Ethical and societal guidelines for Robotics

To introduce our topic, which is a discussion on Roboethics, let us start from robotics as such, and from a statement by yours: You hold that Robotics is a new Science. Is this claim true? Or, is it a wish of some roboticists, who are trying to attribute higher dignity to their studies?

GIANMARCO VERUGGIO
In 2004, roboticists and scholars of humanities gathered in Sanremo, Italy, to lay the foundations of a new applied ethics, which I, as the Chair of the Symposium, had called "Roboethics". This word did not exist before, nor was in any Encyclopedia neither on Google. The two days workshop took place in a historical location, the studying room of Villa Nobel, and the mansion-house where Alfred Nobel lived his last years, and where he wrote his famous testament.

From 2004, five years have elapsed, and today Roboethics is a subject of authoritative discussion and studies; it is the topic of an ad hoc IEEE Robotics&Automation Technical Committee, and headline of many books.

In the next decades in the Western world – in Japan, United States, Europe – humanoids robots will be among us, companions to elderly and kids, assistants to nurse, physicians, firemen, workers. They will have eyes, human voices, hands and legs; skin to cover their gears and brain with multiple functions. Often, they will be smarter and quicker than the people they ought to assist. Placing robots in human environments inevitably raises important issues of safety, ethics, and economics. Sensitive issues could be raised by the so called "robotics invasion" of many non-industrial application sectors, especially with the personal robot; and the surveillance and military applications.

In many instances, I have tried to demonstrate that Robotics is indeed a new science, of a special kind. And that in the making of this new science we can understand in-depth many new fields of physical disciplines, as well as of Humanities. In the main, Robotics is in fact considered a branch of Engineering dealing with intelligent, autonomous machines. It shares knowledge with other disciplines, and it is somehow the linear sum of all these studies. On the other side, some of us regard Robotics as new science, in its early stage. Ultimately – we say – it

is the first time that humanity is approaching the challenge to replicate a biological organism. That is why Robotics holds this special feature of being a platform where Sciences and Humanities are converging – an experiment in itself.

To discuss this matter, let us start from a question: How is a new science born?

Thomas Kuhn says that "under normal conditions the research scientist is not an innovator but a solver of puzzles, and the puzzles upon which he concentrates are just those which he believes can be both stated and solved within the existing scientific tradition".[1]

However, he adds in another locus of the same work, that " (..)I think, particularly in periods of acknowledged crisis that scientists have turned to philosophical analysis as a device for unlocking the riddles of their field. Scientists have not generally needed or wanted to be philosophers".[2]

Let us think of chemistry, of physics, sciences originating from many original and even weird sources, and later on systematized by famous scientists whose mission was to order the knowledge in laws, principles and rules, applying mathematical methodology to structuring the cluster of confirmed experiences and cases. Sciences are syncretic creatures, daughters of rationality, non rationality and of societal forces.

Back to Robotics. As said before, it is the result of melting knowledge from many fields: Mechanics, Automation, Electronics, Computer Science, Cybernetics, and Artificial Intelligence. It also stems from Physics & Mathematics; Logic & Linguistics; Neuroscience & Psychology; Biology & Physiology; Anthropology & Philosophy; Art & Industrial Design. And, the more it develops, the more it floods into other disciplines, exceeding schemes and borders. A proof of the complexity of robotics comes from the 1600 pages of the monumental "Springer Handbook of Robotics"[3], the first encyclopedic volume existing in the literature devoted to advanced robotics, edited by Bruno Siciliano and Oussama Kathib.

There is another important element of development, and it is the boost in robotics' applications, which in turn is controlled by the so-called forces of the market: huge investments are funneled into it, from Japan's Meti 40 billion yen in the humanoids challenge, to the 160 billion dollars in the US Future Combat Systems program.

We are just on the brink of the development of our science, and it is hard to envisage its future. It may

happen that Robotics swallows up other sciences; or that, like the giant red stars, it will explode into many other sciences, originating from the intersections of adjoining fields.

FIORELLA OPERTO

Robotics: Much talking about it, but little known. Actually, despite investments, efforts and results, penetration in our societies and media scoops, Robotics is a science which is still relatively unknown, or little known, and often misrepresented. Seldom is the keyword ***Robotics*** read in the institutional Programmes, being mainly hosted in the ICT cluster, or hidden under different initials.

Sometimes I linger to ponder the under-studied inferiority complex of some engineers which prevents them attributing universal qualities to their work. This so called inferiority feeling derives – as the Italian scholar of studies in history and philosophy of science, Paolo Rossi, says – from ancient times, when *mechanicus* meant a vile and not noble man. Paolo Rossi writes:

> "At the roots of the great scientific revolution of the 17th century is the union between technology and science that has marked, for the good and the not so, the entire Western civilization. This union, that became marked in the 17th and 18th centuries and which perpetrated all over the word, was, however, absent in ancient and medieval civilizations. The Greek term *banausia* means mechanical art or manual labor. In Plato's *Gorgia*, Callicle states that a machine manufacturer ought to be despised; insulted, by being called a banausos; and that no one would consent to the marriage of their daughter to him. Aristotle had excluded the mechanical workers from the citizens'society and had said that they differed from slaves only due to the fact that they care for many individuals' needs whilst a slave only cares for one. The divide between slaves and free individuals tended to be made manifest by the division between techniques and science, the division between practically-orientated knowledge and knowledge dedicated to the contemplation of truth. The disdain with which the slaves were treated was equally transferred to their areas of work. The seven liberal arts of the trivium (grammar, rhetoric and dialectic) and of the quadrivium (arithmetic, geometry, music and astronomy) are so named liberal due to their belonging to free individuals, and not to the non free individuals, or to the slaves who practiced mechanical or manual arts. Knowledge not directed towards a specific end but collected for its own intrinsic value is the only key to discovering the true nature of humankind. The practice of

sophia requires wealth and the presence of life's fundamentalities. Philosophy needs the mechanical arts upon which it is based, however, they are inferior forms of knowledge that are immersed between the material and the sensible and which are linked to manual and practical labor. The wise and learned individuals ideals tends to coincide (as it does in Stoic and Epicurean philosophy and later in Thomas Aquinas' thoughts) with the image of one who dedicates his life to contemplation while waiting for (like the Christian thinkers) the bliss of contemplating God".[4]

GIAMARCO VERUGGIO
In fact, it was in the Italian Renaissance that the profession, and word, "engineer", or, more precisely, geometrician, architect indicated a profession of equally importance as scientist, artist and of socially acknowledged leadership. Maybe, one of the reasons for this underestimate is the paradox of Engineering, which is, on the one hand, an arid, stark, abrupt and operative science; on the other side, the making of a craftsman, often of a true artist.

Even Ove Arup, the leading Anglo-Danish engineer, said that: "Engineering is not a science. Science studies particular events to find general laws. The projecting activity of the engineer uses those laws to solve particular problems. In this, it is closer to the art or handicraft: problems are under-defined and there are many solutions, good, bad and indifferent. The art is finding a good solution through a compromise between media and scopes. This is a creative activity, that requires imagination, intuition and deliberative choice."[5]

I believe that roboticists should get a sense of their creative potential and of the importance of their scientific contributions, with method and rigor. We are perhaps witnessing some hints of getting this sense.

I see, robotics is more known through media exaggerations and novelists' stories, Terminators and Wall-e robots?

GIANMARCO VERUGGIO
It is really true! And, here, you have another "siding-mission" of Roboethics.

In the 18th century, one of the aims of scientists working in the field of electromagnetism was to remove magic from the physical phenomena they were interpreting. And we roboticists have to do just that, freeing our field from the magical conception still dominant today in many layers of the population. We are suffering from the heavy burden of literature and fiction, which overimposes on our products their profile and patterns. A tough life for

Robotics: the less people know about it, the more they talk about, and demand from it.

The general population knows about Robotics what it watches in the Sci-Fi movies, which feed any kind of irrational fear about robots being disobedient, rebelling, good or evil souls, conscious and in love, emotional creatures lacking only the quality of total freedom. Robotics stimulates some of humanity's most fundamental questions. This means that we shall not expect some simple answers to be conclusive. The undertaking of discovering the essence and origin of human' intelligence and self-consciousness is as tough and troubling as the challenge around the unification of physical forces, or the research on the origin of the Universe. Simplistic answers could lead to gross mistakes and we cannot obtain correct answers if we ask the wrong questions.

I had hard times witnessing discussions on the rights for robots; on robotics' superiority to humans; on the development of robots to other biological, top-dog species. The sad side of the story is that often it is us, the roboticists, who are responsible of repeating, or fostering such legends, for narcissism or fashion of being philosophers. I believe that we have to use clear thinking, from now on. We would need other myths, images, and metaphor, which are truly intrinsic and proper to Robotics, and not to the anthropology of the human/automata tragedy and legend. Real robotics is far more exciting that fantasy!

For instance, one of the problems to be addressed, with the support of scholars of humanities, is that in robotics, current uses of words such as knowledge, intelligence, representation, intention, emotion, social agent, autonomy, and humanoid are potentially misleading – insofar as it is thereby suggested that typically human mental properties can be indifferently and unproblematically attributed to technological artifacts, disregarding from the current limitations of state-of-the art robotic systems.

But, ultimately, what is a robot?

GIANMARCO VERUGGIO
A robot is an autonomous machine that is capable of performing a variety of tasks, gathering information about its environment (senses) and using it to follow instructions to do work. Nothing really romantic about it! On the other side, robots are the machines which are more similar to humans than anything we've ever built before, and this makes it easier, for people who don't know the subject, to speak about robots as if they were humans.

This peculiarity has favored the rise of all the legends about robots: That

they will rebel against humankind; that they can "evolve" becoming humans, super-humans, and so on. One day, we could also be witnessing the birth of weird "robot worshipper" sects claiming some nutty visions about robots. But, this misconception could also be generating suspects and diffidence in the traditional cultures that could in turn lead them to raise obstacles to the research and development of the most advanced robotics applications.

Let us discuss concisely one of the most popular myth: the one we could call *Pinocchio principle*, that is the idea that humanoid robots could evolve into humans. Basically, the legend embodied in the *Pinocchio principle* is that reproducing ever more perfectly the human functions coincides with producing a human being. Although it is picked up by many scholars, we recognize in it an acknowledged flaw of reasoning and of composition. In fact, even if we could design and manufacture a robot endowed with symbolic properties analogous to those of humans, the former would belong to another, different species.

Actually, human nature is not only the expression of our symbolic properties, but also the result of the relationships matured during our extra uterine development (we are Nature AND Culture). There is a very important concept that is ***embodiment***, which means that an intelligence develops in a body and that its properties cannot be separated by it. A very enlightening article was written by José Galvan in the December 2003 issue of IEEE Robotics & Automation Magazine, "On Technoethics", where it is said, among other things: "The symbolic capacity of man takes us back to a fundamental concept which is that of free will. Free will is a condition of man which transcends time and space. Any activity that cannot be measured in terms of time and space can not be imitated by a machine because it lacks free will as the basis for the symbolic capacity".

It is quite obvious that when a machine displays an emotion, this doesn't mean that it feels that emotion, but only that it is using an emotional language to interact with the humans. It is the human who feels emotions, not the robot! And attributing emotions to the robot is precisely one of these human emotions.

We humans understand the world around us (people, nature, or artifacts) through emotional interaction. Long interaction can result in attachment (it may also provoke boredom). Interaction stimulates humans, and generates motivations for behaviour. Human interaction with the world always involves emotions.

There are useful objects and aesthetic objects, each of them evoking

different emotions in humans. Machines are also artifacts. Different from the aesthetic objects, machines have been designed and developed as tools for human beings. But usually, machines are passive, so human interaction with them is limited. But when a machine expresses its ability to act in a semi-voluntary way (as in the case of robots, which have been designed and programmed to learning), they have much influence on human emotions because the machine's behaviors may be interpreted by humans as emotional and voluntary. Furhermore, a machine with a physical body is more influential on the human mind than a virtual creature.

In the field of human-robot interaction, there are many studies on all these topics. Mit's kismet is one; also, all the projects involving robot pet therapies (for instance, the robot Paro, designed by Japanese roboticists Takanory Shibata[6], or those robotic playmates which can help children with autism.

It is a truly complex field, because the results depend very much on the cultural context and on the-background of the human actors involved.

From the Roboethics point of view, the sensitive issues concern the human right for dignity and privacy. In the case of robots employed by children, the concern pertains to the domain of the relationship of the kids with the world, their ability to distinguishing robot from living creatures and the danger of technological addiction.

In no way, however, in my opinion, a robot can "feel" emotion, at least, not in the way we do it.

Much of the mess about the robot's consciousness, robot's emotions, and robot's rights are based on the confusion generated by the use of the same words for intrinsically different items. That's why, discussing with philosophers in Europe and United States, we agreed that it could be worth expressing these ontological differences through a specific notation.

This is not a very original idea! For instance, in mathematics, the estimate of the variable x (exact or "truth" value) is referred to as "x hat", while its measure is indicated as "x tilde".

I am an engineer, and I am talking as a scientist, aiming at applying – when reasoning about philosophy of science – the same rigor I should employ in my daily work. For this, I would propose that we indicate with a "star" the properties of our artifacts, to distinguish them from those of the biological beings.

- HUMANS have INTELLIGENCE
- ROBOTS have INTELLIGENCE* (INTELLIGENCE STAR)

This could be a first, very simple way to keep us aware of these ontological differences, and at the same time it can help in avoiding flaws in our reasoning, like this:

- DOGS have four legs
- The THING that I see here has four legs

Therefore

- The THING that I see here is a DOG

For the sake of thruth, it necessary, even when we discuss the philosophy of our science, that we engineers apply the same sharpness as Galileo recommended in his synthesis of the Scientific Methodology: "Necessary demonstrations and sense experiences".

From all this, the necessity for the robotics community to become the author of its own destiny, in order to tackle directly the task of defining the ethical, legal and societal aspects of their researches and applications. Of course not alone, but collaborating with academics in the field of philosophy, of law, and general experts of human sciences. Nor should we feel relegated to a merely techno-scientific role, delegating to others the task of reflecting and taking action on moral aspects. At the same time, it is necessary that those not involved in robotics keep themselves up to date on the field's real and scientifically predictable developments, in order to base the discussions on data supported by technical and scientific reality, and not on appearances or emotions generated by legends.

I understand, from what you said, that Roboethics is, in your view, more than some deontological guidelines for designers and users?

GIANMARCO VERUGGIO

Roboethics is not the "Ethics of Robots", nor any "ethical chip" in the hardware, nor any "ethical behavior" in the software, but it is the human ethics of the robots' designers, manufacturers and users. In my definition, "Roboethics is an applied ethics whose objective is to develop scientific – cultural – technical tools that can be shared by different social groups and beliefs. These tools aim to promote and encourage the development of Robotics for the advancement of human society and individuals, and to help preventing its misuse against humankind.

Actually, in the context of the so-called Robotics ELS studies (Ethical, Legal, and Societal issues of Robotics) there are already two schools". One, let us refer to it as "Robot-Ethics" is studying technical security and safety procedures to be implemented on robots, to make them as safe as possible for humans and the plant. Roboethics, on the other side, which is my position, concerns itself with the global ethical studies in Robotics and is a human ethics.

FIORELLA OPERTO

Roboethics is an applied ethics that refers to studies and works done in the field of Science&Ethics (Science Studies, S&TS, Science Technology and Public Policy, Professional Applied Ethics), and whose main premises are derived from these studies. In fact, Roboethics was not born without parents, but it derives its principles from the global guidelines of the universally adopted applied ethics This is the reason for a relatively substantial part devoted to this matter, before discussing specifically Roboethics' sensitive areas.

Many of the issues of Roboethics are already covered by applied ethics such as Computer Ethics or Bioethics. For instance, problems – arising in Roboethics – of dependability; of technological addiction; of digital divide; of the preservation of human identity, and integrity; the applications of precautionary principles; of economic and social discrimination; of the artificial system autonomy and accountability; related to responsibilities for (possibly unintended) warfare applications; the nature and impact of human-machine cognitive and affective bonds on individuals and society; have already been matters of investigation by the Computer ethics and Bioethics.

A few lines about the "history" of Roboethics can be useful here to understand its aims and scope.

In January 2004, Veruggio, myself, in collaboration with roboticists and scholars of humanities organized the First International Symposium on Roboethics (Sanremo, Italy). Its aim was to open a debate, among scientists and scholars of Sciences and Humanities, with the participation of people of goodwill, about the ethical basis, which should inspire the design and development of robots.

Philosophers, jurists, sociologists, anthropologist and moralists, from many world's Nations as well as robotic scientists, met for two days contributing to lay the foundations of the Ethics in the design, development and employment of the Intelligent Machines, the Roboethics.

In 2005, EURON (European Robotics Research Network) funded the Research Atelier on Roboethics (project leader was School of Robotics) with the aim of developing the first Roadmap of a Roboethics. The workshop on Roboethics took place in Genoa, Italy, 27th February – 3rd March 2006. The ultimate purpose of the project was to provide a systematic assessment of the ethically sensitive issues involved in the Robotics R&D; to increase the understanding of the problems at stake, and to promote further study and transdisciplinary research. The Roboethics Roadmap – which was the result of the Atelier and of the following discussions and dissemination –

outlines the multiple pathways for research and exploration in the field, and indicates how they might be developed. The Roadmap embodies the contributions of more than 50 scientists, scholars and technologists, from many fields of science and humanities. It is also a useful tool to design a robotics ethic trying to embody the different viewpoints on cultural, religious and ethical paradigms converging on general moral assessments.

In the meantime, in the frame of the IEEE Robotics&Automation Society was organized a Technical Committee on Roboethics which is currently co-Chaired by Atsuo Takanishi, Matthias Scheutz, and Gianmarco Veruggio.

GIANMARCO VERUGGIO
One of the most ambitious aims of Robotics is to design an autonomous robot that could reach – and even surpass – human intelligence and performance in partially unknown, changing, and unpredictable environments. Artificial Intelligence will be able to lead robots to fulfil the missions required by the end-users. To achieve this goal, over the past decades roboticists have been working on AI techniques in many fields.

From this point of view, let us consider the fact that one of the fundamental aspects of the robots is their capability to learn: to learn the characteristics of the surrounding environment, that is, a) the physical environment, but also b) the living beings who inhabit it. This means that robots operating in a given environment have to distinguish human beings and living creatures from inorganic objects.

In addition to performing a learning capability about the environment, robots have to understand their own behaviour, through a self reflective process. They have to learn from the experience, replicating somehow the natural processes of the evolution of intelligence in living beings (synthesis procedures, trying-and-error, learning by doing, and so on).

All these processes embodied in the robots produce an intelligent machine endowed with the capability to express a certain degree of autonomy. It follows that a robot can behave, in some cases, in a way, which is unpredictable for their human designers. Basically, the increasing autonomy of the robots could give rise to unpredictable and non predictable behaviours.

Without necessarily imagining some Sci-Fi scenarios, in a few years we are going to be cohabiting with robots endowed with self knowledge and autonomy – in the engineering meaning of these words. This means, for instance, that we could have to impose limits – up to a certain extent –

on the autonomy of the robots, especially in those circumstances in which robots could be harmful to human beings.

In our roboetics studies, we have taken into consideration – from the point of view of the ethical issue connected to Robotics – a time range of a decade, a frame in which it could reasonably be located and inferred – on the basis of the current State-of-the-Art in Robotics – certain foreseeable developments in the field. Moreover, for the above mentioned reason, we have considered it premature to deal with problems inherent in the purely hypothetic emergence of human functions in the robot: like consciousness, freewill, self-consciousness, sense of dignity, emotions, and so on. Consequently, this is why the Roadmap does not examine problems like the need not to consider robots as our slaves, or the need to guarantee them the same respect, rights and dignity we owe to humans. I am convinced that, before discussing robot's rights, we have to ensure human rights to all the human beings on earth.

For instance, we have felt that problems like those connected to the application of robotics within the military and the possible use of military robots against some populations not provided with this sophisticated technology, as well as problems of terrorism in robotics and problems connected with biorobotics, implantations and augmentation, were pressing and serious enough to deserve a focused and tailor-made investigation. It is clear that without a deep rooting of Roboethics in society, the premises for the implementation of artificial ethics in the robots' control systems will be missing.

How can you envisage the definition of a Roboethics guideline protocol, which has to be shared by different cultures?

GIANMARCO VERUGGIO

Roboethics is a work in progress, susceptible to further development and improvement, which will be defined by events in our techno-scientific-ethical future. We are convinced that the different components of society working in Robotics, interested people and the stakeholders should intervene in the process, in a grassroots science experimental case: the Parliaments, Academic Institutions, Research Labs, Public ethics committees, Professional Orders, Industry, Educational systems, the mass-media.

To achieve this goal we need an internationally open debate because, concerning the role of science and technology in law, politics, and the public policy in modern democracies, there are important differences between each of the

European, the American, and the - we could say - oriental approach. But we live in the Age of Globalization and robotics will have a global market, just like computers, videogames, cars or cameras.

In the United States, the general attitude is definitely more science-based than it is in Europe. In the former case, science is said to speak the truth, and the regulatory process is based more on objective scientific data than on ethical considerations. At the same time, the subjective point of view is taken up by the courts, which are now also intervening directly in areas such as risks in society and scientific knowledge, although the current conceptual tools of jurisprudence in the field of science&technology are still very limited. Nonetheless, in the Anglo Saxon culture, “law does not speak the language of science”.

On the other side, in Europe, in the frame of the ongoing process of the culture’s cohesion, the course of regulation and legislation of science and technology assume a character of the foundation of a new political community - the European Union, which is centred around the relationship between science and its applications, and the community formed by the scientists, the producers, and the citizens. We can safely assume that, given the common classical origin of jurisprudence, the latter process could be helpful in influencing other cultures, for instance, the moderate Arab world.

There is a third way to approach issues in science&society it could be called oriental. In fact, in Japan and in the Republic of South Korea, issues of robotics&society have been handled more smoothly and pragmatically than in Europe and in America. Due to the general confidence from their respective societies towards the products of science&technology, the robotics community and the ad hoc ethical committee inside these governments have started to draw up guidelines for the regulation of the use of robotic artefacts. This non-ideological, non-philosophical approach has its pros and cons, but it could encourage scientists and experts in Europe and the United States to adopt a more normative position.

This means that also Roboethics – which is applied ethics, not theoretical – is the daughter of our globalised world. An Ethics which could be shared by most of the cultures of the world, and capable of being translated into international laws that could be adopted by most of the nations of the world.

While we analyze the present and future role of robots in our societies, we shall be aware of the underlying principles and paradigms which influence social groups and single individuals in their relationship with

intelligent machines. Different cultures and religions regard differently the intervention on sensitive fields like human reproduction, neural therapies, implantations, and privacy. These differences originate from the cultural specificities towards the fundamental values regarding human life and death. In different cultures, ethnic groups and religions the very concept of life and human life differs, first of all concerning the immanence or transcendence of human life. While in some cultures women and children have fewer rights than adult males (not even the habeas corpus), in others the ethical debate ranges from the development of a post-human status to the rights of robots. Thus, the different approach in Roboethics concerning the rights in Diversity (gender, ethnicity, minorities), and the definition of human freedom and Animal welfare. From these concepts, other specificities derive such as privacy, and the border between privacy and traceability of actions.

Cultural differences also emerge in the realm of natural vs. artificial. Think of the attitude of different peoples towards the surgical implants or the organs implantation. How could human enhancement be viewed? Bioethics has opened important discussions How is the integrity of the person conceived? What is the perception of a human being?

Last, but not least, the very concept of intelligence, human and artificial, is subject to different interpretations. In the field of AI and Robotics alone, there is a terrain of dispute– let's imagine how harsh could it be outside of the circle of the inner experts.

Because we said that there are big differences in the way the human-robot relationship is considered in the various cultures and religions, only a large and lengthy international debate will be able to produce useful philosophical, technical and legal tools.

At a technical level we need a huge effort by the standard committees of the various international organizations, to achieve safety standards, just like for any other machine or appliance.

At a legal level we need a whole new set of laws, regulating for instance the mobility of robots in the place of work or in public spaces, setting clear rules about the liability and accountability of their operations.

At a philosophical and ethical level, we need to discuss in depth the serious problem of the lethality of robots. This means that humankind has to decide if the license to kill humans should be issued to robots, for instance in military applications.

This is precisely the mission that led us to start and to foster the Roboethics Programme, and to develop

the Roboethics Roadmap. The basic idea was to build the ethics of robotics in parallel with the construction of robotics itself.

Actually, the goal was not only to prevent problems or equip society with cultural tools with enough time to tackle them, but a much more ambitious aim. Indeed I feel that robotics' development is not so much driven by inexistent abstract laws of scientific/technical progress, but moreover by complex relations with the driving forces of the economic, political and social system. And therefore dealing with roboethics means influencing the route of Robotics.

It is certainly a great responsibility, which however cannot be avoided. As the American roboticist George Bekey says : <We roboticists must walk to the future with our eyes wide open>. Indeed in society there cannot be a "Non-choice" stance.

Abstention ultimately ends up favoring the strongest, and in our case, in the current world political, social and economic system, this means one thing only: a development policy driven by the interests of multinational corporations. And, as the French roboticist Philippe Coiffet says: <A development in conformity with a Humanist vision is possible but initiatives must be taken because "natural" development driven by the market does not match with the desired humanist project.>[7]

From the ethical point of view, which kind of approach have you selected in structuring the fundamentals of the ethical issues in robotic?

GIANMARCO VERUGGIO
Given the relative novelty of the ELS issues in Robotics, the recommended ethical methodological approach here is that of the Applied Socio-Ethics.

Lacking an existing body of ethical regulations related to ethical issues in Robotics, scholars in the field (Tamburrini, Capurro et al., 2007) have proposed to sort a high value selection of case-studies in the most intuitively sensitive field on robotics applications (learning robots and responsibility, military robotics, human-robot interaction, surgery robotics, robotics cleaning systems, biorobotics). These cases were analyzed from the following point of view:

a) a technoscientific analysis (risk assessment; stability, sustainability and predictability); dependability assessment;
b) Shared ethical assumptions: liberty, human dignity, personal identity, moral responsibility and freedom (European Charter of Fundamental Rights; UN Chart of Human Right and related documents);

c) General Cultural assumptions (the way we live in Europe, our shared values and future perspectives, the role of technology in our societies, the relationships of European citizenship to technology and robots; our shared notions of social responsibility, solidarity and justice). Successively, a cross- check analysis was carried out between techno-ethical issues and ethical regulations.

Let us look at one case. In the field of service robots, we have robot personal assistants, machines which perform tasks from cleaning to higher tasks like assisting elderly, babies, disabled people, students in their homework, to the entertainment robots. In this sector, ELS issues to be analyzed concern the potection of human rights in the field of human dignity, privacy, the position of humans in control hierarchy (non instrumentalization principle). The right to human dignity implies that no machine should be damaging a human, and it involves the general procedures related to dependability. From this point of view, robotics personal assistants could raise serious problems related to the reliability of the internal evaluation systems of the robots, and to the unpredictability of robots' behavior. Another aspect to be taken into account, in the case of autonomous robots, is the possibility that these were controlled by ill-intentioned people, who can modify the robot's behavior in a dangerous and fraudulent manner. Thus, designers should guarantee the traceability of evaluation/actions procedures, and the identification of robots.

On a different level, we have to tackle the psychological problems of people who are assisted by robots. Lack of human relationships where personal connections are very important (e.g. for elderly care or edutainment applications) and general confusion between natural and aryificial, plus technological addiction, and loss of touch with the real world – in case of kids – are some of the psychological problems involved.

FIORELLA OPERTO

We can underline other kinds of ethical issues involving personal robots. For instance: The emerging market of personal service robots is driving researchers to develop autonomous robots that are natural and intuitive for the average consumer who can interact with them, communicate, work and teach them. Human-Robot interaction is developing along the innovative field of the so-called "emotional" or "social" robots, capable of expressing and evoking emotions. These social robots (employed especially in education, edutainment, care, therapy, assistance or leisure) are produced for the average non-expert consumer, and are supposed to display "social" characteristics

and competencies, plus a certain level of autonomous decision-making ability. They are endowed with: a) natural verbal and non-verbal communication (facial expressions, gestures, mimicking); b) embodiment (that is, in our case, how the internal representations of the world are expressed by the robots' body) and social situatedness; and emotions.

In the process of modelling human schemes of emotions, facial expressions and body language are often used gender, race and class stereotypes drawn from the approach of the empiricist psychology school. From the point of view of ethcal issues in robotics, it should be considered, and possibly avoided, to adopt the discriminatory or impoverished stereotypes of, e.g., race, class, gender, personality, emotions, cognitive capabilities, and social interaction.

The Institut für Religion und Frieden – which is the Editor of this booklet – is promoting a survey on one of the main sensitive aspect of robotics' applications – and of Roboethics: Military robotics. I am aware that you have intervened several times on this issue?

GIANMARCO VERUGGIO

Military research in robotics is being extensively supported, both in the United States and in Europe. Ground and aerial robotic systems have been deployed in warfare scenarios. It is expected that an increasing variety and number of robotic systems will be produced and deployed for military purposes in many developed countries.

While the design and development of autonomous machines opens up new and never faced issues in many fields of human activity, be they service robots employed in caring people (robots companion), or robots used in health care, those autonomous machines employed in war theatres are going to raise new and dramatic issues.

In particular, military robotics opens up important issues of two categories:

a) Technological;
b) Ethical.

Concerning technological issues, these are managed under the so-called Dual Use. Dual Use goods and technologies are products and technologies which are normally used for civilian purposes but which may have military applications. The main legal basis for controls on Dual-Use Goods is the EU Dual-Use Regulation (also known as Council Regulation 1334/2000 to be repealed by Council Regulation 428/2009, adopted 5 May 2009 and published in the OJ of the EU on 29 May 2009, L 134.) (European Commission, External Trade).

In the case of robotics machines, their behavour is affected by issues regarding the uncertainty of the stability of robot sensory-motor processes and other uncertainty questions. For this reason, in robotic systems that are designed to interact with humans, stability and uncertainty issues should be systematically and carefully analyzed, assessing their impact on moral responsibility and liability ascription problems, on physical integrity, and on human autonomy and robotic system accountability issues.

Actually, in modern robots the algorithms governing their learning and behavioral evolution, associated with operational autonomy, give rise intrinsically to the inability to forecast with the needed degree of accuracy each and all the decisions that the robot should take, under the pressure of the operational scenario in which it is employed at that moment.

This window of unpredictability is a well-known issue appearing in every robotics application field; but it involves some dramatic implications when applied to military robotics.

In this field, in fact, we have not only important ethical and humanitarian considerations, but also questions of operational reliability and dependability.

The very same military milieus have several times underlined the danger implied by the lack of reliability of robotics systems in a war theatre, especially when the urgency of quick decisions and the lack of clear intelligence concerning the situation requires the maximum control over its own forces.

This is particularly evident when human-in-the-loop conditions jeopardize timely robotic responses, possibly leading on this account to violations of task constraints and increased risk conditions. In view of current limitations of robotic technologies, robots do not achieve human-level perceptual recognition performances that are crucial, e.g., to distinguish friends or by-standers from foes.

In shaping responsibility ascription policies one has to take into account the fact that robots and softbots – by combining learning with autonomy, pro-activity, reasoning, and planning – can enter cognitive interactions that human beings have not experienced with any other non-human system (Tamburrini, Marino, 2006)

The issue is worsened by the extraordinary complexity of the robot's artificial intelligence control system. This issue makes these intelligent machines vulnerable from the point of view of their software's reliability. We all know, in fact, that no program is free from bugs affecting its behavior. Now, it's one thing when a bug is affecting a word processor

program, but it is different when a program's bug on a robot endangers the human lives the robot is supposed to protect.

The other side of the issues – also stressed by military spokesmen – in military robotics is the high risk of information security gap. Autonomous robot employed in war theatres could be intruded, hacked, attacked by viruses of several types, and become an enemy's tools behind our back.

In some cases, a responsibility gap could also arise, when human adaptation to service robots could cause some phase displacements in human's behavior whose consequences should be carefully considered. The beneficial possibilities provided by robotics remotely and tele operations; by robots serving as human avatars in inaccessible and dangerous areas; the availability through robots to intervene in micro and nanometer ranges could induce in humans the rise of gaps in responsibility (because of the perceived shared responsibility between human and robot) which could lead to disengagement from ethical actions); a gap in knowledge (the so called "videogame syndrome", that is when an operator perceives reality like in a video game), and gaps in actuality and reality.

The second categories of issues are of ethical and social class.

Human life has so high a value to justify a war and to accept the sacrifice of one or more lives to protect a human community.

However, just for this reason, the extraordinary importance and seriousness of the issues has imposed that in civilized societies only and always human beings can decide on the destiny of other human beings, and not automatic mechanisms, as sophisticated as they might be.

Only human beings endowed with the power of reasoning and of free will are endowed with the power of moral responsibility.

Ethical reflection does not justify the exceptions rule that every individual robotic action be submitted to human supervision and approval before its execution.

It is recommended that in human-robot shared action control provisions be made for assigning humans the higher rank in the control hierarchy which is compatible with cost-benefit and risk analyses. Furthermore, it is recommended that robotic systems which are justifiably allowed to override human decisions or to act independently of direct human control or supervision be systematically evaluated from an ethical viewpoint. (Eticboths project, deliverable 5)

For all these consideration, although very briefly summarized, I am deeply

convinced that to attribute a "license to kill" to a robot is a decision of such an extreme gravity that no Nation or community alone can do it by itself. This question must be submitted to a deep and thorough international debate

The further development of a broad ethical framework as an enabling factor for the public to participate in discussions on dual use of robots is highly desirable, together with deliberative technology assessment procedures (for example consensus conferences) backed by technologically informed education initiatives. Suitable policies and actions fostering awareness about the dual use robots are highly recommended at the level of European society. Support of extensive initiatives in dual use problem dissemination and interdisciplinary techno-ethics community building is recommended too.

I am also deeply convinced that an "R" (robot) chapter should be added to the NBC treaties, discussing the guidelines for the use of robots in war theaters. As in the case of many new weapon systems, also in our case, military robotics, we will be witnessing many political, social, and philosophical stands. From the "ban the bomb" (there will be people fighting for "ban robot weaponry" or, "ban the killer robots") to all the nuances of military agreement's proposals, as we have had for the ABC weapons.

[1] Kuhn, Th. The Essential Tension. Tradition and Innovation in Scientific Research, The Third University of Utah Research Conference on the Identification of Scientific Talent, ed. C. W. Taylor Salt Lake City: University of Utah Press 1959.
[2] Kuhn Th., idem.
[3] Springer Handbook of Robotics, Siciliano, Bruno; Khatib, Oussama (Eds.), 2008.
[4] Paolo Rossi, Daedalus sive mehanicus: Humankind and machines, Lecture at the Euron Atelier on Roboethics, Genoa, Feb-Narch 2006. In: http://www.scuoladirobotica.it/lincei/docs/RossiAbstract.pdf.
[5] Ove Arup,1895-1988 http://www.arup.com/arup/policies.cfm?pageid=1259.
[6] http://www.aist.go.jp/MEL/soshiki/robot/biorobo/shibata/shibata.html.
[7] Ph. Coiffet, Conference' speech, International Symposium on Roboethics, 30th - 31st January, 2004, Villa Nobel, Sanremo, Italy, "Machines and Robots: a Questionable Invasion in Regard to Humankind Development".

Ronald C. Arkin: Governing Lethal Behaviour

How and why did you get interested in the fields of robots and ethics?

Several things lead me into the field, some of which are discussed in length in my new book's preface[1].

The first event was the recognition that my research and that of my colleagues was moving out of our laboratories and into the battlefield at a remarkable pace, and that we as scientists must assume responsibility for the consequences of being involved in the creation of this new technology. The second was my participation in the First Symposium on Roboethics in 2004 in San Remo Italy, thereby gaining a broader perspective of the field, including presentations from the Pugwash Institute of Russell-Einstein Manifesto fame, representatives of the Geneva Convention, and the Vatican among many others. I then started writing and presenting my ideas not only in technical conferences but also in philosophical and sociological venues that served to sharpen my thoughts on the subject. My subsequent involvement in our chief professional society (IEEE Robotics and Automation) added momentum, with my co-founding of the Technical committee on Roboethics, and also serving as co-chair of the Human Rights and ethics committee, and Liaison to the IEEE Social Implications of Technology Society. I developed an ethics course for our undergraduates at Georgia Tech entitled Robots and Society which satisfied their ethics requirement and provided fertile interactions from a diverse set of student backgrounds. Finally, my viewing of a particular battlefield video at a Department of Defense workshop that is described and used as a scenario in my book was a tipping point for providing impetus on learning how autonomous systems may be able to perform better than human soldiers in certain circumstances. I wanted to help ensure that robots would make better decisions in the battlefield than these warfighters did.

In the last two decades robots transcended from the production lines into the human society. The use of robots begins to span from the entertainment industry as far as to the care for the elderly. Do you think robots will have a similar impact on the human society as the revolution in telecommunications had?

Certainly the potential is there. I expect that robotics technology will be ubiquitous in the not too distant future, and as telecommunications has already shown, it will significantly affect the ways in which we interact with each other. The future impact on the social fabric from the advent of a robotics age is not yet understood, although we continue to plow ahead unchecked in many different technological venues: warfare, elder and child care, intelligent prostheses, and intimate robotics to name just a few. There is virtually no guidance from an ethical perspective for researchers in our field, unlike bioethics for example.

It seems that since 2004 (First International Symposium on Roboethics in Sanremo, Italy) roboethics at least are seen as an issue among engineers and philosophers alike. What do you think should be done, and having in mind the momentum of the industry, what can be done to meet the challenges in the fields you mentioned?

The most pressing need is continuing international discussion and additional forums to present the emerging issues associated with this new technology, not only to researchers and philosophers, but also social scientists, regulators and policy makers, and the general public among others. These discussions should not be tinged with fear or hysteria, which unfortunately often happens in the lay press, but rather examine these issues carefully and thoughtfully.

It is too early to be overly prescriptive, but we need informed and broad thinking on the subject. I am hopeful that additional venues will be found to lead to some sort of guidelines/regulations/doctrine that can be applied to the field as a whole and specific sub-disciplines as required.

From Karel Čapek's RUR to James Cameron's Terminator, Robots are often portrayed as a menace to human society. Do you think that this is inherent to our concept of robots and therefore influences our view of future developments in this sector? And are there differences between the United States and Europe on the one side and Asian countries like Japan and Korea on the other, where it seems that the society is more open to different roles of robots?

There have been several presentations and papers on this subject by my colleagues in Europe and Japan on this phenomenon. It is clear that the perception of robotics varies from culture to culture with a more Frankensteinian perspective in the West to a more friendly or benign view in the East. Some of this is believed to be related to religious heritage from these different regions, while other influences are

more overt such as Hollywood robots that end up destroying everything in sight (the logic extension of RUR) in contrast to the friendly society-saving robots (e.g., Astroboy, the main character of a Japanese Comic series by Osamu Tezuka from 1952 to 1968, which later on has also been produced for television and in 2009 by Imagi Animation Studios for cinema) of Japan.

This does not affect the research agendas perhaps as much as it does public opinion. The way the popular press treats robotics is remarkable in the U.S. and Europe with its tendency to use pathos and fear mongering blurring the line with objective and factual reporting. While it is important to make people aware of the potential dangers associated with this new technology, it is not fruitful to use Hollywood and other forms of fiction as examples of real consequences of this research. In the East, in contrast, it seems the potential ill effects of robotics are not adequately considered in the lay press, which is also disconcerting.

You have recently spoken at the IEEE (Institute of Electrical and Electronics Engineers) International Conference on Robotics and Automation in Kobe, Japan[2], where a workshop was dedicated to roboethics. Though most of the systems currently in use by the military are in most cases teleoperated, in the last couple of years the question of (semi)autonomous military robots and the ethical implications of their deployment have become of interest to a broader public. Do you see this as something that will have to be dealt with in the foreseeable future?

Certainly. These systems must be designed to comply with international treaties and accords regarding the conduct of warfare. I believe this is feasible and has the potential to lead to the design of autonomous robots capable of lethal force that can outperform human soldiers with respect to ethical conduct in certain constrained situations.

Contrary to a common belief, that robots lacking the capacity of being emotional are a problem, you have argued that being not an emotional being and therefore not being exposed to stress, anger and fear will give robots the potential to behave more ethically than human soldiers. Should the programming of robots be therefore clearly restricted and should we even strive for autonomous systems with emotionlike qualities, as current research does?

We have and are continuing to study the role of emotions in robotics in general, and there is a time and place for them. This is generally, however, not in the battlefield, especially regarding fear and anger

when associated with the use of lethal force. Nonetheless, we are investigating the role of guilt for battlefield robots as a means of reducing unintended collateral damage and we expect to have a technical report available on that subject shortly. For non-military applications, I hope to extend this research into a broader class of moral emotions, such as compassion, empathy, sympathy, and remorse, particularly regarding the use of robots in elder or childcare, in the hopes of preserving human dignity as these relationships unfold in the future.

There is also an important role for artificial emotions in personal robots as a basis for establishing meaningful human-robot interaction. Having worked with the use of artificial emotions for Sony Corporation on their AIBO and QRIO robots, and now with Samsung Corporation for their humanoid robots, it is clear that there exists value for their use in establishing long-term human-robot relationships. There are, of course, ethical considerations associated with this goal due in part to the deliberate fostering of attachment by human beings to artifacts, and a consequent detachment from reality by the affected user. This may also result in the displacement of normal human-human relationships as a by-product. Currently most researchers view this as a benign, or perhaps even beneficial effect, not unlike entertainment or video games, but it can clearly have deleterious effects if left unchecked, and needs to be further examined.

Your study "Governing Lethal Behavior: Embedding Ethics in a Hybrid. Deliberative/ Reactive Robot Architecture" (and the resulting book: Governing Lethal Behavior in Autonomous Robots, 2009) was the first in depth work tackling ethical aspects of the use of military robots on a practical level. What were your motivations and aims for this study and what were the proceedings?

The research in this study was funded by the U.S. Army Research Organization. It was an outgrowth of discussions I had with the military regarding the potential consequences of autonomous systems in the battlefield and I was fortunate enough to have my proposal funded to conduct this work. The motivation was largely to explore the thesis that robotic warfare may lead to a potential reduction in non-combatant casualties as well as a concomitant reduction in other forms of collateral damage, ideally without erosion of mission performance. This is not to say that the systems can be designed to perform in all situations that human soldiers can function; far from it. But rather for certain future situations (e.g., counter-sniper or urban building clearing operations), and although not perfectly, these systems,

due to inherent human failings in the battlefield, will be able to act more conservatively and process information more effectively than any human being possibly could, given the ever increasing battlefield tempo. It is proposed only for future warfare usage, not for existing military engagements or counterinsurgency operations.

The results of this study include the theoretical mathematical formalisms underpinning the approach, the design of a robot architecture potentially capable of ensuring that lethal force is consistent with international ethical requirements in bounded military situations, and some preliminary implementation and testing of the ideas in simulation. These are all documented in the new book.

In the analysis of the Future Combat Systems of the United States Army[3], you have identified the roles of robots either as an extension to the warfighter or as an autonomous agent. How do these two cases differ from each other and what are their ethical challenges?

As defined in our report (which did not deal specifically with the Future Combat System):

- Robot as an extension of the warfighter: a robot under the direct authority of a human, including authority over the use of lethal force.
- Autonomous robot: A robot that does not require direct human involvement, except for high-level mission tasking; such a robot can make its own decisions consistent with its mission without requiring direct human authorization, including decisions regarding the use of lethal force.

The primary difference lies in the locus of responsibility for the robots' actions. The ethical challenges surround the attribution of responsibility if or when things go wrong, an important criterion in Just War theory. If a mistake or a war crime occurs during the use a robot acting as an extension of the warfighter, it seems much more straightforward to blame the actual operator of the system, although this may still be unfair depending upon how the system was designed and other factors. In the case of an illegal action by an autonomous robot, some may choose to try and blame the robot itself when things go wrong. In my opinion, this cannot be the case, and the fault will always lie with a human being: either the user, the commanders, the politicians, the designers, the manufacturers, the scientists, or some combination thereof. Thus responsibility attribution is more difficult to achieve in the autonomy case, but not impossible in my opinion. The design of our software architecture includes a responsibility advisor that strives

to make this attribution process as transparent as possible.

In your research an "Ethical Governor" is designed to regulate behaviour and to ensure that the acts of the robot are ethical. Can we imagine this system similar to the superego of the human mind?

The ethical governor was inspired by James Watts' governor for steam engines (originally designed in 1788), rather than a model of the human mind. His governor was used to ensure that a machine remained within acceptable operational bounds so that it wouldn't destroy either itself or its surroundings. In this sense there is a parallel with the ethical governor, which instead of regulating the speed of the engine, instead regulates the behavioural output of the unmanned system to fall within the bounds prescribed by the internationally agreed upon Laws of War and the specific Rules of Engagement for a given mission.

If military robots are discussed, the focus is often put on systems with the potential to apply lethal force. Do you see a trend towards the use of unmanned systems in lethal and non-lethal engagement (e.g. robots with Taser like weapons)? And will it be functional or even be possible to really "keep the human in the loop"?

Yes, there is a clear trend towards autonomous unmanned weaponized robotic systems. (By the way, there really are no non-lethal weapons, they are rather less-lethal weapons. Tasers have been implicated in the deaths of many people). Due to an ever-increasing battlefield tempo, which has accelerated to the point where humans are hardly capable of making rational and deliberate decisions regarding the conduct of warfare when under fire, there appears to be little alternative to the use of more dispassionate autonomous decision-making machinery. Autonomy will move closer and closer to the "tip of the spear" and target selection will involve less human authority over time. Humans will still be in the loop, but at a highly supervisory level, such as "take that building using whatever force is necessary".

While the traditional military mantras of "warfighter multiplication, expanding the battlespace and increasing the warfighter's reach" will drive the use of autonomous systems to enhance mission effectiveness, there are also manifold reasons to justify their potential use on ethical grounds as well, in my opinion, potentially leading to a reduction in collateral damage and noncombatant casualties.

Given the progressing developments in sensor technology and physiological and behavioural biometrics, projects like the Future

Attribute Screening Technology (FAST) Project of Homeland Security come to mind, do you think that these approaches will have some impact on the development of autonomous systems for the military?

I am not familiar with this particular program, so I cannot comment specifically. But in any case, recognition of combatants is different than recognition of terrorists, which I suspect is what the Homeland Security Program is about. Different recognition criteria are at play. I do not advocate the use of weaponized autonomous systems in situations other than open declared warfare, where civilian populations have been duly warned of possible attack. They are not appropriate in situations where there is a significant civilian population present, as is often the case for many counter-insurgency operations.

Besides Mine-Clearing, IED disposal and surveillance, roles in which unmanned systems will be increasingly deployed in the foreseeable future seem to be also transport and logistics as well as information gathering and processing. From your experience, which trends do you expect in the development of military robots in the next couple of years?

Clearly, hunter-killer UAV (e.g., predator and reaper class) use will continue to be expanded, including for naval operations. Ground troop combat support vehicles and unmanned logistical resupply systems will also be developed. Smaller and more intelligent ground and air microvehicles will play a greater role in military operations. Additional support for battlefield casualties will be handled through robotic technology. Beyond that it's hard to predict as it is often dependent upon the nature of the threat at the time.

With all the research going on in different industries and the – in proportion to other military technologies – relatively low cost of building robotic systems, do you see any specific ethical responsibility of robot researchers, as their work could be used at least derivatively for military applications? And do you think there is the possibility, that we will experience a robot arms race?

In regards to the moral responsibility of robotics researchers, certainly, and education is the key. Most scientists do not foresee the consequences of the technology they are creating. As I often say to my colleagues, that if you create something useful, even if you are not accepting funds from the Department of Defense, that technology will be put to use by someone, somewhere, in a military application. You cannot pretend that you are not involved in this chain of

creation. That is why we must have proactive management of this research by our own field, guided by policy makers, philosophers, lawyers, and the like to ensure that we are ultimately comfortable with the outcomes of our research. The only alternative is relinquishment, as Bill Joy advocates in his "Wired" article "Why the future doesn't need us". I believe there are other alternatives if we choose to act responsibly, but we cannot ignore the looming threats of unchecked technology.

Regarding an arms race, nations always strive to be competitive in the battlefield, as it is an imperative to their existence. So governments will compete to have the best technology, but not only in robotics. These robotic systems are not weapons of mass destruction, as is the case of chemical, biological or nuclear weapons, and thus are not subject to similar developmental restrictions, at least currently. But asymmetry is an important factor in winning wars, and there is an inherent quest for self-preservation at a national level leading towards substantial investment worldwide in this technology, if not properly considered and managed at an international level. International discussions on the appropriate use of this technology are overdue.

[1] Governing Lethal Behavior in Autonomous Robots: http://www.amazon.com/Governing-Lethal-Behavior-Autonomous-Robots/dp/1420085948/ref=sr_1_1?ie=UTF8&s=books&qid= 1244630561&sr=8-1.
[2] http://www.icra2009.org.
[3] https://www.fcs.army.mil.

John P. Sullins: Aspects of Telerobotic Systems

How and why did you get interested in the field of military robots?

It was not intentional. My PhD program focused on artificial intelligence, artificial life and consciousness. During my studies I was persuaded by the works of Rodney Brooks and others, who were arguing that embedding AI and robotic systems in real world situations is the only way to gain traction on the big issues troubling AI. So, I began studying autonomous robotics, evolutionary systems, and artificial life. Right away I began to be troubled by a number of ethical issues that harried this research and the military technological applications it was helping to create. Just before I finished my doctorate the events of September eleventh occurred closely followed by a great deal of interest and money being directed at military robotics. Instead of going into defence contract research, as a number of my peers were doing, I decided to go into academic philosophy as this seemed like the best angle from which to speak to the ethics of robotics. Like the rest of us, I have been swept up by historical events and I am doing my best to try to understand this dangerous new epoch we are moving into.

In your work you have engaged questions regarding ethics of artificial life, ethical aspects of autonomous robots and the question of artificial moral agency. Where do you see the main challenges in the foreseeable future in these fields?

In the near term the main issue is that we are creating task accomplishing agents, which are being deployed in very ethically charged situations, be they AI(Artificial Intelligence), ALife (Artificial Life), or robotic in nature.

In ALife work is proceeding on the creation of protocells, which will challenge our commonsense conception of life and may open the door to designer biological weapons that will make the weapons of today look like the horse does now to modern transportation technology.

Autonomous robotics has two main challenges, the most imminent challenge is their use in warfare, which we will talk more about later, but there is also the emergence of social robotics that will grow in importance over the coming decades. Social robots are machines designed as companions, helpers, and as sexual objects. I believe

that a more fully understood concept of artificial moral agency is vital to the proper design and use of these technologies. What worries me most is that in robotics we are rushing headlong into deploying them as surrogate soldiers and sex workers, two activities that are surrounded by constellations of tricky ethical problems that even human agents find immensely difficult to properly navigate. I wish we could have spent some additional time to work out the inevitable bugs with the design of artificial moral agents in more innocuous situations first. Unfortunately, it looks like we will not have that luxury and we are going to have to deal with the serious ethical impacts of robotics without delay.

Concerning the use of robots by the military, Ronald Arkin has worked on an ethical governor system for unmanned systems. Do you think similar developments will be used in other application areas of robots in society? Especially the impact of robots on health care and care for the elderly concerns ethically sensitive areas.

Yes, I do think that some sort of ethical governor or computational application of moral logic will be a necessity in nearly every application of robotics technology. All of one's personal interactions with other humans are shaped by one's own moral sentiments. It comes so naturally to us that it is hard to notice sometimes unless someone transgresses some social norm and draws our attention to it. If we expect robots to succeed in close interactions with people we need to solve the problem Arkin has addressed with his work. Right now, our most successful industrial robots have to be carefully cordoned off from other human workers for safety reasons, so there is no pressing need for an ethical governor in these applications. But when it comes to replacing a human nurse with a robot, suddenly the machine is through into a situation where a rather dense set of moral situations develops continuously around the patients and caregivers. For instance, one might think that passing out medication could be easily automated by just modifying one of the existing mail delivery robots in use in offices around the world. But there is a significant difference in that a small error in mail delivery is just an inconvenience, whereas a mistake in medication could be lethal. Suppose we could make a fool proof delivery system and get around the last objection, even then we have a more subtle problem. Patients in a hospital or nursing home often tire of the prodding, poking, testing and constant regimen of medication. They can easily come to resist or even resent their caregivers. So, a machine dropped into this situation would have to be able to not only get the

right medication to the right patient but then will need to also engage the patient in a conversation to try to convince him or her that it is interested in the well being of the patient and wants only what is best for him or her, listen attentively and caringly to the patients concerns and then hopefully convince the patient to take the medication. We can see that this simple task is imbedded into a very complex and nuanced moral situation that will greatly task any known technology we have to implement general moral intelligence. Therefore I think the medical assistant sector of robotics will not reach its full potential until some sort of general moral reasoning system is developed.

A lot of the challenges concerning the use of robots in society seem to stem from the question of robot autonomy and especially from the question of robots possibly becoming moral agents. Where do you see the main challenges in this field?

This is a great question and I have much to say about it. I have a complete technical argument which can be found in the chapter I wrote on Artificial Moral Agency in Technoethics, in the Handbook of Research on Technoethics Volume one, edited by Rocci Luppicini and Rebecca Addell. But I will try to distil that argument here. The primary challenge is that no traditional ethical theory has ever given serious concern to even non human moral agents, such as animals, much less artificial moral agents such as robots, ALife, or AI, so we are existing in a conceptual void and thus most traditional ethicists and theologians would find the concept unthinkable or even foolish. I think it is important to challenge this standard moral certainty that humans are the only thing that count as moral agents and instead entertain the notion that it is possible, and in fact desirable, to admit nonhumans and even artefacts into the club of entities worthy of moral concern. If you will allow me to quote myself from the work I cited above, "…briefly put, if technoethics makes the claim that ethics is, or can be, a branch of technology, then it is possible to argue that technologies could be created that are autonomous technoethical agents, artificial agents that have moral worth and responsibilities – artificial moral agents."

Let me explain myself a bit more clearly. Every ethical theory presupposes that the agents in the proposed system are persons who have the capacity to reason about morality, cause and effect, and value. But I don't see the necessity in requiring personhood, wouldn't the capacity to reason on morality, cause and effect, and value, be enough for an entity to count as a moral agent? And further, you probably do not

even need that to count as an entity worthy of moral concern, a "moral patient" as these things are often referred to in the technical literature. So, for me a thing just needs to be novel and/or irreplaceable to be a moral patient, that would include lots of things such as animals, ecosystems, business systems, artwork, intellectual property, some software systems, etc. When it comes to moral agency the requirements are a little more restrictive. To be an artificial moral agent the system must display autonomy, intentionality, and responsibility. I know those words have different meaning for different people but by "autonomy" I do not mean possessing of complete capacity for free will but instead I just mean that the system is making decisions for itself. My requirements of intentionality are similar in that I simply mean that the system has to have some intention to shape or alter the situation it is in. And finally the system has to have some moral responsibility delegated to it. When all of these are in place in an artificial system it is indeed an artificial moral agent.

If we speak about a moral judgement made by a machine or artificial life-form, what would be the impact of this on society and human self-conception?

There are many examples of how it might turn out badly to be found throughout science fiction. But I do not think any of those scenarios are going to fully realize themselves. I believe this could be a very positive experience if we do it correctly. Right now, the research in moral cognition suggests that human moral agents make their decisions based largely on emotion, guided by some general notions acquired from religion or the ethical norms of their culture, and then they construct from these influences their exhibited behaviour. Working on artificial moral agents will force us to build a system that can more rationally justify its actions. If we are successful, then our artificial moral agents might be able to teach us how to be more ethical ourselves. We are taking on a great responsibility, as the intelligent designers of these systems it is ultimately our responsibility to make sure they are fully functioning and capable moral agents. If we can't do that we shouldn't try to build them.

We are not guaranteed success in this endeavour, we might also build systems that are amoral and that actively work to change the way we perceive the world, thus striping ourselves of the requirements of moral agency. This is what I am working to help us avoid.

You have argued that telerobotic systems change the way we perceive the situation we are in and that this factor and its effect on

warfare is insufficiently addressed. Where do you see the main ethical challenges of this effect and what could be done to solve or at least mitigate these problems?

The main issue is what I call telepistemological distancing: how does looking at the world through a robot colour one's beliefs about the world? A technology like a telerobotic drone is not epistemically passive as a traditional set of binoculars would be. The systems of which the drone and pilot are part of are active, with sensors and systems that look for, and pre-process, information for the human operators' consumption. These systems are tasked with finding enemy agents who are actively trying to deceive it in an environment filled with other friendly and/or neutral agents, this is hard enough for just general reconnaissance operations but when these systems are armed and targets are engaged this obviously becomes a monumental problem that will tax our telepistemological systems to the limit. It does not stop there, once the images enter into the mind of the operator or soldier, a myriad social, political, and ethical prejudgments may colour the image that has been perceived with further epistemic noise.

As we can see, there are two loci of epistemic noise; 1) the technological medium the message is contained in and 2) the preconditioning of the agent receiving the message. So, if we are to solve or mitigate these problems they have to be approached from both of these directions. First, the technological medium must not obscure information needed to make proper ethical decisions. I am not convinced that the systems in use today do that so I feel we should back off in using armed drones. The preconditioning of the operator is a much harder problem. Today's soldiers are from the X-Box generation and as such come into the situation already quite desensitized to violence and not at all habituated to the high level of professionalism needed to follow the strict dictates of the various ROEs, LOW, or Just War theory. A recent report by the US Surgeon General where US Marines and Soldiers were interviewed after returning home from combat operations in the Middle East suggests that even highly trained soldiers have a very pragmatic attitude towards bending rules of engagement they may have been subject to. As it stands only officers receive any training in just war theory but drones are now regularly flown by non officers and even non military personnel such as the operations flown by the CIA in the US, so I am worried that the pilots themselves are not provided with the cognitive tools they need to make just decisions. To mitigate this we need better training and very close command and control maintained on these

technologies and we should think long and hard before giving covert air strike capabilities to agencies with little or no public accountability.

As far as CIA UAV operations are concerned, one can witness a continuous increase. As you mentioned there are various problems connected with them. To single out just one: do you think the problem with the accountability of the actions – i.e. the question of the locus of responsibility – could be solved in an adequate manner?

This is a very hard problem that puts a lot of stress on just war theory. A minimal criteria for a just action in war, is obviously that it be an action accomplished in the context of a war. If it is, then we can use just war theory and the law of war to try to make some sense of the action and determine if it is a legal and/or moral action. In situations where a telerobot is used to project lethal force against a target, it is not clear whether the actions are acts of war or not. Typically, the missions that are flown by intelligence agencies like the CIA are flown over territory that is not part of the overall conflict. The "War on Terror" can spill out into shadowy government operators engaging an ill defined set of enemy combatants anywhere on the globe that they happen to be. When this new layer of difficulties is added to the others I have mentioned in this interview, one is left with a very morally suspect situation. As an example we can look at the successful predator strike against Abu Ali al-Harithi in Yemen back in 2002. This was the first high profile terrorist target engaged successfully by intelligence operatives using this technology. This act was widely applauded in the US but was uncomfortably received elsewhere in the world, even by those other countries that are allied in the war on terror. Since this time the use of armed drones has become the method of choice in finding and eliminating suspected terrorists who seek sanctuary in countries like Pakistan, Yemen, Sudan, Palestine, etc. It is politically expedient because no human intelligence agency agents are at risk and the drone can loiter high and unseen for many hours waiting for the target to emerge. But this can cause wars such as these to turn the entire planet into a potential battlefield while putting civilians at risk who are completely unaware that they are anywhere near a potential fire-fight. While I can easily see the pragmatic reasons for conducting these strikes, there is no way they can be morally justified because you have a non military entity using lethal force that has caused the death and maiming of civilians from countries that are not at war with the aggressor. I am amazed that there has not been sharp criticism of this behaviour in international settings.

Negotiations and treaties will no doubt be needed to create specific rules of engagement and laws of war to cover this growing area of conflict. Yet, even if the major players can agree on rules of engagement and laws for the use of drones that does not necessarily mean the rules and laws obtained will be ethically justified. To do that we have to operate this technology in such a way that we respect the self determination of the countries they are operated in so that we do not spread the conflict to new territories, and we must use them with the double intention of hitting only confirmed military targets and in such a way that no civilians are intentionally or collaterally harmed. I would personally also suggest that these missions be flown by trained military personnel so that there is a clear chain of responsibility for any lethal force used. Without these precautions we will see more and more adventurous use of these weapons systems.

One of the problems you have identified in UAV piloting is, that there is a tendency for these to be controlled not only by trained pilots, typically officers with in-depth military training, but also by younger enlisted men. Also do you see the future possibility to contract UAV piloting to civil operators? What would be the main challenges in these cases and what kind of special training would you think would be necessary for these UAV operators?

Yes, there is a wide variety of UAVs in operation today. Many of them do not require much training to use so we are seeing a trend emerging where there are piloted by younger war fighters. Personally, I prefer that we maintain the tradition of officer training for pilots but if that is impossible and we are going to continue to use enlisted persons, then these drone pilots must be adequately trained in the ethical challenges peculiar to these technologies so they can make the right decisions when faced by them in combat situations.

Since the larger and more complex aircraft like the Predator and Raptor, are typically piloted from locations many thousands of miles away, it is quite probable that civil contractors might be employed to fly these missions. That eventuality must be avoided, at least when it comes to the use of lethal force in combat missions. The world does not need a stealthy telerobotic mercenary air force. But, if we can avoid that, I do think there is a place for this technology to be used in a civil setting. For instance, just recently a Raptor drone was diverted from combat operations in Afghanistan and used to help locate survivors of the earthquake in Haiti. Certainly, that is a job that civil pilots

could do. Also, these machines are useful for scientific research, fire patrols, law enforcement, etc. All of which are missions that would be appropriate for civilians to accomplish. The ethical issues here are primarily those of privacy protection, expansion of the surveillance society, and accident prevention. With that in mind, I would hope that civil aviation authorities would work to regulate the potential abuses represented by these new systems.

Regarding the impact of telerobotic weapon systems on warfare, where do you see the main challenges in the field of just war theory and how should the armed forces respond to these challenges?

Just war theory is by no means uncontroversial but I use it since there are no rival theories that can do a better job than just war theory even with its flaws. It is, of course, preferable to resolve political differences through diplomacy and cultural exchange, but I do think that if conflict is inevitable, we must attempt to fight only just wars and propagate those wars in an ethical manner. If we can assume our war is just, then in order for a weapons system to be used ethically in that conflict, it must be rationally and consciously controlled towards just end results.

Telerobotic weapons systems impact our ability to fight just wars in the following ways. First they seem to be contributing to what I call the normalization of warfare. Telerobots contribute to the acceptance of warfare as a normal part of everyday life. These systems can be controlled from across the globe so pilots living in Las Vegas can work a shift fighting the war in the Middle East and then drive home and spend time with the family. While this may seem like it is preferable, I think it subtly moves combat into a normal everyday activity in direct confrontation with just war theory that demands that warfare be a special circumstance that is propagated only in an effort to quickly return to peaceful relations. Also, telerobots contribute to the myth of surgical warfare and limit our ability to view one's enemies as fellow moral agents. That last bit is often hard for people to understand, but moral agents have to be given special regard even when they are your enemy. Just war attempts to seek a quick and efficient end to hostilities and return to a point where the enemy combatants can again respect one another's moral worth. For instance, look how many of the European belligerents in WWII are now closely allied with each other. The way one conducts hostilities must not be done in a way that prevents future cooperation. Telerobotic weapons seem to be doing just the opposite. The victims of these weapons have claimed that they are cowardly and that far from

being surgical, they create devastating civilian casualties. These allegations may or may not be true, but they are the image that much of the world has of those countries that are using these weapons fanning the flames of intergenerational hatred between cultures.

So what you are saying is, that the current method of using UAVs might actually endanger one of the principles of just war theory, the probability of obtaining a lasting peace (iustus finis), in other words the short term military achievements might curb the long term goals of peace?

Yes, that is exactly right. People who have had this technology used against them are unlikely to forgive or reconcile. When these technologies are used to strike in areas that are not combat zones they tend to fan the flames of future conflict even if they might have succeeded in eliminating a current threat. This can cause a state of perpetual warfare or greatly exacerbate one that is already well underway. For instance, we can see that the use of remote controlled bombs, missiles and drones by both sides of the conflict in Palestine are not ending the fight but are instead building that conflict to new highs of violence.

The armed forces should respond to this by understanding the long-term political costs that come with short-term political expediency. Right now, a drone strike that causes civilian casualties hardly raises concern in the home audience. But in the rest of the world it is a source of great unease. It is also important to resist the temptation to normalize telerobotic combat operations. I would suggest backing off on using these weapons for delivery of lethal force and move back to reconnaissance missions. And yes, I do know that that will never happen, but at least we should use these weapons only under tight scrutiny, in declared combat zones, with the intent both to justly propagate the conflict and eliminate non combatant casualties.

One question connected to the normalization of warfare through telerobotics, is the so called shift-work fighting. Where do you see the main challenges in the blending of war and civilian life and how could this be countered?

I need to be careful here so that I am not misunderstood. I do understand that these technologies take the war fighters that would have had to risk their own lives in these missions out of danger and put in their place an easily replaceable machine. That is a moral good. But what I want to emphasize is that it is not an unequivocal good. Even if our people are not getting hurt, there will be real human agents on the other end of the cross hairs.

Making a shoot or don't shoot decision is one of the most profound a moral agent can be called on to make. It can not be done in an unthinking or business-as-usual way. When we blend war fighting with daily life we remove these decisions form the special moral territory they inhabit in just war theory and replace it with the much more casual and pragmatic world of daily life. Realistically I do not think there is anyway to counter this trend. It is politically expedient from the viewpoint of the commanders, it is preferable to the individual war fighters, and there does not seem to be any international will to challenge the countries that are using UAVs in this way. As the technology advances we will see more and more naval craft and armoured fighting vehicles operated teleroboticaly and semi autonomously as well. For instance, this is a major plank of the future warfare planning in America and quite a bit of money is being directed at making it a reality. It is my hope though, that these planners will take some of these critiques seriously and work to keep the operators of these future machines as well trained and professional as possible and that they operate them with no cognitive dissonance. By that I mean the operators should be well aware that they are operating lethal machinery in a war zone and that it is not just another day at the office.

I understand, that in your speech at the IEEE International Conference on Robotics and Automation 2009 in Kobe, you have also presented recommendations for the use of telerobotic weapon systems. What should be our top priority at the moment?

The Conference in Kobe was very interesting. Roboticists such as Ronald Arkin are working hard on designing systems that will act like "ethical governors" in the hope that future autonomous and semi autonomous military robots will be able to behave more ethically than humans do in combat situations. I believe the top priority right now should be to tackle this idea seriously so we can make sure that these ethical governors are more than just an idea but an actual functioning part of new systems. The main sticking point right now is that at least theoretically, a system with a functioning ethical governor would refuse orders that it deemed unethical, and this is proving to be a difficult technology to sell. If I can be permitted one more top priority it would be to investigate some of the claims I have made to provide more detailed information. Is telepistemological distancing real? Do drone pilots view the war as just a kind of super realistic video game? The military has the funds and personnel to carry out these studies and without this data we cannot rationally and consciously use

these weapons and therefore cannot use them ethically.

To mitigate the most detrimental negative effects of telepistemological distancing, there are five aspects one might consider:

1) Constant attention must be paid to the design of the remote sensing capabilities of the weapon system. Not only should target information be displayed but also information relevant to making ethical decisions must not be filtered out. Human agents must be easily identified as human and not objectified by the mediation of the sensors and their displays to the operator. If this is impossible, then the machine should not be operated as a weapon.
2) A moral agent must be in full control of the weapon at all times. This cannot be just limited to an abort button. Every aspect of the shoot or don't shoot decision must pass through a moral agent. Note, I am not ruling out the possibility that that agent may not be human. An artificial moral agent (AMA) would suffice. It is also important to note that AMAs that can intelligently make these decisions are a long ways off. Until then, if it is impossible to keep a human in the decision loop, then these machines must not be used as weapons.
3) Since the operator his or herself is a source of epistemic noise, it matters a great deal whether or not that person has been fully trained in just war theory. Since only officers are currently trained in this, then only officers should be controlling armed telerobots. If this is impossible, then these machines should not be used as weapons.
4) These weapons must not be used in any way that normalizes or trivializes war or its consequences. Thus shift-work fighting should be avoided. Placing telerobotic weapons control centres near civilian populations must be avoided in that it is a legitimate military target and anyone near it is in danger from military or terrorist retaliation.
5) These weapons must never be used in such a way that will prolong or intensify the hatred induced by the conflict. They are used ethically if and only if they contribute to a quick return to peaceful relations.

Roger F. Gay:
A Developer's Perspective

How and why did you get into the field of robotics, how has it changed in the last ten years, and what are the goals of your company?

I did not get involved in robotics until 2003 or 2004 while looking for applications of some ideas I had about improving AI in the 1980s. The problems I addressed back then were still with us in 2004 and I noticed that the technology available now makes application of my old ideas much easier; thus, commercially interesting. Helping to make robot systems work better and to make robots smarter seemed a logical place to start. My direct participation in the robotics industry started with my association with Peter Nordin, whose work in AI and learning systems lies at the heart of our commercial activities. Much of my work since then has been devoted to business development, although I have been involved in conceptualization and some high-level design. Such a fate awaits many engineers after a certain age.

It is clear that a great deal has happened in the past 10 years. iRobot's famous autonomous vacuum sweeper, Roomba only came on the market in 2002 and the company went public in 2005 due to its overwhelming success. I'm sure this will be part of the historians' account of robot history – a major turning point for the industry. Analysts have been saying that the robotics industry will grow to be larger than the automotive industry. I'm one of the greater optimists who thinks that we don't need to wait too long to see that happen.

Many of the mobile robots in use today are still largely controlled remotely by human operators. In activities such as mine clearing and some surveillance work, they are tools that allow workers to keep their distance while doing dangerous jobs. Over the past 10 years, governments around the world have been pouring a great deal of investment into robotics research, initially driven and still with heavy involvement from the military. This was well-timed and has resulted in steady progress in the related science and technology. Particularly when it comes to progress in the technology for making robots smarter and capable of performing a greater range of tasks, even insiders who shouldn't be surprised cannot help but be a little amazed. It

seems to me that the expanding public interest in robot ethics is a direct result of this rapid progress. There are various estimates about how fast progress will occur in the future – how soon we'll have certain kinds of intelligent robots in our living rooms etc. – but whether or not such progress will occur seems now only debatable at the outermost fringes.

The Institute of Robotics in Scandinavia AB (iRobis) served as a technology transfer unit that brought Peter's work and that of others out of university laboratories and into first commercial form for complete robot software systems development. Peter and I are now committed to putting the software in the hands of end-product developers. This will likely involve a new company start-up. We face an educational challenge in that our software is used and performs much differently than traditional ("old-fashioned") software systems. Interest in learning systems and genetic programming in particular, including their application in robotics has grown exponentially, which is helpful. During the last couple of years, some of the largest companies in the world have started R&D programs in the field. We also keep noticing a wealth of other possible applications for a powerful "cognitive" system. How much we can do is a matter of time and money.

What are your goals for your cognitive software system "Brainstorm"?

One of our goals is to decide on a new name for the commercial product. I'll take advantage of any part of the readership that has maintained interest this far and ask that they may send suggestions if they wish.

Our initial vision has been to provide an advanced learning and adaptive software system for robots. We will provide that to companies that want to create robots and take them to market. Our primary goal at this point is to get "Brainstorm" into the hands of end-use developers. We can make arrangements for training or even joint development efforts. In some special cases, we may even be able to develop initial prototypes for potential customers.

In the near term, I've mentioned that we have an educational goal to achieve. It's still a little odd for many engineers and business decision-makers go accept the idea of letting their robots learn behavior rather than having it rigidly programmed in line-by-line. It can also be difficult to imagine letting a machine adapt – change its behavior – while in operation (optional). What if its behavior changes in a bad way? I do not see fear of the technology in the people I speak with. But the approach is new to many of them and these are perfectly reasonable issues. The simple

answer – and there is one – is that it's not yet time to fire all the engineers. Although development can be much faster and robots smarter, it still takes capable people to design and develop and test before sending a product to market. Developers will still have much more than sufficient control over what is created, not just to assure product quality, but to use their own creative energies to produce useful and interesting machines.

Much of the history of machine learning actually lies outside of robotics. Genetic programming (GP) in particular has been applied to many "thought" problems. For example, GP systems read Internet material and provide specialized summaries of interest to their users and have even created patentable inventions in electronics. This has created one of our nicer challenges, although it still is an educational challenge. When we first tell people about our robotics software, they often want to know what specific tasks it has been developed to perform. I often respond by asking – what do you want it to do? In the world of traditional robotics, where advanced behavior can take years to develop, this can seem like an odd question. We are crossing a threshold from a situation in which technical capabilities drive product development decisions to one in which we are ready to ask what people want.

How can we imagine "genetic programming"? What is it used for in the development of robots? What is the difference to other approaches of AI programming?

The idea was taken from the concept of evolution. In the genetic programming approach (GP), a "population" of programs is created. All the programs run, and feedback is used to measure performance. The best performers are allowed to "survive" and are modified by processes that were inspired by genetics. One of them is a recombination of elements of two "parent" programs into a single child program. Just enough random changes are made to keep the possibilities open. This approach has been quite successful in guiding improvement in each successive generation, which is one of the reasons it is practical for use in the real world. Randomly creating and testing all possible programs for example, until one that does what you want it to do is created, would be impractical.

It is a very powerful technical approach. It is used to create "Turing complete" programs, which is to say that there are no logical limitations to the programs that can be created. It is capable of creating "arbitrarily complex" programs – in a good way. That is, there are no limitations on the complexity of the program that is needed.

Peter Nordin has been a pioneer in genetic programming for decades and much of his work is related to robotics. Starting in the 1990s, he had the opportunity to consolidate this effort in The Humanoid Project at Chalmers University in Sweden. One of the developments was the basic architecture for GP robotics software systems used in Brainstorm. Brainstorm is not simply a GP processing engine. It is the mind of a robot, capable of dealing with many things. It consists of several layers to deal rapidly and directly with low-level processing through higher level "thinking" and problem-solving processes. Built-in simulation allows the GP system to build and test its programs without physically carrying out tasks. It can first imagine how it will do something before doing it. (This also means that robots do not need to physically perform through generations of populations of programs to produce a working one.)

Within The Humanoid Project, GP was applied in hundreds of robot projects on a variety of hardware platforms. Humanoid robots learned to walk, four-wheeled robots learned to drive themselves, the world's first flapping wing robot took flight. A four-legged robot learned to walk, broke a leg, and automatically learned to walk efficiently with three legs. Robots have learned hand-eye coordination, grasping movements, to mimic human behavior, and to navigate, among other things. Higher level cognitive processes take care of such tasks as planning, evaluating safety, and problem solving.

GP is the only approach capable of producing arbitrarily complex, Turing complete programs. Brainstorm is the first robotics software system to carry the label "complete cognitive system." When we install Brainstorm on a robot (notice I don't mention a specific physical type of robot), it learns about itself, learns to move, and wanders through its environment learning about it as it goes. By knowing about itself and its environment, it is able to first determine how to deal with its environment and then carry out the programming that has been created in its own imagination.

Where are the major differences between your work and other approaches like the artificial brain projects, which also use evolutionary algorithms?

I do not know a lot about the artificial brain projects. From the name, and what I have read, it seems clear that we are much less interested in modeling the brain. Our interest is in a working robotic mind instead; a cognitive system for machines. There can be some incidental overlap in structure because nature is often quite logical in its designs, but modeling the brain is

not our goal. I've heard some good things about some of the brain projects. There are some very smart people involved. But, I think even by their estimates, it will be a very long time before a functional artificial humanoid brain exists.

I should also mention that from the time of Darwin to the present day, some very smart people have theorized that our minds might use an evolutionary process when we think. People naturally rationalize. Thoughts that don't make sense do not survive (even if they're right). Thoughts emerge that make sense to the person doing the thinking (even when they're wrong). For simpler things this process seems effortless – or at least to some extent it is "subconscious." In higher-level problem solving, you might be aware of simpler ideas growing in complexity as they are examined in your imagination. You consciously discard ideas that seem like they won't work and add ideas that might.

Could genetic programming be a step towards a recursive self-improving artificial intelligence (Seed AI)?

Yes, I think so. I do not sense at present, any general consensus on what a step toward strong AI is supposed to look like, but since it hasn't happened yet, I think the floor is still open to the widest range of opinion. In GP, we still tell the system what we want it to do – at least how to measure results. This is the basis for determining whether a program is better or worse than others – which survive and which perish. Above, I stated that engineers still control the character and quality of the results. The specification of what results are considered good is a very important part of that control. This is a valuable characteristic of the software for companies that want to assure that their robots aren't going to go out of control or start a rebellion and try to take over the world, etc. The current necessity of it is also one of those things that seem to put a wall between where the technology is and strong AI.

At this point I suppose I should repeat that I'm one of the greater optimists. The interaction between design and evolution fascinates me. There are some interesting ways for GP to infer logic from examples and other input techniques being brought into the mix, like showing a robot what you want and teaching it words (and concepts). It makes sense to me that expansion of the information sources robots can use and the ways in which they learn combined with GP's ability to create new will lead to something more than the sum of the parts. Yes – I think GP is a step.

What impact could genetic programming have on complex artificial

intelligence in the field where robots act as moral agents and are confronted with ethical decision-making?

There is a very optimistic discussion on using GP to approach robot ethics in Wallach and Allen's book, Moral Machines: Teaching Robots Right from Wrong. They describe a "top down" plus "bottom up" strategy in their thoughts about how autonomous moral agents might develop. Using GP, this involves the interaction between design and evolution that fascinates me. We have the possibility of experimenting with different philosophies of morality and to combine them, letting the robot evolve its own way of responding to moral issues. The "top-down" part is in the fitness functions, the programs that are designed to measure performance. This is our way of specifying good and bad.

We have already demonstrated the use of our system in evolving safe behavior. This implies that we're already in the field of ethics. Some "textbook" ethical questions involve the operation of a machine that puts human life in danger. Instead of asking what the human driver or observers should do, we let the software evolve its own reaction. We haven't tried any textbook ethical dilemmas yet, but we have looked into a robot's imagination and watched it consider the sacrifice of a jeep and itself to dispose of a bomb.

As I write this, a team is on its way to Barcelona to set-up and run an experiment with Brainstorm aimed at setting a course toward designing autonomous moral agents. There is enough interest among us to have added the experiment to the tail end of a larger project. I would be very interested in seeing the work continue. Two graduate students are involved in the effort, which may lead to some interesting thesis results if they decide to keep this focus.

The discussion around the Barcelona experiment has already become quite interesting. How, for example, might the robot's knowledge of self in combination with its ability to imagine be used to improve its moral judgments? Can we substitute a human model for the robot's self to create a form of artificial empathy? From where we are now, I can easily imagine a meaningful exploration in a larger focused project.

UPDATE: At the end of the work done at iRobis, we were able to squeeze in an initial experiment in robot ethics even though our project was not specifically funded to do that. We used a REEM-B humanoid robot at Pal Robotics in Barcelona and provided software that allowed the robot to learn how to please the human it was interacting with. The

robot learned (rather than being programmed with the knowledge) to recognize a can of Coke (Coca-Cola), that Coke will quench human thirst, and that it makes a person happy to receive one when they are thirsty. It created its own set of rules for pleasing someone based on that learned knowledge.

The experiment was included in a Swedish documentary that provided a broader look at the robot ethics and RoboEthics discussion.[1] Noel Sharkey provided a pessimistic perspective against the demonstration and Peter Nordin's positive vision. Unfortunately, time ran out before the robot faced the planned ethical choice between providing a Coke to quench the human's thirst and quenching its own thirst for electrical power.

The experiment reinforced my optimism about the short-term potential for advances in learning ethical behaviour. It was rather clear and simple, as initial experiments should be; especially when there is little time to do it all. And simple isn't bad. Engineers face a lot of complexity and a simple yet powerful idea is gold. In a very short time, a general engine for learning relationships between its own behaviors and how humans are affected by it and applying the learned knowledge was created and demonstrated.

What's been demonstrated is a shift from the need for human programmers to develop and program the logic and knowledge required to create autonomous moral agents. It has been shown that there is a way for robots to learn about the effects of their behavior using simple determinants for what is a good outcome and a bad one. It seems to me that the case for optimism is extremely clear and concrete. Robots can learn ethical behavior much the way humans do; with the advantage of learning ethics in controlled circumstances and being tested sufficiently to assure the quality of outcomes.

As I explained above, the GP learning approach is Turing-complete and capable of producing arbitrarily complex programs. Logically, there is no practical limit to what can be accomplished.

For the use in robots you have put forward an "ethical regulator mechanism". How could such a system work?

In Brainstorm Responds to Robot Ethics Challenge[2], I describe something of the idea from the 1980s mentioned above, and its potential for application as an ethics regulator. I used a rather generic title for the idea – HLL (High Level Logic). It was initially envisioned as a concept for creation of more powerful expert systems and was a few years ago suggested to a large number of AI scientists and roboticists as having

potential for development of a standard component for many AI systems, including autonomous robots.

HLL includes “experts”, manager(s), and at least one executive related in a hierarchy similar to many human organizations. Executives set goals and assign them to managers. Managers formulate plans and have the authority to approve or disapprove actions. Both executives and managers have specific responsibilities in formulating and controlling acceptable behavior. Experts with specialized knowledge can play a supportive role involving details, such as whether an action would violate the Geneva Convention.

HLL also provides a structured approach to robot-robot and robot-human interaction. For example, a human commander could modify a robot’s executive orders and then allow the robot to carry out the orders autonomously. Given the same structure in a group of robots, it was easy to imagine executives assigning tasks to other robots – chain of command. Each robot’s own executive would be aware of the robot’s capabilities, which could include sufficiency in ethics related to a particular command. In this way, an individual robot’s executive could potentially refuse to carry out an order when it is incapable of properly performing the task; instantly informing commanders that another decision is needed. A structured approach to sharing knowledge specifically as needed, automatically, is also in the vision.

There were plans to build HLL and integrate it with Brainstorm during a project that is now at its end. About the time the project started however, Microsoft offered its robotics development kit and the technical team decided to start by using it to deal with some of the lower level and service oriented mechanisms that HLL would have provided. Peter Nordin’s initial design already included high level processing in ways that nicely integrated with or directly use GP processing. HLL got shifted off the table. I built an initial prototype in 2007 that includes the basic structure. But so far it’s only been run independently with a very simple robot simulation.

UPDATE: I have started a 6 month project that includes making HLL available in an Open-Source project. A cleaned up version of the simple prototype built at iRobis should be online by the end of August (2010) along with a description of desired improvements. The first offering will include a very (very) simple robot simulation. I hope it will one day be used in development of ethical processing. At least small demonstrations should become very simple as the basic system matures. (Some simple demonstrations wouldn’t be terribly difficult from the start.) Of

course, it would also be quite nice to have HLL applying some learned ethical behavior as well.

Do you think human society is ready for autonomous systems in their daily life?

I'm sure that I want a washing machine to wash my cloths rather than doing it by hand. Same goes for the dishes. Better still if I don't need to be involved in either activity. Let someone else do it, or some thing. Humans tend to put enormous effort into making life easier and I doubt acceptance will pose an insurmountable problem. I think history can tell us much about what problems we should expect. When increased automation happens quickly for example, it can cause unemployment. But adjustments have always been made, smooth or not. When adjustments lead to higher paying jobs and lower prices, workers will run out and buy the new gadgets. Ask Henry Ford. Ask the stockholders in iRobot, which went public after only a few years due in part to acceptance of their autonomous vacuum sweepers and floor cleaners. In the broader view, I believe the age of robots will be welcomed by society in a fashion not unlike that of the acceptance of automobiles and computers. Aside from the particular benefits robots will provide, the potential for industry is enormous. Society always seems to like it when economic times are good – and the quality of life benefits that brings. What we need are plenty of good robots at reasonable prices.

Generally humans are less forgiving if machines make mistakes than if humans do. Will the human society be able to cope with robots which choose their actions according to their goal autonomously?

I'm not sure that I agree with the question's premise. Maybe it's partly because I'm an engineer. When I look at a cute, fuzzy little baby seal robot snuggling someone, I'm still very much aware of the machine parts and electronics that lie beneath the artificial skin. I could easily destroy a machine with the only consideration being cost verses benefit. Forgiveness isn't much of an issue. Not so with a fellow human. Be that as it may, I believe human society will cope in one way or another. Even in the longer term – if we're talking about – maybe even robots that are smarter than we are. There will always be those among us who will work to solve problems rather than giving up. I can however imagine recalls to fix problems, such as with automobiles, and the possibility of "grounding" robots until fixes are made – as well as investigations like the FAA conducts after airplane disasters. I also think manufactures will be aware of potential economic liabilities, which – aside from our own

humanity – will help guide decisions about the products that are offered. Safety isn't a new issue in design, manufacture, and sale of machines.

The question of responsibility – who will be held responsible for actions of a (semi)autonomous robot? Is there a need for additional legislation?

I'm a bit pessimistic about the possibility of additional legislation having a positive effect (although I should mention that I don't know much about Austrian law). I think the best guidance is to look at what is already established and by working with the understanding that robots are machines. In the near future, whether a manufacturer or an operator should be held responsible depends on the details. What happened? In concrete circumstances, it should usually be much easier to determine who was at fault after something goes wrong. The difficulties will not be unlike those of centuries of liability cases in human history. Established precedents should still hold validity. The common law approach offers the benefit of dealing with new circumstances as they arise, based on a concrete view of what actually happened; whether a manufacturer delivered a faulty product, whether maintenance was performed improperly, whether an informed operator chose to take a risk, or whether something happened purely by accident – unpredictable and beyond human control.

My view is seasoned by engineering experience. It is first principle in product development that we create things that people want. In most of my personal experience, this has always meant creating useful things on purpose. The path is still one of deciding what useful things we want robots to do, designing, building and testing products before they go to market. I understand that your question comes from consideration of future robots with behavior that is more truly autonomous. That gives rise to our interest in robot ethics. Optimistic as always, I believe the technology of ethics for robots can grow alongside increased autonomy. We should be looking at this as part of the equation for maintaining a balance.

A lot has been written on the use of robots in elderly care and in the entertainment industry mainly concerning on how this will influence interpersonal relations. What do you think is the future of robots in and their impact on the human society?

I've spent time as a hospital patient and wouldn't rate it highly as a stimulating social experience. Some of the stories I've heard about abuse of the elderly in care facilities make my teeth curl. I look forward to the day when robots can take over many routine duties and are

vigilant and capable enough to recognize when something is wrong. This is in some way an expansion on the idea of hooking people up to monitors, but may be less physically intrusive. This doesn't mean that people in care should be entirely isolated except for contact with machines. Human specialists could focus more directly on social and psychological needs. I wouldn't underestimate the positive value of psychological stimulation from machines, however. Benefits have been shown in controlled circumstances. We also need to use our imaginations to create benefits. Technology might for example, more reliably inform someone when a friend is going to a common area in an elderly care facility and wishes company. It could potentially keep family members better informed about the state of their relatives in care. Again – an important question is – what do you want it to do?

On a more general basis, what do you think about robots as moral agents? What is to be expected in the next decade?

One can imagine robots standing motionless, doing nothing, in the presence of a human in need. So long as they are not causing the problem, there would be little distinction in this regard between the robot and an automobile or washing machine. It could be better of course, if robots were capable of helping people in need, even choosing to perform "heroic" acts to save a human from tragedy.

As I said above, I think designing ethics into robots is part of the equation for balancing increasing autonomy. Put in human terms, greater autonomy should be balanced with greater personal responsibility. As robots become more intelligent, more capable of "thinking" for themselves, we need mechanisms to control the quality of their decisions that are equal to the task. I take this as part of a design philosophy, separate from the issue whom courts hold liable when things go wrong. A manufacturer can be held liable for not including sufficient ethical safeguards in a robot's design.

Some aspects of moral behavior are obviously quite necessary. For example, if we want to build robots that are physically capable of killing and put them into domestic service, we don't want them to go around killing people. In fact, we will very definitely want them to avoid behavior that could result in harm. We can't build them as simple utilitarian creatures, single-mindedly concerned about specialized tasks. If we did, we might end up with what is now only a sci-fi nightmare – robots disposing of living creatures because they get in the way.

Accurately predicting what will actually happen in the future, in a particular time period especially, is a lot harder than discussing what is possible. To a pretty large extent, what actually happens during the next 10 years will depend on who gets money to do what and how much. I will predict an increase in interest and funding for research of work on moral agents. This prediction is based only in part on what I have said so far. I believe that research into developing autonomous moral agents can yield a great deal of general value in the field of AI. After all, we use the same mind to process moral questions as we do others.

In the field of military robots ethical questions have been raised. Some questions are tackling issues at hand other issues seem decades away. How do you see your role as a developer in this context?

In modern design, we tend to create enabling technology – technology that can be used for the creation of numerous end-products for a variety of purposes. Brainstorm, and GP technology generally, is an advanced example. If you build a fitness function specifying what you want to happen, a program can automatically be built to do it. We intend to put this technology in the hands of end-product developers, where final decisions about product design will be out of our control. I would be more than happy to include the best tools for ethics in the development package. Our present effort focuses on the use of existing Brainstorm technology and demonstrating fitness functions for ethical decision making. Expanding even just this effort would be of value. I have noticed over my years in the software industry, the rapid adoption of solutions presented as examples in software tutorials. Aside from the research value in doing the work, the more examples we can produce, the better.

I also appreciate being able to address your questions, whatever they may be. The interest in interdisciplinary discussions regarding robot ethics is quite beneficial in my opinion. I am pleased to participate.

It has been argued that, if properly programmed, robots could behave more ethically than human soldiers on the battlefield. How can we imagine something like a machine readable copy of the Geneva Convention?

Not all weapon systems should be measured against human performance. I read an argument recently regarding an autonomous delivery system, capable of completing a bombing mission on its own. Although the record is not perfect, smart technologies have been better at finding and hitting military targets with less incidental damage.

There is a larger set of considerations to this case, but my point here is that system performance should be compared to appropriate alternatives.

But let's consider, hypothetically at least, a goal of building an ultimate autonomous humanoid soldier. We want this breed to obey the Conventions rather than being a Terminator type robot that decides to wipe out the human race. They might even play an important role in United Nations peace keeping missions. If they're going to think for themselves and decide actions related to the Conventions, then they will need to have knowledge of the rules of the Conventions, one way or another. Somewhere in the chain of research and development, this knowledge needs to get into a form that machines can use.

The suggestion made in my article assumes that putting Geneva Conventions in machine readable form is not a complex undertaking. Based on the parts I am familiar with, it does not appear that it would be. Neither would formulating rules that computers could process.

The greater technical challenge is in getting the robot to recognize circumstances well enough to reliably apply the rules. For example: Can a robot distinguish between a green military uniform and green civilian clothing? What if combatants on the other side dress to blend with the civilian population? Can it distinguish between a combatant ready to fight and one who is surrendering? The challenges have led to suggestions that robots capable of autonomous behavior should be banned from combat roles – where they could replace human combatants.

There are many possibilities for autonomous and semi-autonomous machines that are more limited in their capabilities than an ultimate humanoid soldier. What approach can be taken to create efficient and objective criteria to assure that autonomy is sufficiently ethical? ("Efficient" in form for real-world use.)

What we need is a systematic approach that integrates development of robot ethics capabilities directly with the flow of R&D work on autonomy. My idea for immediate use of machine-readable Conventions is rather basic, involving just the sort of thing that R&D engineers often do. Build, test, and focus engineering effort on solvable even if challenging problems. Continue research in areas where solutions are still farther out. Is the robot's performance better, worse, or equal to human performance? Keep track of the roles and circumstances that can be supported technically, in view of the Conventions. Maintain a balance between measurable technical capabilities and the roles and

circumstances of autonomous machine behavior in deployment.

I have imagined an Open / Open Source project that would first focus effort on creating a basic or generic (sort of) machine-readable encoding of the Conventions. What I mean is that it should not be geared especially toward any particular processing system. Developers should be able to use it efficiently while making their own choices regarding how the data is processed and how it is used. One team might choose a rule processing system while another may use it in conjunction with a learning system. It could also be used to formulate tests in simulation and in quality assurance systems that would also help formulate functional and technical requirements.

Having an Open project seems a good idea. It would allow for rapid dissemination to all interested parties and could make use of feedback from a larger body of users. Of critical importance is general agreement on the validity of results. Perhaps this point becomes clearer in the next paragraph.

Extending such a project could lead to detailed answers to robot ethics issues. The project could play a central role toward developing international technical standards; standard tests and benchmarks, working out how to measure performance of systems in particular roles at particular levels of autonomy. A certain amount of technology could also be developed in support of applying the standards, running the tests. It's a safe bet that the first machines to pass tests won't pass all the tests. The complete ultimate humanoid soldier that properly handles all aspects of the Conventions is a ways off yet.

The idea responds to those who suggest banning all autonomous machine capabilities as weapons, by suggesting technical support for an acceptable deployment strategy. Bring autonomy into a fight to the extent that it is well-tested and quality assured. Will autonomous robots be better than humans? Yes, robots designed to obey the rules will be better if testing standards require it. Those that aren't, won't be deployed (by nations that accept the standards). Taking a systematic approach that integrates ethics directly into the R&D process would push development of better ethical performance. Ethics would become a systematic integral part of technical development, with results measured according to international standards.

Let me take that one step further. Let's say we do all this and the UN finds the standards and testing an acceptable way to determine whether autonomous machine technology can be used to support more

dangerous potential peace-keeping missions – or even peace-creating missions. Can machine autonomy help to create and maintain peace? It's a thought that brings my own focus back to a question I've asked more than once. What do you want robots to do?

---

[1] English translation available here: http://isr.nu/robots/SVT_Barcelona_EN.doc.
[2] available on the Internet: http://mensnewsdaily.com/2008/12/10/brainstorm-responds-to-robot-ethics-challenge.

Author Information

Colin Allen
Professor of Cognitive Science and History & Philosophy of Science in Indiana University's College of Arts of Sciences, a member of the core faculty in Indiana University's Center for the Integrative Study of Animal Behavior and author (with Wendell Wallach) of the book *Moral Machines. Teaching Robots Right from Wrong* (2009).

Jürgen Altmann
Researcher and lecturer at the Technische Universität Dortmund, one of the founding members of the International Committee for Robot Arms Control. Since 2003 deputy speaker of the Committee on Physics and Disarmament of Deutsche Physikalische Gesellschaft (DPG, the society of physicists in Germany). Currently directs the project on "Unmanned Armed Systems – Trends, Dangers and Preventive Arms Control" located at the Chair of Experimentelle Physik III at Technische Universität Dortmund.

Ronald C. Arkin
Regents' Professor and Associate Dean for Research at the School of Interactive Computing at the Georgia Institute of Technology and author of the book *Governing Lethal Behavior in Autonomous Robots* (2009).

Peter Asaro
Philosopher, member of the Faculty of the Department of Media Studies and Film at the New School University, New York and affiliated with the Center for Cultural Analysis at Rutgers University, New Jersey. He has written several articles on new technologies and their ethical challenges as well as on military robots and just war theory.

George Bekey
Professor Emeritus of Computer Science, Electrical Engineering and Biomedical Engineering at the University of Southern California and Adjunct Professor of Biomedical Engineering and Special Consultant to the Dean at the California Polytechnic State University. He is well known for his book *Autonomous Robots* (2005) and is Coauthor of the study *Autonomous Military Robotics: Risk, Ethics and Design* (2008).

John S. Canning
Has had a leading role in the weaponization and safety of unmanned systems, having substantially participated in the OSD-lead effort to produce the *Unmanned Systems Safety Guide for DoD Acquisition* document. Previous assignments include Science

Advisor to ADM Robert Natter (ret), Commander, Fleet/Forces Command, Norfolk, VA; Science Consultant to VADM Diego Hernandez (ret), Commander, THIRD-FLEET, Pearl Harbor, HI; NSWCDD's Chief Engineer for CVNX; and Threat Cell Chair for RADM Tim Hood (ret), PEO (Theater Air Defense) for Tactical Ballistic Missile Defense. For this last item, Mr. Canning initiated a group effort that led to being awarded the CIA's Seal Medallion.

Gerhard Dabringer
Editor of this volume. Historian, Research associate at the Institute for Religion and Peace (Vienna) of the Catholic Military Chaplaincy of the Austrian Armed Forces. His main field of research is military ethics.

Roger F. Gay
Co-founder of Institute of Robotics in Scandinavia AB (iRobis) where he was a key contributor in the creation of its R&D program and currently VP for Partnerships at JPN Group in Sweden, which is involved in continuing development and marketing of self-learning software that supports creation of compete robotic systems that automatically develop and maintain their own intelligent controls, adapt, and solve problems. Educated in engineering, his 30 year career has spanned software engineering and product development, marketing, project management, policy analysis, journalism / commentary, and business.

Armin Krishnan
Visiting Assistant Professor for Security Studies at the University of Texas at El Paso. His book *Killer Robots* (2009) focuses on aspects of legality and ethicality of autonomous weapons.

Fiorella Operto
President at the Scuola di Robotica(Genova) and scholar professor of philosophy. She has specialised in ethical, legal, and societal issues in advanced robotics and has been working in collaboration with important scientific laboratories and research centres in Europe and in the United States. Recently she has co-operated with the Robotics Department of the National Research Council in Italy in promoting the knowledge and understanding of the new science of Robotics.

Noel Sharkey
British computer scientist, Professor of Artificial Intelligence and Robotics and Professor of Public Engagement at the University of Sheffield as well as Engineering and Physical Sciences Research Council Senior Media Fellow. He has held a number of research and teaching positions in the US (Yale, Stanford) and in the UK (Essex, Exeter and Sheffield). He is founding editor-in-chief of the Journal

Connection Science. Besides his academic contributions he is also widely known for setting up robot control and construction competitions for children and young adults and for his numerous appearances on the BBC as expert on *Robot Wars* and *Techno Games* and as co-host for *Bright Sparks*.

Peter W. Singer
Senior Fellow and Director of the 21st Century Defense Initiative at the Brookings Institution, who is considered one of the world's leading experts on changes in 21st century warfare. His new book "Wired for War" (2009) investigates the implications of robotics and other new technologies for war, politics, ethics, and law in the 21st century.

Robert Sparrow
Senior Lecturer at the School of Philosophy and Bioethics at Monash University, Australia. His main fields of research are political philosophy (including Just War theory), bioethics, and the ethics of science and technology. He is currently working on a research project on the impact developments in military technology have on the military's core ethical commitments, the character of individual warfighters, and on the application of Just War theory.

John P. Sullins
Assistant Professor of Philosophy at Sonoma State University. He has substantially contributed to the fields of philosophy of technology and cognitive science as well as to the fields of artificial intelligence, robotics and computer ethics. In addition John P. Sullins is a Military Master at Arms and directs the Sonoma State University Fencing Master's Certificate Program.

Gianmarco Veruggio
CNR-IEIIT Senior Research Scientist and President at Scuola di Robotica(Genova). He is serving the IEEE Robotics and Automation Society as Corresponding Co-chair of the Technical Committee on Roboethics, as Co-Chair of the Human Rights and Ethics Committee, and as a Distinguished Lecturer. Among others, he was the General Chair of the *First International Symposium on Roboethics*, Sanremo, January 2004. In 2009 he was presented with the title of Commander of the Order to the Merit of the Italian Republic.

Also available:

Ethica. Jahrbuch des Instituts für Religion und Frieden

2009: Säkularisierung in Europa – Herausforderungen für die Militärseelsorge
2008: Der Soldat der Zukunft – Ein Kämpfer ohne Seele?
2007: Herausforderungen der Militärseelsorge in Europa
2006: 50 Jahre Seelsorge im Österreichischen Bundesheer. Rückblick – Standort – Perspektiven
2005: Familie und Nation – Tradition und Religion. Was bestimmt heute die moralische Identität des Soldaten?
2004: Sicherheit und Friede als europäische Herausforderung. Der Beitrag christlicher Soldaten im Licht von „Pacem in Terris"
2003: Das ethische Profil des Soldatenvor der Herausforderung einer Kultur des Friedens. Erfahrungen der Militärordinariate Mittel- und Osteuropas
2002: Internationale Einsätze
2000: Solidargemeinschaft Menschheit und humanitäre Intervention – Sicherheits- und Verteidigungspolitik als friedensstiftendes Anliegen

Ethica. Themen

Gerhard Marchl (Hg.): Die EU auf dem Weg zur Militärmacht? (2010)
Petrus Bsteh, Werner Freistetter, Astrid Ingruber (Hg.): Die Vielfalt der Religionen im Nahen und Mittleren Osten. Dialogkultur und Konfliktpotential an den Ursprüngen (2010)
Werner Freistetter, Christian Wagnsonner: Friede und Militär aus christlicher Sicht I (2010)
Stefan Gugerel, Christian Wagnsonner (Hg.): Astronomie und Gott? (2010)
Werner Freistetter, Christian Wagnsonner (Hg.): Raketen – Weltraum – Ethik (2010)
Werner Freistetter, Bastian Ringo Petrowski, Christian Wagnsonner: Religionen und militärische Einsätze I (2009)

Broschüren und Behelfe

Gerhard Dabringer: Militärroboter. Einführung und ethische Fragestellungen
Christian Wagnsonner: Religion und Gewalt. Ausgewählte Beispiele
Joanne Siegenthaler: Einführung in das humanitäre Völkerrecht. Recht im Krieg
Informationsblätter zu Militär, Religion, Ethik (dt, eng, frz)
Informationsblätter zu Franz Jägerstätter (dt, eng, frz)
Informationsblätter zum Humanitären Völkerrecht (dt, eng, frz)

ISBN: 978-3-902761-04-0