WORKING DRAFT - Do Not Cite Without Permission
For Presentation at: **WeRobot 2015,** University of Washington, April 10, 2015.

# Regulating Robots: Approaches to Developing Robot Policy and Technology

Dr. Peter Asaro
Assistant Professor, The New School
Visiting Fellow, Center for Information Technology Policy, Princeton University
Affiliate Scholar, Center for Internet and Society, Stanford Law School
asaro@newschool.edu

## Introduction

We stand at the cusp of a robotics revolution with the potential to permeate and transform our social and physical world to an even greater extent than the computer and smartphone revolutions of recent decades. Insofar as we can apply lessons learned from those technological revolutions, and improve on our processes of innovation and mechanisms of regulation, we have an opportunity to shape this emerging revolution towards desirable social goals. Failing to do so will mean abdicating responsibility, and allowing powerful political and commercial interests, and the technocratic values of engineers, to shape the social values that are built into these technologies and which will reshape the social and physical world according to those values.

There has been increasing discussion about how best to approach the question of developing regulatory policy alongside technological innovation in the field of robotics. From the individualistic approach of traditional engineering ethics and design values,[1] to a corporate approach to self-policing and internal ethics review boards, as Google and Deep Mind have embarked upon,[2] to proposals for a national robot regulatory agency,[3] to attempts at international regulation of certain kinds of robotic systems.[4] In this paper I will examine the unique challenges of regulating robots and autonomous systems facing each approach, and reflect on the necessary requirements of a comprehensive framework for regulating robots.

Running through each approach to regulation are concerns about the predictability of complex systems and unanticipated risks, as well as the challenge of establishing and articulating the fundamental values, public interests and norms which engineers, firms and policies ought to

---

[1]Knight, Heather (2014) "How Humans Respond to Robots: Building Public Policy through Good Design," *Brookings Institute Report*, July, 2014, pp. 20.

[2]Lin, Patrick and Evan Selinger (2014) "Inside Google's Mysterious Ethics Board," *Forbes*, February 3, 2014.

[3]Calo, Ryan (2014) "The Case for a Federal Robotics Commission," *Brookings Institute Report*, September, 2014, pp. 18.

[4]Asaro, Peter (2012) "On Banning Autonomous Lethal Systems: Human Rights, Automation and the Dehumanizing of Lethal Decision-making," *International Review of the Red Cross*, 94 (886), Summer 2012, pp. 687-709.

strive to meet. Meeting these concerns and challenges will require a creative mix of expertise, experimentation, testing and the participation of the public in discussing the values integral to the development of advanced robotic technologies. The paper will reflect on the ways in which this might be achieved to the advantage of a broad range of stakeholders.

The aim of this paper is to consider the necessary elements of a general framework for regulating robots. Of course, the problem of "regulating robots" is very complex, and covers a broad variety basic and advanced technologies, and many application domains. Robotics involves the integration of sensors, software and actuators, and many of its component elements might be used in other non-robotic applications, or combined in very different types of applications. Thus, it is difficult to separate robotics from other technologies with which it is integrated. Robotics can also be applied in many practical domains, from small drones, to self-driving cars to surgical robots, to factory robots, to personal and home assistants, each with it own set of concerns and challenges. A more fundamental challenge is that the extent of the future technical capabilities of robotics is only partially foreseeable, as are its potential applications and their social impacts. Of course, many if not all, technologies embody, project, and sometimes enforce social values and politics.[5] So we could easily make the same arguments and draw the same conclusions about the social regulation of any technology.

In its most abstract form, technology is simply a system of skills, tools and practices.[6] The introduction of any new techniques, skills or technological tools necessarily implies a transformation in practices, whether at the level of individual tasks or at the levels of task management or institutional organization. The question is whether these transformations conform to existing norms, or challenge those norms. Those technologies that significantly impact upon or transform norms, I call disruptive technologies. More precisely, there are four possibilities:

1) Existing norms may continue to govern practices after the introduction of a new technology, largely unchanged.
2) Existing norms may be diminished or dissolved after the introduction of a new technology.
3) New norms may emerge to govern the practices after the introduction of a new technology.
4) Existing norms may be transformed or reconceived.

According to the received view of norm formation, norms follow after the establishment of new practices. When considering the changes in norms governing new technologies, we are presented with a chicken-and-egg problem. If we have not yet implemented a new technology,

---

[5]Winner, Langdon (1980) "Do Artifacts Have Politics?" *Daedalus*, Vol. 109, No. 1, Winter 1980, pp. 121-136.

[6]In ancient Greek *techne* meant the performance of a craft, what in modern English we might call technique. Of course the performance also requires the use of the body and tools, and in modern English we often use "technology" to refer to sophisticated tools. Yet with sophisticated technologies, such as computers, we are really concerned with their performance, which entails the practice of skills as well as the functionality of the tools. Thus, we should think of technologies as consisting of the function of material tools and the skilled practices of people in conjunction, as well as their systematic organization.

we cannot observe what the new norms are (if we limit norms to already-recognized and accepted behavior). We can examine existing norms, and try to determine if the use of a new technology would challenge or violate those norms. If so, we might try to regulate that technology, and try to ensure that the norm remains in effect. Or we might find that it does not violate those norms, and so we should not need to regulate it, and the existing norms ought to be sufficient until proven otherwise. But there are further possibilities. We might find, after a technology is adopted, that new sets of practices emerge, followed by new norms.[7] We may also find that the capabilities of a new technology actually manifest situations in which we recognize norms that had always been tacitly assumed, but never articulated or codified because it was previously unnecessary. I will argue in this paper for a proactive approach to norm innovation. That is, we should be actively interrogating norms that may be threaten by disruptive technologies, with the aim to innovate them in ways that both improve them and render them more resilient to impending disruptions.

Robotics is an interesting case precisely because it will be disruptive of a broad range of social practices and norms, much like information and computing technologies (ICTs) have been over the past four decades. And robotics will be more visible than ICT, allowing for and encouraging a more open and public debate about what values they should embody, and how they ought to be regulated.[8]

The potential and existent means for regulating robots are numerous and varied, making it difficult to know where to begin. With little to pick out robots as a particular or unique class of things to be regulated, we still find robots falling into a variety of regulatory regimes determined by their applications. Robots have been operating in factories for decades. These industrial robotics fall under a variety of national and international standards requirements, such as ANSI/RIA,[9] NIST and ISO,[10] and workplace safety regulations, such as OSHA[11] in the US. Surgical robots[12] and personal care robots[13] have recently been the subject of an ISO standards

---

[7]Among these are "best practices" and other forms of soft law.

[8]Calo, Ryan (2011) "The Drone as Privacy Catalyst," *Stanford Law Review Online*, Vol. 64 (December 12, 2011), pp. 29-33.

[9]Robotic Industries Association (2013) "New Robot Safety Standard," blog post, May 29, 2013, Link. See also: ANSI/RIA R15.06-2012.

[10]See: ISO Technical Committee 184/SC 2 on Robots and Robotic Devices.

[11]OSHA Instruction PUB 8-1.3 SEP 21,1987 Office of Science and Technology Assessment. See also: OSHA Directive STD 01-12-002.

[12]Virk, Gurvinder S. (2013) "Dr Robot needs standards too! With more robots used in medical care, International Standards are vital for safety," International Electrotechnical Commission, *e-tech blog,* May 2013. See also:

[13]Cole, Emmet (2014) "New International Standards Boon to Personal Care Robotics: Will early certification boost mainstream acceptance or breed false confidence?," *Robotics Business Review*, January 9, 2014. See also: ISO 13482:2014

process. As many of these systems utilize the electromagnetic spectrum for wireless communications, and a variety of commercial computer interfaces, they also fall under a host of standards for everything from WiFi and USB, to safety standards for the lasers utilized in range finding and LIDAR, to network protocols, data file formats and compression standards. But these sorts of regulations do little to address our concerns with how robotics might impact social values such as unemployment, privacy, safety, and accountability of individuals acting at a distance through robots.

Two of the leading robotic technologies, small unmanned aerial drones and self-driving cars, also happen to fall under the jurisdiction of existing large regulatory bureaucracies: the Federal Aviation Administration (FAA), and the National Highway Traffic Safety Administration (NHTSA) and a federation of Departments of Transportation (DOT), respectively. While it is clear that these bodies have an interest in regulating flying robots and driving robots, what is less clear is how they should go about creating those regulations, and whether they have sufficient expertise and a broad enough range of participating stakeholder interests and social values in order to do so.

As we look at the history of disruptive technologies such as computerization, and recognize that we have not yet fully come to terms with their impact on social values, such as privacy, one cannot help but wonder if we could have done better. Knowing what we know now, could we have developed better privacy practices in the early days of internet, perhaps imposing more secure communications, a users' bill of rights? Could we impose those practices now? What lessons can we learn from history to inform how we develop policy for robots? This paper aims to begin outlining the key challenges facing various regulatory approaches that might be applied to robotics, as well as provide the a sense of the goals and elements that a good regulatory framework ought to incorporate.

## Three Cases

Rather than discuss the issue in broad and abstract terms, I will ground this study in several cases of recent and emerging regulation of robotic technologies, which represent different approaches and scales. From this, I hope to illuminate both the key challenges of regulating robots, and also the necessary elements for responding to these challenges. These three cases are chosen because they represent the broadest classes or approaches to technology regulation: national regulatory bodies, international regulation, and corporate and individual self-regulation.

The first case will examine the efforts to develop regulations for small aerial drones[14] to fly within the national airspace of the US. This case is interesting because it falls, at least partially, within the jurisdiction of an already-established federal regulatory body, the FAA. Despite this, there have emerged various challenges and problems with addressing the regulatory issues that

[14]The FAA has chosen to call these Unmanned Aerial Systems (UAS), though they are also known as Unmanned Aerial Vehicles (UAVs), Remote Piloted Vehicles (RPVs), or simply drones. I will refer to these primarily as drones, except when discussing the technical definitions offered by the FAA.

drones raise. So it will be helpful to start there by examining both what has been accomplished in that process, but also in examining what has not worked in that process.

The second case will look at efforts aimed towards internal review and self-regulation through the case of Google's acquisition of the DeepMind machine learning company, and their proposed internal ethics board. In conjunction with this, I will discuss other aspects of self-regulation of technology innovation, including engineering ethics, value-centered design methodologies, and the creation of industry standards.

For the third case, I will look at efforts, which I have been personally involved in, to develop an international treaty prohibiting the development and use of autonomous weapons systems. Unlike national regulation, international regulation lacks any central constitutional authority. As a result it is far less comprehensive than national regulatory apparatus, even if it is successful.

**FAA Drone Regulations**

It is helpful to briefly review the history of the FAA's efforts to regulate small drones, and its various shortcomings. Rather than fault the competence or earnestness of the FAA, the aim is to elucidate what may be structural problems with the regulatory framework within which the FAA operates, especially as it runs up against economic and political pressures, or problems requiring expertise beyond its own capacities.

Small drones, those under 55 pounds, fall under the designations established for model aircraft. A 1981 circular[15] issued by the FAA provides guidelines for the operators of such aircraft, including staying below 400 feet, not flying near airports, doing pre-flight checks, and other guidelines. Compliance was voluntary, but local model airplane communities did a good job of self-regulating with these rules, and training newcomers to the hobby about them.

In 2007, the FAA announced a rule that required drones flown for commercial purposes to obtain a special permit in advance. These permits were given out almost exclusively to government agencies. The first person to be fined under this rule was Ralph Pirker for selling footage of his flight over the University of Virginia campus to the University for a promotional video. Pirker and his lawyer successfully challenged the fine in court, where a federal judge struck it down on the basis that the FAA had not followed proper rule-making process in issuing the rule. The FAA appealed, but the suit was settled out of court before the appeal was heard.

In February of 2012, the FAA Modernization and Reform Act became law.[16] A key part of that Act included the modernization of air-traffic control to incorporate GPS data in tracking aircraft, as well as a mandate for the FAA to promulgate rules for allowing unmanned aerial systems (UAS) into the national airspace by September of 2015.

---

[15] See: FAA AC 91-57 - Model Aircraft Operating Standards.

[16] See: Public Law 112 - 95 - FAA Modernization and Reform Act of 2012.

It is worth noting some of the major provisions of the mandated UAS rules, and their relation to the powerful interests that shaped those provisions. First, it is clear that the driving influence behind the effort was a long-standing consortium of government agencies, including the Department of Defense,[17] Homeland Security, NOAA, and NASA[18] who were already using drones for training and data-collection within the national airspace. The military had flown drones within restricted military airspace for decades, but found itself needing to move drones between bases, and return them from overseas deployments, and moving them under their own power would require access to the national airspace. The science agencies wanted to collect atmospheric, meteorological and other data within the national airspace.

It was the efforts of the government agencies that resulted in the 2007 rules provided for special certifications of authorization (COAs) on a per-flight basis.[19] But even this was becoming burdensome and the government agencies sought to normalize drone operations in the national airspace, as well as create test sites for new designs, another major provision of the 2012 Act. Under the provisions of the Act was a research program to develop a sense-and-avoid system that would enable drones to avoid collisions with each other and with other aircraft. The concept was to start with ground-based systems for tracking aircraft and detecting potential collisions and communicating that information to the drones, and then moving towards miniaturizing the systems in order for these systems to be carried on-board the drones. It is also clear from the Act that a particular model for implementing drone technology was envisioned–remote-piloted aircraft of significant size (comparable to existing aircraft) requiring similar certification of the airframe and pilot training, and posing similar risks to air-traffic as current aircraft. The six test sites were meant to be designated areas where dangers to other aircraft and people on the ground would be minimal, and permits could be obtained to perform test flights of experimental test drones.

Also lobbying congress on this issue was AUVSI, a large industry group representing a broad spectrum of business and commercial interests in developing and unmanned vehicle systems, including drones. The Academy of Model Aeronautics (AMA), an association of model aircraft hobbyists, was perhaps the only grassroots organization playing a part in shaping policy at that point. The efforts of the AMA and AUVSI led to a provision in the 2012 Act that excepted small drones (under 55 pounds) flown for hobby and recreation, and would lead to the development of a separate set of rules for small drones which were much more like model aircraft than existing private and commercial manned aircraft, which was directing much of the FAA's UAS rulemaking effort.

In February of 2015, the FAA issued a Notice of Proposed Rulemaking for drones.[20] Prior to that, a number of commercial interests had pressed for special permission from the FAA for

---

[17] See: OSD UAS Airspace Integration Plan, March, 2011.

[18] See: Civil UAV Capabilities Assessment, NASA, 2004.

[19] See: FAA webpage "Busting Myths about the FAA and Unmanned Aircraft"

[20] See: Small UAS Notice of Proposed Rulemaking (NPRM).

flying drones. The first of these was a group of Hollywood film studios who sought permission to fly cameras on drones, on essentially closed sets with professional pilots.[21] Based on a provision in section 333 of the 2012 Act.[22] that allowed the Director of the FAA to grant such special permission, it was granted. Similar permissions were subsequently granted to a number of other business interests.

At the US Senate hearings on this process in March of 2015, Amazon complained that it took too long to receive their test license, and they had already abandoned the test model in favor of another by the time they were authorized to fly it. To which the Senators replied that they would make sure the process was speedier and more efficient in the future.

The course of the development of rules for drones in the national airspace has clearly been shaped by the influence of powerful stakeholders–the government consortium and industry trade groups. The issuing of permits and special permissions, as well as the process of locating the six test sites, have heavily favored commercial and political interests with the resources and influence to achieve their goals.

Of course, the FAA was devised in order to enable the commercialization of the airspace, and to designate it as a public commons and shared resource. The FAA was created in the same act that nationalized the airspace above private properties in order to enable aircraft to fly without acquiring the permission of each individual property owner in advance. So it should be of no surprise that it is heavily influenced by commercial interests. This fact is not unique to the FAA, and similar influences and relationships can be seen in other federal regulatory agencies including the FDA, FCC, FEC, SEC, and others.

The concept of regulatory capture involves a regulatory agency coming under the effective control of the institutions it is meant to regulate. This can be through overly-cozy personal relationships, including working side-by-side for so long that regulators adopt and internalize the goals of the institution they are supposed to regulate. Or the revolving door of professional careers that move easily between or back-and-forth between the regulatory agency and the regulated institutions. This is a real danger, but not the only challenge facing this form of regulation.

Even when regulatory agencies escape regulatory capture, they are still heavily influenced by the political and commercial interests of the largest players in a regulatory domain. This is not so much a failing as a feature of the design of such frameworks. Some efforts have been made in these frameworks to provide for public advocates. The Consumer Financial Protection Bureau was specifically created to counter this in the domain of personal banking. But these efforts have had mixed results, not least because they run up against so much political pressure from the large institutions they are supposed to regulate.

---

[21] Verrier, Richard, "FAA gives drone exemption to Hollywood production firms," *Los Angeles Times*, September 25, 2014.

[22] See: FAA website "Section 333".

In fact, there is a mechanism built into the FAA rulemaking process (and that of many other federal agencies) which allows for public hearings, and soliciting public comment on the rulemaking proposals. While a not really a form of direct participation in the making of the rules themselves, the admission of public comment allows a certain feedback mechanism. But it is unclear what weight, if any, is given to public opinion and comment at any phase of the rulemaking process, as this remains largely opaque and not subject to review.

In the current process, the public reaction has focused on a few issues, though it is still unclear how the FAA will respond. One issue of great concern to hobbyists, and an issue the FAA has explicitly asked for public comment on, is the designation of an even smaller class of drones, micro UAS, which would weigh under 4.4 pounds, that might be allowed to fly over people and without special licensing. This seems likely to be accepted, based on safety studies from bird strikes (at this weight they pose little risk to manned aircraft).

Another is enthusiasm for flying first-person-view (FPV) aircraft beyond line-of-sight. FPV is a means of flying the aircraft by looking through a camera mounted on the aircraft (with the pilots using goggles or a screen). This simulates being inside the aircraft, and with powerful consumer radios these aircraft can fly as far as ten miles away from the pilot. Currently the proposed rules require the pilot, or a spotter, to maintain visual contact with the aircraft and thus line-of-sight, even if the pilot is using FPV. Given that these aircraft can fly well beyond the distance of line-of-sight, many FPV pilots are unhappy with this rule. Certainly the larger drones will use this technology, such as remote-piloted long-haul cargo aircraft, so the question is whether such FPV flights will fall under the much more rigorous requirements for large drones, or whether they can be accommodated in the small and micro drone regulations.

Not discussed at all by the small drone rules are fully autonomous drones. Drones which fly solely under the control of computers, with no pilot at the controls or using FPV, are essentially prohibited. While many universities and now companies such as Amazon, Google and Facebook entering into the development of such systems, this is a significant gap. Companies like the Swiss-based SenseFly have already marketed a system for planes to fly under algorithmic control to systematically scan and area and develop 3D models of it. And there are numerous potential commercial applications of such autonomous drones.

By far the most contentious issue facing the FAA effort to regulate drones is with how to approach the privacy issues. At the time of the 2012 Act, the ACLU was already calling for the FAA to include privacy as an aspect of public safety, and part of its responsibility to protect people on the ground.[23] The FAA balked at this, claiming they lacked the jurisdiction and expertise to deal with the issue. And yet the FAA was also claiming to regulate commercial uses of small drones, including the sale of video footage, implying jurisdiction over how aerial video footage is used, and a growing expertise in how to regulate data from aerial surveillance. Recently they have even been issuing take-down notices for YouTube video footage from

[23] Stanley, Jay and Catherine Crump (2011) "Protecting Privacy From Aerial Surveillance: Recommendations for Government Use of Drone Aircraft," ACLU Report, December, 2011.

drones.[24]

A looming problem which I discussed in my WeRobot 2014 paper (Asaro & Cooper),[25] is the regulation of airspace close to the ground over private property. While the FAA has traditionally regulated the "navigable airspace" this varies depending on the type of aircraft in question and the local airspace classification–typically navigable airspace includes airports and over 1,000 feet above ground for airplanes, but almost any altitude, barring obstructions such as buildings powerlines and trees, for helicopters. Small drones further problematize this concept, as they can fly very close to the ground and even in obstructed and confined spaces, such as wooded areas and even inside buildings. The air-rights over private property are a nebulous concept, which is now being called into question. Some restriction on drones buzzing around your yard seems necessary, but what should the limits be, and what can be done about drones hovering just beyond the legal limit (whatever it is)? Certain existing local laws prohibiting peeping toms, paparazzi, and stalking may be applicable, but may also require demonstrating the intentions behind the use of drones that may not be easy to determine when a drone is spotted, or to prove in court.

And yet the proposed rules in the recent notice of rulemaking include no specific privacy provisions. Just ahead of issuing their notice, however, the President's Office issued an executive order covering government use of drones, which did specify guidelines for the collection of data, and the treatment of data from drones that contains personally identifiable information (PII). This move appears calculated to deflect criticisms over the lack of privacy provisions in the FAA rules. Yet such an executive order only applies to federal government agencies, and thus private and commercial drone operators will face no new privacy restrictions, despite the very real threats to privacy that drones pose.

States and local governments have stepped in to fill the void left by the lack of federal regulations on drones, especially in the privacy domain. There have been a flurry of such regulations appearing in different locales. Some of these simply require police and government agencies to acquire a warrant from a judge in order to follow or film a citizen with a drone, (Virginia). Others prohibit the use of drones for hunting (Alaska and Oregon). Others proposed rules actually outlawing drone photography (Nevada), or outlawing street photography altogether (Arkansas). Some municipalities have asserted the rights of property owners and individuals to disable drones flying above or near their property or person (Oklahoma[26]).

The proliferation of local drone regulations is a bit of a free-for-all. One could argue that those models which work best will eventually win out. But in the meantime it will also cause great

---

[24]Koebler, Jason (2015) "The FAA Says You Can't Post Drone Videos on YouTube," *Motherboard*, March 12, 2015.

[25]Asaro, Peter and Diana Marina Cooper (2014) "Robots, Micro-Airspaces, and the Future of 'Public Space'," WeRobot 2014, Unpublished manuscript.

[26]http://www.npr.org/blogs/thetwo-way/2015/02/10/385239519/-drone-shoot-down-bill-advances-in-oklahoma

confusion among the public and drone operators as to what the rules are, especially as they travel to new locales. There have already been numerous cases where the local police are confused as to what types of use are permissible and how to determine the use of a given drone.

In addition to concerns over how large federal regulatory agencies operate, we should also be aware that these operations occur in a context where state and local jurisdictions will promulgate laws and rules of their own. This seems necessary and desirable, just as local authorities decide where to put stop signs, and what the speed limits are for a given street, even though federal regulations determine what the signs must look like. But there will certainly be battles over the limits of what is permissible under local authority (drone racetracks that permit speeds over the federal 100mph limit? Or fully autonomous drones?). We might think of these as semi-autonomous zones, like the Salt Flats where they set land speed records, and the FAA itself has asked for public comment on the creation of "innovation zones" based on its test-site model.

**Self-Regulation & Artificial Intelligence**

In 2013, Google acquired the UK-based machine learning company DeepMind. One stipulation of that acquisition, insisted on by the three founders (Demis Hassabis, Shane Legg and Mustafa Suleyman), was the creation of a internal ethics review board within Google.[27] While few details about this review board have been made public, it appears that it is intended to be an integral part of the company, not just the DeepMind subsidiary. It also appears that it will be tasked primarily with considering the ethical issues stemming from advanced artificial intelligence. Since we do not really know how the internal ethics board at Google will be arranged, I want to use this example as a point of departure to discuss an array of regulatory mechanisms that we might loosely call "self-regulation" and how they might apply to robotics.

I include professional ethics under the self-regulation category. There is a long tradition within the engineering profession to ensure the safety of their constructions. Engineering ethics training is required for all engineering students, and the professional engineering societies go further to include codes of ethics as well. Much of the focus of these efforts is on individual responsibility, and standing up to outside pressures to sign-off on faulty or shoddy work, or dangerous designs. The aim is safety, and the presumption is that the individual knows right from wrong, and can ensure good design and workmanship through their individual actions. Surely if all the engineers do this, the thinking goes, we would have good and safe designs.

Whether this approach will work for a technology (like robotics) or not depends on the kinds of values that we want engineering to embody in its products. While it might work well to ensure that a bridge can handle an adequate load and remain safe under normal usage, this approach is not well suited to grappling with multiple competing values. Should the bridge have more lane space for cars or for bikes and pedestrians? This requires consulting with the community who

[27] Lin and Selinger, *op cit.*

will use the bridge to determine what values[28] they want it to support, not reflecting on one's own views of engineering virtue. Similarly, if the safety fault in a complex system depends on the interaction of many components, which are being engineered by different teams, it may be difficult to detect the potential fault, or to act to ensure that it is corrected. Combined with the fact that it can be difficult to determine the cause of such a failure even after the fact, has been called the "problem of many hands" wherein the diffusion of individual responsibility in both the design and in accountability results in individuals distancing themselves from their own personal responsibility in large group projects.[29]

Other approaches to engineering ethics have sought to rectify this short coming with better design methodologies, including value-centered design, human-centered design, and participatory design. The first two design methodologies have also been influential in the emerging field of human-robot interaction, where we find researchers trying to figure out how users perceive and interact with robotic systems in order to design robots that are better able to interact with people.[30] Such approaches have the advantage of being able to address a much broader set of values, and follow through the design process with attention and respect for those values. They do not themselves have much to say about where those values come from, however. Participatory design does concern itself with eliciting values from users, and often doing so through their interaction with prototypes. It can also be quite cumbersome to implement, and it is not always clear who the users are, and how their participation should be structured.[31]

One way of looking at ethics boards is as a bureaucratic instantiation of the individual ethics model from engineering ethics. That is, rather than leaving it up to each individual, which can be a burdensome demand on their time or politically difficult to realize, an ethics board is devised which both grapples with the most difficult questions and ensures ethical standards are met across the organization. But many of the same shortcomings can also be found: what values does the board consider, and where do those come from?

Can we take any lessons from the proliferation of bioethics and medical ethics boards, or even Institutional Research Boards (IRBs) in hospitals and research institutions? While these have been successful, they have not been spaces where we see robust innovation in values, norms or ethical ideas. IRBs mostly serve to ensure that research participants give informed consent, and protect vulnerable populations. Bioethics and medical ethics boards largely ensure that an approved policy is being implemented, and research and hospital practices conform to it, and are

---

[28]Or whose values it should support, if we wish to frame it as a political question. Engineers might wish to ignore this question, and simply "build to specification". But in many cases the expertise of engineers is needed to explain which alternatives can be realized, and at what costs, and thus they are involved in the political process whether they like it or not.

[29]Nissenbaum, Helen (1996) "Accountability in a Computerized Society," *Science and Engineering Ethics,* **2**(1), pp. 25-42.

[30]Knight, *op cit.*

[31]Asaro, Peter (2000) "Transforming Society by Transforming Technology: The Science and Politics of Participatory Design," *Accounting, Management and Information Technologies*, **10**(4), pp. 257-290.

rarely tasked with actually devising the policy. When they are, it is usually a matter of balancing legal obligations (legal rights of patients, treatment of lab animals) with the mission of the institution, including perhaps its religious ideals (or taking a position on a controversial issue such as end-of-life treatment or using stem cells). Such ethics boards have been useful for ensuring that ethical policies are implemented and realized, but not spaces for the innovation of new norms.

Ethics boards are conceived as being "internal" to an institution. This makes them accountable to the leadership of that institution. Which raises the question of whether such boards are able to hold the leadership accountable to the values they aim to promote, or whether they merely serve an advisory function (indeed many have the title "advisory board"), with the ultimate decision resting with the institution's leadership. Unless an internal ethics board has real authority within the institution, it cannot really be expected to shape the institution itself. The other alternative is an "external" board. This could be made of external experts, but still merely advisory, or again given some real authority within the institution.

Related to the question of internal versus external is the question of public perception and transparency. Insofar as ethics boards do their work behind closed doors, they provide no transparency, and offer little evidence upon which public trust might be built. If the aim is merely the expert advice that these boards might offer, this might be sufficient. But if the aim of an institution is to convince the public that its work is ethical, or is serving particular values, they made need to demonstrate this in a more transparent way.

This most challenging issue remains that of innovation around norms and values. While a group of experts comprising such a board may actually be innovative in this regard, they probably still need the proper incentivization for this to occur. Beyond that, we might also be concerned with where those values and norms are coming from. Insofar as they are internal and advisory to an institution, why should we expect them to develop norms and values beyond the interests and aims of the institution? If they are external, how do we ensure that there is not a "capture" of the board, as its members align their independent interests over time with those of the institution? Where is the participation of affected stakeholders in the development and articulation of those norms?

The other examples of self-regulation that we see are the promulgation of "best practices," voluntary compliance, industry standards, and formal standards. These processes, for the most part, involve a institution or group of institutions coming together to agree upon a shared set of regulations or standards, and then adhering voluntarily to those standards. These approaches work best when there is a technological advantage to compliance, as there is with many standards. It makes sense for companies to comply with technical standards, such as USB3 or TCP/IP, because they can ensure their products are interoperable, and they can win the trust of consumers this way. Of course, there can be vicious politics behind the standards (e.g. HD-DVD vs Blu-Ray, or Beta vs VHS) when one company stands to gain a market or IP advantage depending on which standard is adopted. Or companies can follow their own off-beat standards if they think it advantageous (such as Apple's constantly changing power and video adapter plugs). That is why the best standards are typically those developed by independent institutions which do not have a financial stake in the ownership of the standard (such as IEEE, NIST, ANSI,

or ISO).

With some exceptions, the primary weakness of most of the self-regulatory frameworks is that they lack teeth, and they are not usually innovative. Technical standards can be innovative, and can be effectively regulated, but only with interest and effort. Internal ethics boards are not by default very innovative and tend to only seek minimal forms of compliance. However, they have the potential to be rethought and redesigned as places where innovation happens. In order for them to have real teeth, they probably need independent or external supervision and accountability. How else can we count on them to make tough decisions that effect company profits, unless we can ensure that they answer to something outside of company leadership.

**Banning Autonomous Weapons**[32]

The development of systems that are capable of targeting and firing weapons without human supervision and control is well within current technological capability. Indeed, there are various systems that already do this in various ways. However, the use of autonomous weapons by militaries is highly constrained (including some missile defense systems, gun sentries, and missiles guided by target recognition algorithms). We are, however, a the verge of a revolution in these autonomous technologies, where we could start seeing a broad variety development and deployed in the coming years, from autonomous submarines, bombers, jet fighters, and tanks, to autonomous systems designed for urban warfare and policing civilian populations.

For the past few years, I have been working to raise awareness and international discussion about this issue. The aim of myself and my colleagues in the International Committee for Robot Arms Control and the Campaign to Stop Killer Robots is work towards an international consensus and treaty that would prohibit the development and use of such fully autonomous lethal weapons. The issue has been taken up by the United Nations, and will be discussed a second meeting of the Convention on Certain Conventional Weapons (CCW) in Geneva, immediately following WeRobot.

In an obvious way, the development of such an international treaty demonstrates yet another framework for regulating robots at an international level. Like the others it has its limitations, and also its own lessons to offer. One major shortcoming is that such a treaty would only hold states accountable for the development of autonomous systems for use by their militaries. It would still permit the development and use of such systems for domestic policing and private security. Such uses still fall under national law. There are also well known challenges to enforcing international law in general, and arms control regimes in particular. And unsurprisingly, the United Nations, and international diplomacy, as a bureaucracy is less than ideal for the timely development of regulations covering a fast moving technological development.

---

[32]The text of this section replicates portions of my paper for the WeRobot anthology: Asaro, Peter (2015) "*Jus nascendi,* Robotic Weapons and the Martens Clause," in Ryan Calo, Michael Froomkin and Ian Kerr (eds.) *Robot Law*, Edward Elgar Publishing.

Those shortcomings aside, I believe the case of regulating autonomous weapons provides a good example of innovation in norms, which should parallel innovation in technology as a requirement for good regulation. That is, as technologies destabilize existing norms, we need innovation around what norms ought to replace them, and these should influence the design and development of new technologies, rather than always following behind technological disruption and beholden to the changes wrought by technological innovation. In other words, norm creation should be proactive, despite the traditional accounts of norm-creation following a pattern of established practices, which in turn follow the adoption of new technologies. Norm creation needs to take greater precedence in the process of technological change.

It is easy to frame the current debate over autonomous weapons as a utility (or risk) calculation which weighs the potential risks of harms from a new technology against its potential benefits. But it would be better to view it as a question of how we might best regulate this new class of technologies, which might take many forms and have various sorts of capabilities which challenge existing norms. That is, while autonomous weapons do pose many obvious risks, it would be difficult to regulate or mitigate these based on capabilities which are as yet unknown. Rather, it makes sense to focus on the norms that will be challenged or violated by this new technology, and try to determine how best to protect those norms from being undermined, or strengthen them rather than allow them to be diminished by emerging practice. Because of the new capabilities these technologies will bring, it is also important to consider what nascent norms may be in place which have not previously been recognized or articulated because the need had not previously been questioned.

In the case of autonomous weapons, what is really new is the automatic selection of targets and automated decisions to use violent force against those targets. While Sharre[33] and others have pointed out that there are various weapons systems already in use that have simple implementations of these capabilities, it is not at all obvious that the practices and norms governing the use of current systems will adequately scale to the sophistication and complexity that robotic and autonomous weapons systems appear poised to achieve in the coming years and decades. Anderson and Waxman[34] agree that new practices and norms must emerge to govern these new technologies, but would rather leave this process to self-regulation and the dissemination of best practices and soft law. So the debate is not whether new norms will emerge, but where they should come from, how they should be articulated, who should articulate them, and whether and where they should be codified.

In discussing these issues in various forums over the past few years, what seems increasingly clear is that there are some nascent norms and principles regarding autonomous weapons which are widely shared, and which are beginning to take shape. What emerged from the discussions at the Experts Meeting convened by the Convention on Certain Conventional Weapons at the

---

[33] Paul Scharre, *Robotics on the Battlefield, Part I: Range, Persistence and Daring,* Center for New American Security Report, pp. 47. Downloaded from: http://www.cnas.org/sites/default/files/publications-pdf/CNAS_RoboticsOnTheBattlefield_Scharre.pdf

[34] Anderson, Kenneth and Matthew C. Waxman (2012) "Law and ethics for robot soldiers," *Policy Review*, 28 April 2012, available at: http://ssrn.com/abstract=2046375.

United Nations in Geneva in May of 2014,[35] was a growing convergence towards a new norm or principle, that of "meaningful human control." There remains some disagreement and confusion, at least in terminology, over the various definitions of automatic, autonomous, and the nature of intention, agency and responsibility in highly automated and autonomous systems. Yet despite this terminological disagreement, there was broad-based agreement that military attacks and the use of violent force in war should always be kept under "meaningful human control." What exactly constitutes meaningful human control, and how we might evaluate whether a given weapons system conforms to a possible requirement for it, are matters that remain to be articulated.

However, it seems quite clear that we have something that looks very much like an emergent principle. It is difficult to call this a norm proper, even though it is the case that states currently do act so as to keep weapons under meaningful human control. It is less clear whether they actually believe that they have a shared obligation to do so. That belief and obligation might emerge in the future as practices develop around new autonomous weapons technologies. But it is also possible that such a norm might not emerge in practice in the absence of regulations or guidelines. The real risk here, from a moral and legal perspective, is that practices based in political and military expediency might give rise to weak norms, or no norms at all, and the use of weapons without any meaningful human control may become acceptable as a *de facto* norm. In that case, we might find ourselves in a world where the risk of unintended consequences (or the ability to strategically obscure one's intentions) in using such weapons is substantial, while the responsibility and accountability for such use by states and individuals is limited or easily avoidable in practice.

It has been argued that no professional militaries would want to use such weapons, because they would have unpredictable results.[36] But this is rather speculative, as such systems are not yet available, and the behavior and beliefs of states and their militaries may change as new technologies begin to enable new kinds of operations, tactics and strategies. Once states begin to see some advantage to using such weapons, they may be reluctant to impose any regulations, and in the event of widespread use there may be little basis in customary law to reign them in. Even if a handful of countries with advanced professional militaries were to observe a stricter set of norms, if a significant number did not then a norm, in the traditional sense, may fail to be established.

As a matter of analogy, it is helpful to look at the concept of "superfluous injury and unnecessary suffering" in international humanitarian law (IHL). This phrase first appeared in the St. Petersburg Declaration of 1868. Prior to its statement, it was only implicitly or vaguely recognized as a norm which had not been articulated clearly before. That is, militaries already recognized that one should not intentionally cause injury or suffering beyond what was necessary to achieve a military objective. So while it is permissible to kill or wound in order to achieve an objective, causing additional injury or increased suffering which serves no military

---

[35] See: CCW website on Lethal Autonomous Weapons, http://www.unog.ch/80256EE600585943/(httpPages)/6CE049BE22EC75A2C1257C8D00513E26?OpenDocument

[36] Sharre, *op cit.*

purpose ought to be avoided. It could be argued that the norm had existed all along, but it was the introduction of new technologies–exploding and incendiary bullets, in that case–which inspired the articulation and codification of this normative principle.

In some sense the prohibition on the use of these new types of bullets, and the recognition of this nascent norm emerged together. In another sense, the recognition of the norm and its underlying principle were fomented by unease at what these new weapons represented for the future of warfare. That is, we might consider that there was something in the moral sensibility and conscience of the delegates to the St. Petersburg meeting that the use of such weapons is morally wrong, but also that they may not have been able to state clearly exactly what was wrong about them before attempting to find the appropriate language for their moral sense. It took some effort to work through what their moral conscience was, and to express this in words, as well as to reach consensus on how to codify this in law, and to articulate the prohibitions on weapons that it implied.

Thus, the principle of meaningful human control would appear to be something that has historically been taken for granted–assumed but never stated. Weapons always required humans to decide when and where to use them. While booby traps and mines challenged this implicit assumption to some extent, their use was not sufficient to motivate the articulation of this new principle, though it did spawn various practices, norms and even treaties. Autonomous weapons pose a much greater challenge to our previous assumptions about the determination of targets and decisions to use violent force. As such, they have motivated a discussion and reflection on the principles of humanity and the dictates of public conscience. It is from this reflection that the principle of meaningful human control has emerged. And indeed, given the uncertainty over how such systems might be developed in the future, it is all the more important to clarify the fundamental principles.

Meaningful human control, as it has thus far been articulated, contains several elements.[37] First, it is fundamentally humanist in its insistence on explicitly human control of targeting and firing decisions. If any new principle might be convincingly derived from the "principles of humanity" as expressed in the Martens Clause, surely it would be a principle that ensures human control over the violence of war, political and strategic decision-making,[38] and war itself.[39]

---

[37]The phrase "meaningful human control" was coined by Richard Moyes, and its initial articulation first appeared in the briefing papers of the UK-based NGO Article 36:
Article 36 (2013) "Killer Robots: UK Government Policy on Fully Autonomous Weapons," Briefing Paper, April, 2013, 5 pp. Downloaded from: http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf
Article 36 (2013) "Structuring Debate on Autonomous Weapons Systems," Briefing Paper, November, 2013, 3 pp. Downloaded from:
http://www.article36.org/wp-content/uploads/2013/11/Autonomous-weapons-memo-for-CCW.pdf
Article 36 (2014) "Key Areas for Debate on Autonomous Weapons Systems," Briefing Paper, May, 2014, 4 pp. Downloaded from: http://www.article36.org/wp-content/uploads/2014/05/A36-CCW-May-2014.pdf

[38]Heather Roff (2014) "The Strategic Robot Problem," *Journal of Military Ethics*, **13** (3), pp. 211-227.

[39]Mark Gubrud, "The Prinicple of Humanity in Conflict," ICRAC Blog, November 19, 2012. Downloaded from: http://icrac.net/2012/11/the-principle-of-humanity-in-conflict/

It is not yet clear what constitutes "control" of a weapon exactly. So it seems that more discussion of this would be prudent. While a strict definition is not necessary, a shared understanding should be a goal of further discussion. Consider, again by analogy, the principle that prohibits "superfluous injury and unnecessary suffering." These seem fairly clear as a matter of principle–injuries and suffering that go beyond military necessity. Yet as a matter of practice, or judging individual weapons, it is far from clear how one might apply this principle to a new weapon under review. How does it apply to pain-inducing weapons that do little physical damage, for instance? In practice, of course, states have developed norms in the application of the principle, shoring it up with more explicit treaties or guidelines as necessary. At the very least, control implies that the effects and potential consequences of using a weapons system must be reliable and predictable to an extent that the human can exert some form of control over it. It also implies that the performance of the system must conform to the intentions of the operator such that it is possible to distinguish when a system is under control, and when an operator has lost control.

Another key aspect of meaningful human control is the "meaningful" part. While it is difficult to define exactly what meaningful control consists of, the concept itself is clear and it seems a reasonable standard, if not an obvious one. In part it aims to prevent weapons systems that use humans instrumentally as approval mechanisms. For example, ordering a soldier to press a "fire" button every time a light comes on would imply that a human is technically "in control" of the weapon system even if they effectively have no meaningful control over what the system is targeting, or how and when it is using lethal force against those targets. But it also implies more than this, in that meaningful control also entails taking responsibility for the use of the weapon system, and being accountable for the consequences of that use.

In order to ensure that a system retains human control, and thereby upholds the principles of humanity and does not trivialize human actions in the process, the requirement for control to be meaningful is necessary. For the killing of a human to be meaningful, it must be intentional. That is, it must be done for reason and purpose. Philosophically, intentionality requires understanding the meaning and significance of an act. While autonomous systems may be programmed to operate in a certain way, given a certain set of conditions, they cannot understand the significance of their operations. This is in part why they cannot make legal or moral judgements. But this also relates to the question of human dignity. If a combatant is to die with dignity,[40] there must be some sense in which that death is meaningful. In the absence of an intentional and meaningful decision to use violence, the resulting deaths are meaningless and arbitrary, and the dignity of those killed is significantly diminished.

Meaningful human control also offers some positive guidance on how systems ought to be designed to interface with humans. It obliges engineers and designers to consider how the use of violent force is a form of symbolic and intentional action, as well as a functional performance. As such, an interface ought to provide its user with the potential to make meaning and take

[40]While combatants are not entitled to a dignified death under IHL, arguably every human remains entitled to dignity under Human Rights Law, even if they are liable to be killed as enemy combatants.

meaningful actions, as well as to perform tasks. Under Human Rights Law, there would also be a duty upon states to ensure that such weapons could not violate human rights outside of armed conflict–by ensuring that meaningful human control is maintained. Moreover, it couples the making of meaning with control over the system–to the extent the system is automatic, the meaning of its activities are dependent upon higher-order levels of organization, which may not be able to address what is significant in a given situation. Meaning-making is a distinctly human capacity, and artificial systems will lack such capabilities for the foreseeable future. The use of anthropomorphizing language can often confuse the real capabilities of systems, which is why it is imperative that we make principles like meaningful human control clear.

More generally we can see in the emergence of this new norm and potential principle a model for how we might proceed in developing regulations for robots. That is, when we can foresee that a new technological capability or application could destabilize an existing norm that society values, then we should think about how to innovate that norm. What are the consequences of giving it up? What did the old norm do? What more could it have done? What does a new norm need to do?

**Reflections**

The biggest issue facing the regulation of robots, and really all disruptive technologies, is that we need to develop new norms at the same rate that we adopt new technological capabilities and applications. The simple fact of the matter is that disruptive technologies are transforming social practices at a relentless pace. In order to develop good regulatory policy we need not only the political will, which might be difficult to muster or to align, we also need shared norms that we desire the technologies to conform to. Absent the norms, the political will is either directionless or misapplied. Sometimes the political will drives policy in the absence of public and expert dialogue. This is the case in some of the state and local drone legislation, which fixates on often sensational notions of drone use, such as hunting, while ignoring critical privacy issues. So too can traditional norms get extended well beyond their usefulness or reasonable applicability. The Supreme Court's test for the reasonable expectation of privacy is an example of a test that may have made sense in a bygone era. The current environment of rapid technological innovation and integration renders acts of public viewing that were once reasonable into serious invasions of privacy when retained, indexed, combined in archives and made searchable.

A big part of articulating new norms involves dialogue, among experts, in the public sphere, in the cultural domain, among policymakers, and between and among all of these. The latest films and pop culture can and should inform public opinion, as should the technological optimism of innovators, and pessimism of skeptics. We seem to live in a paradoxical time in which we are better able to communicate with one another than ever before in history, and yet politics are becoming more dysfunctional and divisive than ever before. I am more inclined to believe this is a product of deliberate efforts to be divisive by those who seek economic and political advantage, rather than due to a natural inclination of a society empowered with communication technologies. The public sphere can and should be aplace where new norms are incubated and critically discussed.

Is the problem one of balancing conflicting interests? Sometimes, but it seems more often the issue is that we simply lack relevant norms. If we simply try to extrapolate traditional norms onto new technologies with greater and more extensive capabilities, it rarely works and we often find that those traditional norms break down. Privacy from aerial surveillance was not such a big deal when few could afford airplanes and cameras used film. Now that flying digital cameras are cheap and becoming ubiquitous, a new norm is needed. The drafters of the Geneva Conventions did not envision computers making targeting and firing decisions outside of human control. Extending the norms established to regulate human behavior in war to autonomous weapons would result in a variety of serious problems. Producing superintelligent AI, or automating the majority of the world's population out of employment without dealing with the risks and social implications of doing so, is simply reckless.

There are those who would argue that we should wait for the technologies to develop more fully, discover their capabilities and limitations, and then come up with regulations in an incremental and *ad hoc* manner. This strategy means that the development of norms will always be in some sense reactionary, and based in compromise and expediency. It will also ensure that the systems and regulations that result are not well thought out in advance of the technology. It will mean that society will always be fighting an uphill, rear-guard battle against those aspects of disruptive technologies which fail to meet social desirable ends.

There are those who would argue that we cannot develop norms for technologies, before we understand what those technologies can do, and how people will use them. They argue that norms and regulation should follow innovation and practice, and focus on best practices that have been learned through experience. But this is a recipe for social engineering that asks for forgiveness rather than permission.

There are those who would argue that my approach is Luddite, and seeks merely to slow down technological innovation so that norms can catch up. By why should we not invest more in the development of norms? Where is the research funding for that? The schools, departments and degrees? If we want good robot law, we will need more robot ethicists, philosophers and lawyers.

The value of technological innovation is not intrinsic or absolute. The economic forces guiding the adoption of new technologies are not sufficient on their own to ensure that technologies realize any social goods. We need people thinking about and discussing the norms that should regulate our robotic future.