UIC John Marshall Journal of Information Technology & Privacy Law

Volume 32 | Issue 2 Article 1

Spring 2016

Autonomy of Military Robots: Assessing the Technical and Legal ("Jus In Bello") Thresholds, 32 J. Marshall J. Info. Tech. & Privacy L. 57 (2016)

Remus Titiriga

Recommended Citation

Remus Titiriga, Autonomy of Military Robots: Assessing the Technical and Legal ("Jus In Bello") Thresholds, 32 J. Marshall J. Info. Tech. & Privacy L. 57 (2016)

ARTICLES

AUTONOMY OF MILITARY ROBOTS: ASSESSING THE TECHNICAL AND LEGAL ("JUS IN BELLO") THRESHOLDS

REMUS TITIRIGA, PHD[*]

PROFESSOR, INHA LAW SCHOOL

SOUTH KOREA

> "The claws were bad enough in the first place--nasty, crawling little death-robots. But when they began to imitate their creators, it was time for the human race to make peace--if it could!"[1]
>
> Philip K. Dick, *Second Variety*

---

[*] Remus Titiriga, PhD, LLM, MSc is professor at INHA Law School in Incheon, South Korea. With a double background, in Computer engineering and Law, he is fluent in 3 languages (including English and French). He has almost 7 years of international teaching experience within French, American, South Korean universities.

He has published a book: La comparaison, technique essentielle du juge européen at L'Harmattan, Paris, France, 2011. Other of his researches, in Legal Methodology, ICT regulation, International Security or International Economic Integration are visible at: http://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=1600093.

[1] *Second Variety,* LITHIVE http://lithive.com/books/26.

ABSTRACT

While robots are still absent from our homes, they have started to spread over battlefields. However, the military robots of today are mostly remotely controlled platforms, with no real autonomy. This paper will disclose the obstacles in implementing autonomy for such systems by answering a technical question: What level of autonomy is needed in military robots and how and when might it be achieved, followed by a techno-legal one: How to implement the rules of humanitarian law within autonomous fighting robots, in order to allow their legal deployment? The first chapter scrutinizes the significance of autonomy in robots and the metrics used to quantify it, which were developed by the US Department of Defense.

The second chapter focuses on the autonomy of "state-of-the-art" robots (e.g.; Google's self-driving car, DARPA's projects, etc.) for navigation, ISR or lethal missions. Based on public information, we will get a hint of the architectures, the functioning, the thresholds and technical limitations of such systems. The bottleneck to a higher autonomy of robots seems to be their poor "perceptive intelligence."

The last chapter looks to the requirements of humanitarian law (rules of "jus in bello"/rules of engagement) to the legal deployment of autonomous lethal robots on the battlefields. The legal and moral reasoning of human soldiers, complying with humanitarian law, is a complex cognitive process which must be emulated by autonomous robots that could make lethal decisions. However, autonomous completion of such "moral" tasks by artificial agents is much more challenging than the autonomous implementation of other tasks, such as navigation, ISR or kinetic attacks.

Given the limits of current Artificial Intelligence, it is highly unlikely that robots will acquire such moral capabilities anytime soon. Therefore, for the time being, the autonomous weapon systems might be legally deployed, but only in very particular circumstances, where the requirements of humanitarian law happen to be irrelevant.

I. THE MEANING OF AUTONOMY IN ROBOTS AND WAYS TO QUANTIFY IT

A) DEFINING AUTONOMY IN ROBOTS

The term "robot" is based on the Czech word "*robota*," meaning "serf or slave," and came into being in Karel Capek's 1921 play *R.U.R.* (*Rosumovi Univerzální Roboti* or Rossum's Universal Robots). Today, a robot is defined as "a mechanical creature which can function autonomously."[2] This concept of autonomy is correlated but different from *automation*. While both processes can be executed independently, from start to finish without human intervention, there is a qualitative distinction between them.

An automated system normally operates with no human intervention, but is not self-directed and lacks decision-making capabilities. It only replaces routine processes with software/hardware that follows a step-by-step order, which usually requires human supervision. In a certain way, an automated system is rigid, blind, and, one might say, stupid.

An autonomous system has the aim to emulate human cognitive processes rather than to simply eliminate them. Therefore, autonomy requires three main characteristics (which also help to identify whether a machine is truly autonomous): (1) The "frequency of human operator interactions" that the machine needs in order to function; (2) The machine's ability to function successfully despite "environmental uncertainty"; and (3) The machine's level of assertiveness to each of various operational decisions that let the machine to complete its mission.[3]

An autonomous system may also learn or acquire new knowledge, such as adopting new methods to accomplish its tasks or adjusting to changing surroundings. Over the last twenty years there have been many attempts to identify and quantify the "levels of autonomy" in robots. The degree of autonomy of a robot has been considered, at least in the beginning, from the viewpoint of human interaction/interface with a robotic system.

B) SHERIDAN'S LEVELS OF AUTONOMY

During his work for the National Aeronautics and Space Admin-

---

[2] Robin R. Murphy, Introduction to AI Robotics 2 (2000), reproduced in Benjamin Kastan, AUTONOMOUS WEAPONS SYSTEMS: A COMING LEGAL "SINGULARITY?" 1 J. L. TECH & POL'Y 45, 49 (2013).

[3] William Marra and McNeil, Sonia, *Understanding "The Loop": Regulating the Next Generation of War Machines*, 36 HARV, J. L. & PUB. POL'Y 1, 3 (2012).

istration (NASA), Thomas Sheridan created both a categorization and vocabulary to express the state of human-machine interaction at any given moment during a mission.[4] His classification is structured into levels and implies that autonomy is a delegation of a complete task to a computer, that a system operates on a single level of autonomy for any given task, and that these levels are discrete and represent steps of growing difficulty.[5]

At Level one, a machine is automated. Levels two through four emphasize the allocation of the decision-making capacity between human and machine. Levels five through nine offer an initial decision-making power to the machine and confer to human operators special levels of approval or veto power. At Level 10, a machine is fully autonomous.

Most other tentative classifications for quantifying robots' autonomy seem to follow the steps of Sheridan's early work. The United States Department of Defense (US DOD), for example, has funded a number of studies about the "levels of autonomy" in robots in order to aid their development.[6] Further analysis of these studies is provided in later sections.

C) AIR FORCE RESEARCH LAB SCALE OF AUTONOMY

The Air Force Research Lab (AFRL)'s autonomy frame considers 11 levels of autonomy, specifically for Unmanned Aerial Vehicles (UAV):

Remotely piloted vehicle;
Execute pre-planned mission remotely;
Changeable mission;
Robust response to real-time faults/events;
Fault/event adaptive vehicle;
Real-time multi-vehicle coordination;
Real time multi-vehicle cooperation;
Battle space knowledge;
Battle space single cognizance;
Battle space swarm cognizance; and
Fully autonomous[7]

There is a perspective, common in aviation, to consider the fighter

---

[4] THOMAS B. SHERIDAN, TELEROBOTICS, AUTOMATION, AND HUMAN SUPERVISORY CONTROL (1992).

[5] *Id.*

[6] DEF. SCI. BD., TASK FORCE REPORT: THE ROLE OF AUTONOMY IN DOD SYSTEMS (July 2012) at 4, *available at* https://fas.org/irp/agency/dod/dsb/autonomy.pdf (hereinafter TASK FORCE REPORT).

[7] Eric Sholes, *Evolution of a UAV Autonomy Classification Taxonomy*, AEROSPACE CONFERENCE DIGEST (2007), available at http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4161585 (on file with author).

pilots of a plane as acting according to the OODA decision loop.[8] It was natural to consider the autonomy of UAVs, on the pathway of Sheridan's view of the interactions between humans and machines, from the perspective of the OODA loop.[9] The greater a machine's ability to observe, to orient, to decide and to act by its own, the greater its autonomy would be.[10]

This approach introduces an additional refinement since the machine's level of autonomy in relation to humans is considered at different phases of the OODA loop. For example, a machine might exhibit greater autonomy in observing its environment and orienting itself, but might be dependent on humans at the decision or action stage.

Figure 1: Autonomy Spectrum and the OODA Loop[11]

In Figure 1 above, which is a subsystem of AFLR scale - developed horizontally, one can see how a machine might operate at Level 10 at the *observe* stage and thus be cognizant of all objects in its environment. However, the machine might only achieve Level 5 at the *decide* stage, enabling it to avoid collisions with objects in its environment, but still needing human assistance to realize more significant objectives.

D) Autonomy of UGV (Unmanned Ground Vehicle) according to American Army Science Board

Some other metrics have been developed for measuring the autonomy of UGVs (Unmanned Ground Vehicle).[12] In this respect, the most

---

[8] The concept was developed by fighter pilot and strategist John Boyd, for decision-making in air combat. It is called the OODA (observe, orient, decide, act) loop, where: observe means to collect current information from as many sources as possible; orient means to analyze the information, and use it to update the current reality; decide means to determine a course of action; act means to follow through with a decision. FRANS P.B. OSINGA, SCIENCE, STRATEGY AND WAR: THE STRATEGIC THEORY OF JOHN BOYD, 235 (2006).

[9] *See* Paul Scharre, *Robotics on the Battlefield Part I: Range, Persistence and Daring*, CENTER FOR A NEW AMERICAN SECURITY, 13 (2014), *available at* http://www.cnas.org/range-persistence-daring (stating that "[m]achines that perform a function for some period of time, then stop and wait for human input before continuing, are often referred to as 'semiautonomous' or 'human in the loop.' Machines that can perform a function entirely on their own but have a human in a monitoring role, with the ability to intervene if the machine fails or malfunctions, are often referred to as 'human-supervised autonomous' or 'human on the loop.' Machines that can perform a function entirely on their own and humans are unable to intervene are often referred to as 'fully autonomous' or 'human out of the loop.'"

[10] *Id*, at 18.

[11] Marra & McNeil, *supra* note 3 at 26.

[12] *See generally* Hui-Min Huang, et al*., Autonomy Levels for Unmanned Systems (ALFUS) Framework, Volume II: Framework Models*, NIST Special Publication 1011-II-

sophisticated initiative was accomplished by the American Army Science Board Study.

Figure 2. Autonomy frame of the Army Science Board Study[13]

The observe and orient stages from the AFLR's OODA loop were fused into the orient perception/situation awareness stage of this scale, as seen in Figure 2 above. The matching between the two frames is not quite one-to-one, but this new scale gives a clear hint of the autonomy requirements for an UGV.

E) FURTHER REFINEMENTS IN MEASURING AUTONOMY: THE IMPORTANCE OF COMPLEXITY OF ENVIRONMENT AND TASK

From a close inspection of UAV or UGV metrics, it becomes clear that the environment in which autonomous systems must operate is important in establishing the level of autonomy within the corresponding scale. It is intuitively clear that robotic systems operating at sea (or undersea) are faced with a different and less complex environment than the robots crossing crowded city roads.

However, a closer look at both metrics reveals another important issue for the level of autonomy; they both deal with measures of autonomy only in the navigation of specific environments (air or ground). The autonomous task/mission considered in both cases is the navigational task. Therefore, it would be meaningless to refer to a machine as "autonomous" or "semi-autonomous" without identifying the relevant task (mission). In fact, a machine that might be "fully autonomous" for one task, such as navigation along a route, might be fully human-controlled for another task, such as gathering information.[14] For example, a UAV might have autonomous control over its flight path, but it might only be remotely operated for firing a missile to an enemy—the human operator would retain the absolute control of when and at whom to fire (the kinetic attack task). In conclusion, the task/mission should be considered as well in creating an inclusive frame for levels of autonomy in robots.

Huang attempted to develop a more complete framework and metrics for robot autonomy by taking into account all three dimensions: (1) Human interaction/interface; (2) Task/mission complexity; (3) Environmental difficulty.[15] According to this framework, the robot's autonomy must be determined by levels on three axes as depicted below (Figure

---

1.0 Version 1.0 (2007), *available at* http://www.nist.gov/el/isd/ks/upload/ALFUS-BG.pdf.

[13] *Id.* at 28.

[14] Scharre, *supra* note 9 at 13.

[15] Huang, et al., *supra* note 12 at 7.

3).

Figure 3. Autonomy Levels for Unmanned Systems (ALFUS) model of autonomy[16]

For example, the environmental complexity axis in Figure 3 takes into account the environmental difficulty in measuring the autonomy. It would be a huge difference between navigating an unstructured but static environment such as a desert crossing, versus a dynamic but structured environment such as city roads. An added difficulty would be encountered in environments where maps provide little guidance or where both cooperative and hostile agents proliferate. The model developed by Huang is a clear advancement because it clarifies the multidimensional nature of autonomy. However, the authors have yet to determine exactly how to compute a general autonomy level (i.e., by determining the average of the scores along each of the axes).

A recent Final Report of the Defense Science Board (DSB) Task Force, regarding the role of autonomy, questions the whole effort by recognizing that:

> […] many of the DoD–funded studies on 'levels of autonomy'...are not particularly helpful to the autonomy design process. These taxonomies are misleading…Cognitively, system autonomy is a continuum from complete human control of all decisions to situations where many functions are delegated to the computer with only high-level supervision and/or oversight from its operator. Multiple concurrent functions may be needed to evince a desired capability, and subsets of functions may require a human in the loop, while other functions can be delegated at the same time. Thus, at any stage of a mission, it is possible for a system to be in more than one discrete level simultaneously.[17]

One must agree with the ideas of the study. However, our goal is to understand the challenges and the thresholds for implementing autonomy within robots. To achieve that, by a qualitative assessment of performances in "state-of-the-art" (military) robots, the metrics developed by the DoD are indispensable.

The next chapter will start this analysis in relation to autonomous navigation. The results will allow an extrapolation for autonomous performances required from robots executing more challenging tasks, such as ISR missions or lethal missions.

---

[16] *Id*, at 23.

[17] TASK FORCE REPORT, *supra* note 6 at 13.

II. UNDERSTANDING THE TECHNICAL THRESHOLDS FOR IMPLEMENTING AUTONOMY IN MILITARY ROBOTS

1. UGV AS ROBOT NAVIGATING THE URBAN ROADS: THE CASE OF GOOGLE'S SELF DRIVING CAR

There is little information about the architecture of Google's self-driving car, but certain elements might be gathered from the history of the project. Its origins are in DARPA's Challenges, two major US competitions leading to the development of autonomous ground vehicles.

The 2005 Grand Challenge required autonomous vehicles to cross-sections of California's Mojave Desert.[18] The vehicles were provided with GPS coordinates of way-points along the path, while the terrain was completely unknown to the designers, and the vehicles moved autonomously at speeds of 20 to 30 mph.[19] In 2007, the Urban Challenge required autonomous vehicles to travel in a simulated urban environment (a mock city at George Air Force base in Victorville, California), in the presence of other vehicles and signal lights, while complying with traffic laws.[20]

The winner of the latter race was Carnegie Mellon's team, led by Chris Urmson, followed by Stanford's team, led by Sebastian Thrun.[21] Thrun was the initiator of Google's car project and was later replaced by Urmson, who is still its leader today.[22] It could be reasonably presumed that the initial project was largely inspired by Stanford's competition car.[23]

a) The decision steps of Google's car during urban driving

Google's car software architecture is designed as a data-driven pipeline in which individual modules process information asynchronously. The time delay between entries of sensor data into the pipeline to a vehicle's actuators is approximately 300ms.

Google's car might follow through six steps in its driving decisions on urban roads.[24] The first step for the car is to locate itself.[25] The on-

---

[18] THE DARPA URBAN CHALLENGE: AUTONOMOUS VEHICLES IN CITY TRAFFIC, 307 (Martin Buehler et all eds., 2009).

[19] *Id.* at 233.

[20] *Id.* at 651.

[21] *Driven to Innovate*, CARNEGIE MELLON UNIVERSITY, (last visited Nov. 14, 2015), www.cmu.edu/homepage/computing/2010/fall/driven-to-innovate.shtml

[22] *Id.*

[23] *See Id.*

[24] *See* Eric Jaffe, *The First Look at How Google's Self-Driving Car Handles City Streets*, THE ATLANTIC CITY LAB (Apr. 28, 2014), http://www.citylab.com/tech/2014/04/first-look-how-googles-self-driving-car-handles-city-

board computer of the car collects sensor data from radar, lasers, and cameras, and integrates them to orient itself in the world *via GPS*, and in the streets with *special embedded 3D maps*.[26] These maps are 3D digitizations of the physical world, including extremely small details like the position and the precise height of every curb, traffic sign, etc. Special Google teams travel in advance and pre-scan the roads that the car will travel and then create these 3D maps.[27]

In the next step, the car determines the obstacles on the road.[28] The car collects sensor information from its radar, lasers, and cameras. Based on this data (and after comparing it with the 3D map), the vehicle determines and identifies the obstacles in urban environments as static obstacles or moving obstacles.[29]

In the third step, the car "classif[ies] this information as actual objects that might have an impact on the car's route — other cars, pedestrians, cyclists, etc. — and to estimate their size, speed, and trajectory."[30] Technological advancements in **machine vision** (in relation to deep learning algorithms based on artificial neural networks) have facilitated the car's ability to classify the objects around it.[31] While in the beginning it was difficult to distinguish a car from a pedestrian, the system is now able to make the difference.[32]

In the next step, the information enters into a probabilistic prediction model which evaluates what these objects are doing now and estimates what they will do next.[33] The prediction is based, mostly, on consistent constraints for objects, which are located on the road. For example, at every intersection where a driver has choices for changing lanes, several hypotheses are created. Whereas some drivers' behavior could be easily described and programmed, the engineers also allow the car *to learn* from other drivers' behavior.

The great technological advancement which allowed Google's car to drive on urban roads relates to machine learning; this learning is the

---

streets/8977/ (showing the 6 steps taken from a comprehensive description of the system made in The six steps were taken from a comprehensive description of the system).

[25] *Id.*

[26] *Id.*

[27] *Id.*

[28] *Id.*

[29] Eric Jaffe, *The First Look at How Google's Self-Driving Car Handles City Streets*, THE ATLANTIC CITY LAB (Apr. 28, 2014), http://www.citylab.com/tech/2014/04/first-look-how-googles-self-driving-car-handles-city-streets/8977/

[30] *Id.*

[31] *Id.*

[32] *Id.*

[33] *Id.*

ability of computer systems to interpret data and to solve problems on their own.[34] Google likely uses machine learning algorithms with some reinforced learning to create models of people or vehicles on the road. The machine learning process may be equally accomplished within virtual simulations of street traffic.[35] The engineers can even register circumstances when the human driver turned off the autonomous driving and determine what would have happened if the car had acted alone.[36]

b) The (real) autonomy level of Google's car and its meaning

Although Google's car could be considered at Level 7 of autonomy—on the scale for UGV, developed by American Army Science Board Study (see figure 2)—a deep assessment of the car's abilities would lower its autonomy to Level 5, at most. In fact, the false higher autonomy level is linked to the "ruse" of using the 3D pre-scanned maps.[37]

While all autonomous cars of today rely on basic electronic maps for navigation and lane-centering, Google's car uses, as we have seen, far more detailed 3D maps. The combination of GPS with the fine aligned 3D pre-scanned maps lowers tremendously the needed performance in "perceptual processing" of data about urban environment, since the car does not need to process entire surrounding scenes from scratch.[38]

Metaphorically speaking, the need to "understand/recognize" the world's elements just disappears for Google's car. Its "perceptual-

---

[34] Alexis C. Madrigal, *The Trick that Makes Google's Self-Driving Cars Work*, THE ATLANTIC (May 15 2014), http://www.theatlantic.com/technology/archive/2014/05/all-the-world-a-track-the-trick-that-makes-googles-self-driving-cars-work/370871/

[35] Google has re-built the complete California 172,000-mile road system in its software, including accurate simulations of weather, traffic, pedestrians, etc. Sebastian Anthony, *Google has built a Matrix-like simulation of California to test its self-driving* cars, EXTREMETECH (August 22, 2014), http://www.extremetech.com/extreme/188482-google-has-built-a-matrix-like-simulation-of-california-to-test-its-self-driving-cars.

[36] Alexis C. Madrigal, *supra* note 34.

[37] Lee Gomes, *Urban jungle a tough challenge for Google's autonomous cars,* MIT TECHNOLOGY REVIEW (July 24, 2014), available at: http://www.technologyreview.com/news/529466/urban-jungle-a-tough-challenge-for-googles-autonomous-cars/.

[38] *See* Comment to *How do Google's self-driving cars work?,*QUORA (Jul. 13, 2011), http://www.quora.com/How-do-Googles-self-driving-cars-workSee (stating that "[…] laser sensors are used to create a 3d point cloud of the surroundings. In this scan it is actually quite easy to extract lane markers due to the fact that lane markers produce a higher intensity value in the laser scan. In addition to the lane detection they also have the ability to do full alignment of the point clouds. In order to do this they first drive through a patch of road collecting point cloud of the surroundings for many successive frames. These point clouds are then aligned algorithmically[.] This creates a full 3d model of that patch of road and you can then align the 3d model to satellite imagery. […] [W]hen they drive through that road again they can take the [new] laser scan and align the scan to the existing 3d model to find and estimate where they are also[.]"; *see also* Madrigal, *supra* note 34.

visual"[39] processing system should not deal with the complexity of surroundings (traffic markings, trees, different obstacles, etc.). It will only detect and process data about "new" elements-obstacles, which do not appear on the 3D map.[40]

As such, the system has clear limitations.[41] The most important one is the use of "brute (computing) force" to circumvent its blindness. Therefore, it is difficult to imagine it as being scaled up to cover the entire world. However, we could try a thinking experiment: to attempt to 3D pre-scan all the roads of the world. Google's street view project, multiplied by tens of thousands, gives us a hint of the magnitude of required effort.

Furthermore, if the robots will enter our homes to make deliveries or to execute menial tasks, for example, they will need finer and higher resolution 3D maps of the inside of all the buildings.

With this approach, the robots will become autonomous only if we can get an almost perfect and systematically updated virtual copy of the real world. Perhaps the solution is to give any object of the future significant deep 3D mapping abilities. These objects will do the "hard work." The Tango program of Google, which integrates 3D scan abilities to smart phones, is just a small step in that direction.[42]

However, this sort of highly regular, geometrical, and artificial virtual copy might exist as long as the real world lives in peace. The war, by its very own nature, is "fog": uncertainty, disruptions, and destructions that will instantly turn off such a highly detailed virtual copy of the world.[43] Military robots cannot rely on this approach and must find another pathway toward autonomy.

---

[39] This is only a metaphor, since the main sensor of the car is a laser (LIDAR).

[40] *See* Alexei Oreskovic, *Silicon Valley debate on self-driving cars: do you need a map?,* REUTERS (March 6, 2015),http://www.reuters.com/article/2015/03/06/us-autos-selfdriving-technology-analysis-idUSKBN0M20EK20150306)(stating that "[t]he map-based approach allows you to drive accurately in a controlled environment, where you know that things don't change much[.]"

[41] The existing 3D maps, the know-how and tremendous resources provide Google with an obvious advantage. However, these 3D maps can easily become obsolete if, for example, fresh fallen snow or even rain might change the landscape. And if a traffic light, a stop sign or an intersection has changed, even slightly, the database of 3D maps must be updated too. *Id.*

[42] The "Project Tango" prototype is an Android smartphone-like device that tracks the 3D motion of the device, and creates a 3D model of the environment around it. *About Project Tango*, GOOGLE, https://www.google.com/atap/project-tango/about-project-tango/ (last visited August 3, 2015).

[43] BARRY D. WATTS, CLAUSEWITZIAN FRICTION AND FUTURE WAR, INSTITUTE FOR NATIONAL STRATEGIC STUDIES, 1-2 (2004), *available at* http://www.clausewitz.com/readings/Watts-Friction3.pdf

c) The significance of Google's car limitations for military UGV navigation: the restrictions in visual intelligence of robots

We have seen that Google's car navigates in the *dynamic but structured environment* of urban roads. We have also discovered that Google's car, as a state-of-the-art system, does not have the perception or situational awareness to operate by itself in such an environment. It can only overcome this threshold by combining the navigation based on GPS with the pre-scanned highly detailed 3D maps.

The autonomous navigation requirements of a military UGV, an autonomous military truck, for example, are much greater than those of Google's car. Most of the time, military UGVs operate in environments that are both highly dynamic and structured/unstructured, where existing maps provide little guidance; where its GPS is not always accessible because of the jamming or decoy; and where hostile agents are highly active. A military UGV could not use any pre-scanned 3D maps and might, eventually, rely on inertial navigation. Given the limitations of Google's "state-of-the-art" perceptual intelligence, it must be clear that today's military UGV are not up to the (navigational) task.[44]

Experts confirm that "UGV navigation in urban environments, in dense foliage, off road and with people remains nascent" since there is a "lack of high-speed obstacle detection in complex terrain."[45] Where "these specialized range sensors permit rapid identification of surfaces for navigation, it is not sufficient to permit the UGV to determine the difference between…a bush that it can run over, tall weeds that indicate a drop off into a creek bed underneath, and the presence of a rock among the weeds that would damage it."[46] In fact, the actual systems are far from the visual abilities of birds or even flying insects which are able to perform well without using predetermined waypoints or an external position reference system.

As seen previously, in relation to GPS waypoints and 3D maps, Google's self-driving car relies on range sensors and laser scans (LIDAR) to navigate.[47] The visual processing is realized only during the later stage of driving decision, when the car must differentiate categories of mobile obstacles (cars, bicycles, passengers, etc.). That was the only solution since the visual systems of today are only classificatory systems[48] with no *deep field vision.* Their performances are far behind

---

[44] For example, a Carnegie Mellon's car, which used only visual navigation system, required sensors mounted on street lights to make its 2013 autonomous journey to the airport, not on the urban roads. Oreskovic, *supra* note 40.

[45] TASK FORCE REPORT, *supra* note 6 at 37.

[46] *Id.*

[47] Jaffe, *supra* note 24.

[48] "What people are doing today in computer vision isn't really vision—they're do-

those of their biological counterparts.[49]

This current limitation of "visual intelligence" in the autonomous navigation systems is indirectly acknowledged by a recent DARPA announcement of a Fast Lightweight Autonomy (FLA) program.[50] This four-year program tries to give small UAVs (without GPS) the abilities of birds or flying insects that navigate at high speeds in cluttered environments.[51] It investigates software approaches that will enable revolutionary improvements to state of the art visual processing; such improvements include perception approaches such as rapid evaluation and recognition of previously visited areas using landmark recognition, and algorithms to localize and navigate relative landmarks or other visually distinctive features. [52] If successful, the developed algorithms could impact a wide range of unmanned systems in navigation through cluttered environments (such as a UGV in a city).

More important scientific breakthroughs in visual intelligence would be necessary to provide the robots with "humanlike" navigation abilities. This is the idea of the Intelligence Advanced Research Projects Activity (IARPA), which recently announced MICRONS, a program trying to reverse-engineer human brain algorithms in sensory information processing.[53]

IARPA seeks to significantly improve artificial intelligence and machine learning technologies since "today's state of the art algorithms are brittle and do not generalize well […][while] in contrast, the brain is able to robustly separate and categorize signals in the presence of significant noise and non-linear transformations, and can extrapolate from single examples to entire classes of stimuli." [54] The MICRONS program is based on "targeted neuroscience experiments that interrogate the op-

---

ing image classification. That's a subset of vision. 'Here's an image; what is it?' Caleb Garling, *Jeff Hawkins on Firing up the Silicon Brain,* WIRED, http://www.wired.com/2015/05/jeff-hawkins-firing-silicon-brain/, (last visited June 5, 2015).

[49] Very complex mechanisms are implicated in insects' navigation. Bees possess a mechanism for learning the solar ephemeris (the position of the sun in the sky) for the time of day, as measured by an internal clock of some sort. They have also mechanisms for dead reckoning, which requires the integration of two types of information (direction and speed). In this case it is the 'visual flow' that seems to be used as the measure of distance travelled. PETER CARRUTHERS, THE ARCHITECTURE OF THE MIND, 96 (2006).

[50] *Fast Lightweight Autonomy (FLA)*, FEDERAL BUSINESS OPPORTUNITIES, *available at* http://www.darpa.mil/Our_Work/DSO/Programs/Fast_Lightweight_Autonomy_%28FLA%29.aspx, (last visited April 18, 2015)

[51] *Id.*

[52] *Id.*

[53] *Machine Intelligence from Cortical Networks (MICrONS),* IARPA, http://www.iarpa.gov/index.php/research-programs/microns (last visited April 18, 2015).

[54] *Id.*

eration of mesoscale cortical computing circuits, taking advantage of emerging tools for high-resolution structural and functional brain mapping."[55]

This program has the potential to significantly advance the visual intelligence of robots. However, at this time, the complete success of both DARPA and IARPA initiatives is uncertain.

In conclusion, the military UGVs' ability to navigate the dynamic but structured environments of urban roads does not exist today, or in the foreseeable future, due to limitations in perceptual and visual intelligence of robots. Nevertheless, one may still assess the autonomous navigation in other possible settings, of (slightly) structured and/or static environments. Such spaces were the perceptual and visual processing is less needed might be found during navigation on the sea (or undersea) or in navigation through air. Intuitively, a robotic system with the current capabilities of Google's car would be able to autonomously navigate in such environments.

2. TOWARD AUTONOMOUS NAVIGATION IN OTHER UXV

a) The first real autonomous UMV: the ACTUV

DARPA financed the development of the Anti-Submarine Warfare Continuous Trail Unmanned Vessel (ACTUV), an unmanned maritime vehicle (UMV), optimized to track the most quiet diesel electric submarines.[56]

Certain countries use cheap diesel-electric submarines as anti-access/access-denial (AA/AD) components against US carriers.[57] With ACTUV, the US Navy will be equipped with an anti-submarine detection system (an autonomous submarine "hunter"), which will be much cheaper than any diesel submarine.[58]

This program integrates highly autonomous (with light remote supervisory control) navigational tasks within missions spanning a range of thousands of kilometers and for months in duration.[59] It includes autonomous compliance with maritime laws and conventions for safe navigation, autonomous system management for operational reliability,

---

[55] *Id.*

[56] *Anti-Submarine Warfare (ASW) Continuous Trail Unmanned Vessel (ACTUV) "Sea Hunter,"* NAVAL DRONES, http://www.navaldrones.com/ACTUV.html (last visited April 18, 2015); Tamir Eshel, *Autonomous ASW: The Predator Becoming a Prey*, DEFENSE UPDATE (Jan 1, 2013), http://defense-update.com/20130101_saic_develops_an_unmanned_submarine_hunter.html.

[57] *Anti-Submarine Warfare*, *supra* note 56.

[58] *Id.*

[59] *Id.*

and even autonomous interaction with an intelligent adversary.[60]

Besides its autonomous navigation tasks, the ACTUV employs non-conventional sensor technologies for autonomous ISR (Intelligence, Surveillance, and Reconnaissance), allowing it to track the submarines over their entire operating activity, even by following them into harbors.[61]

DARPA allowed Leidos, a national security, health, and engineering company, to build a mock-up ACTUV as a "Trimaran" made of carbon composites and equipped with navigation and piloting sensors, electro-optics, and long/short-range radar.[62]

b) Toward navigation autonomy in UAVs

The most common UAVs today reveal characteristics both of automation and autonomy. For instance, "the Global Hawk reconnaissance drone has the ability to take off and land unassisted."[63] For other functions, the human operator can choose among different levels of autonomy.[64] These systems are at Level 0, or at most 1, on the AFLR's autonomy scale.[65]

However, some state of the art systems might reach higher autonomy levels. Since the air is a less complex environment than the one Google's car must navigate through, highly autonomous navigation seems attainable by today's most advanced UAVs.

That is the case for BAE System *Taranis*, a British demonstrator program, which, in addition to radar invisibility, includes artificial intelligence and high levels of autonomy.[66] The system showed its capacity to autonomously take off, navigate to the target, detect a target, generate a plan for flying to the target, search for it, and then return to base.[67] It relies on onboard electronic maps to identify a target's position when GPS signals are jammed.[68] The vehicle also carried out simulated attacks and post-attacks damage assessment.[69]

Another highly autonomous system is the Northrop Grumman X-

---

[60] *Id.*

[61] *Id.*

[62] On November 18, 2014, Leidos announced that a test of a surrogate vessel mimicking the final configuration of the ACTUV, completed 42 days of at-sea demonstrations to fulfil collision regulations. *Id.*

[63] Marra & McNeil, *supra* note 3 at 28.

[64] *Id.*

[65] *Id.*

[66] *What we do, BAE SYSTEMS, http://www.baesystems.com/en/product/taranis* (last visited Apr. 18, 2015).

[67] *Id.*

[68] *Id.*

[69] *Id.*

47B,[70] which is designed for carrier-based operations. Developed under DARPA's supervision, as part of the United States Navy's Unmanned Combat Air System Demonstration (UCAS-D) program, it has successfully performed a series of land-based and carrier-based demonstrations.[71] Both of these systems are, at most, at Level 5 on the AFLR autonomy scale.

3. THE SEARCH FOR AUTONOMY IN ISR (INTELLIGENCE, SURVEILLANCE AND RECONNAISSANCE) FOR UXV

a) UMVs

As seen previously, certain UMVs with autonomous navigation capabilities are able to autonomously track the quietest diesel submarines over their entire operating envelope. Therefore, ACTUV will necessarily have high autonomous ISR capacities in relation to its main mission. However, the ISR capabilities required for discovering and discriminating targets within an undersea environment are far lower than those required, for instance, in a city environment. This illustrates the current feasibility of this project.

b) UAVs

In the case of UAVs, the ground ISR's requirements for a plane are far greater. However, promising work has been done in human interaction with computer vision processing, allowing the reduction of manpower and cognitive workload in ISR missions.[72] The automatic/autonomous ground-based surveillance is the goal of DARPA's programs such as Gorgon Stare, Argus-IS,[73] and Mind's Eye.[74]

---

[70] *X-47B UCAS Makes Aviation History…Again!,* NORHTROP GRUNMAN, http://www.northropgrumman.com/Capabilities/X47BUCAS/Pages/default.aspx, (last visited April 18, 2015).

[71] *Id.*

[72] For the manpower requirements of the system. *See* Mark Thompson, *Manning Unmanned Systems*, TIME (Sept. 10, 2012), http://nation.time.com/2012/09/10/manning-unmanned-systems/.

[73] The ARGUS-IS (Autonomous Real-Time Ground Ubiquitous Surveillance Imaging System) is an advanced camera system that uses hundreds of cell phone cameras in a mosaic, to video and auto-track every moving object within an area. Damien Gayle, *The incredible U.S. military spy drone that's so powerful it can see what type of phone you're carrying from 17,500ft*, DAILYMAIL (Jan. 28, 2013), www.dailymail.co.uk/sciencetech/article-2269563/The-U-S-militarys-real-time-Google-Street-View-Airborne-spy-camera-track-entire-city-1-800MP.html.

[74] "The DARPA Mind's Eye program seeks to develop in machines a capability that currently exists only in animals: visual intelligence. In particular, this program pursues the capability to learn generally applicable and generative representations of action be-

The Mind's Eye project attempts to analyze images with software capable of recognizing human activities in videos and elaborating predictions.[75] It is possible that the system "translates" certain tactically significant actions detected from the air (by Gorgon Stare and Argus-IS systems) in tags brought to the attention of human operators.

Given the limits of today's computer visual intelligence, for example, with Google's car, it is obvious that such ground ISR systems will have "the man in the loop," and will remain, at most, semi-autonomous.

4. TOWARD UXVS AS AUTONOMOUS WEAPON SYSTEMS (AWS)

a) UMVs as AWS

We have seen that ACTUV will be an autonomous UMV, both in maritime navigation and in tracking the quietest diesel submarines. Adding a lethal autonomous mission to such a submarine hunter will transform it into an autonomous submarine destroyer, a truly Autonomous Weapon System. From a technical viewpoint and given the actual limitation in state-of-the art systems, such a step seems easy to accomplish and might be achieved within the foreseeable future.

b) UAVs as AWS

The current UAVs are not autonomous in the firing of weapons because their weapons are always fired in real-time by human controllers. Presently, there are no known plans or reasons to take the human element out of the weapons firing loop.

However, there is a real need for armed and autonomous UAV or UCAV (Unmanned Combat Air Vehicles).[76] In the Western Pacific, there is a build-up of anti-access/area-denial (A2/AD) capabilities based on diesel submarines (as already mentioned), on missiles, communications jamming, cyber-warfare tools, or anti-satellite weapons.[77] Among them, the anti-ship cruise or ballistic missiles may force the US Navy's air-

---

tween objects in a scene, directly from visual inputs, and then reason over those learned representation". H.L.H. de Penning, et al., *A Neural-Symbolic Cognitive Agent with a Mind's Eye* (2012), *available at* https://www.aaai.org/ocs/index.php/WS/AAAIW12/paper/viewFile/5265/5652

[75] *Id.*

[76] For a compelling analysis of future naval conflicts, *see* Andrew F. Krepinevich, *Maritime Competition in a Mature Precision-Strike Regime* (April 13, 2015), *available at* http://csbaonline.org/publications/2015/04/maritime-competition-in-a-mature-precision-strike-regime/.

[77] Dave Majumdar, *Essay: The Legal and Moral Problems of Autonomous Strike Aircraft*, USNI NEWS (Aug. 21, 2014), news.usni.org/2014/08/21/essay-legal-moral-problems-autonomous-strike-aircraft.

craft carriers to stand a great distance (more than 1,000 nautical miles) from an enemy's seacoast.[78] In this scenario, the existing (manned) stealth aircrafts may not have the range or the survivability required to operate within this space.[79]

One option for the US Navy is to develop a long-range unmanned strike aircraft with stealth technology capable of penetrating the thickest of enemy's air defenses. Since an advanced enemy might deny or degrade communications through jamming and decoys, such an aircraft must be fully autonomous. It should navigate independently of communications with human controllers and should detect and make decisions to release weapons without long distance consent by a human. This explains the need for the Unmanned Carrier-Launched Surveillance and Strike (UCLASS) program.[80] In one of its versions, still under discussion,[81] the UCLASS aircraft would be an UCAV with air–ground/sea attack capabilities. We will examine the necessary autonomous performance of such UCAV in relation to state-of-the-art robots' abilities.

Such an aircraft requires defensive capacities against surface-to-air missiles that might be encountered. In a normal combat plane, human pilots decide whether to engage this kind of threat.[82] However, human pilots make this decision based on sensor information processed by the aircraft's computers. The Lockheed Martin F-22 Raptor or F-35 Joint Strike Fighter both illustrate this concept primarily in beyond visual range air-to-air combat.[83] Both planes join data from the aircraft's sensors into a trail file that the computer identifies as hostile, friendly, or unknown.[84] Hence, the pilot is entirely reliant on computers to establish the appropriate combat identification. It might be just a small technological step to let such a system engage targets autonomously, with no human intervention.

The next step would be to build an UCAV capable of autonomously

---

[78] *Id.*

[79] *Id.*

[80] *Unmanned Carrier Launched Surveillance and Strike (UCLASS) Program*, www.navaldrones.com/UCLASS.html, (last visitied Nov. 24, 2015). Certain UAV systems, such as the BAE Taranis and Northrop Grumman X-47B, from the autonomous navigation perspective, seem to attain such performnces as (semi) autonomous weapon systems.

[81] There are two competing ideas for UCLASS: a semi-stealthy aircraft with sufficient endurance to operate within normal carrier air wing operations, provide ISR and light strike in lightly contested environments; and a more capable aircraft with air-to-air refuelling capability designed to operate in denied airspace for penetrating surveillance and strike missions. *Id.*

[82] Majumdar, *supra* note 77.

[83] *Id.*

[84] *Id.*

engaging an enemy's combat aircrafts.[85] In a recent article, *Air & Space Power Journal,* U.S. Air Force Captain Michael Byrnes defends the idea of a future unmanned autonomous air-to-air fighter, called the FQ-X.[86] Byrnes' opinion is based on the suggestion that air-to-air combat presents an extremely "sterile" environment, while acquiring and processing the data against a relatively empty background, the sky, is rather simple.[87]

We discover, again, the inverse correlation between environmental simplicity and the level of autonomy attainable for a given task in state-of-the-art robotic systems. Intuitively, the environment in visual air-to-air combat is not more complex than that of Google's car navigation of urban streets. The moving obstacles and 3D liberty movements of UAVs require certain technical capabilities. But the use of machine learning algorithms in simulated combat environments may be a few steps ahead of the technological advancements seen in Google's car. This is why such a UCAV in air-to–air engagements appears to be attainable.

The air-to-ground arena appears much more challenging.[88] Besides target location requirements or enemy camouflage, there are difficulties in "acquiring and processing sensor data against [...] a cluttered backdrop of the Earth's surface and all of the natural and manmade objects layered upon it" because this "[s]urface attack is [...] extremely context dependent."[89]

This complex environment requires certain ISR high abilities and the sort of perceptual-visual intelligence that, as we saw already, is still out of reach for foreseeable technologies. However, one can imagine ground/sea environments less complex than those of cities; for example, the visuals over the sea or over the seashores. This reduced environment complexity would lower the necessary perceptive-visual intelligence from UCAVs with autonomous air-to-ground/sea attack capabilities.

Given these restrictions, the first autonomous UCAV will most likely be a stealth UAV with certain ISR capabilities. The following step might be reached by a UCAV with air-to-air combat abilities. Most likely, the initial UCAV with autonomous air-to-ground attack capacity will be initially used for anti-ship and subsequently for anti-shore combat missions. In brief, the implementation of the UCLASS program, as a

---

[85] In air-to-air combat an UAV might have a tactical advantage since certain manoeuvres are impossible for human pilots, given the multiple G limiting factor.

[86] Michael W. Byrnes, *Nightfall machine autonomy in air-to-air combat*, AIR & SPACE POWER JOURNAL, (May–June 2014), 71, note 30, *available at* http://www.airpower.maxwell.af.mil/digital/pdf/articles/2014-May-Jun/F-Byrnes.pdf.

[87] *Id.*

[88] Majumdar, *supra* note 77.

[89] Byrnes, *supra* note 86.

highly able UCAV option, requires substantial effort, but it is within today's technological reach.

However, in addition to these technological hurdles, there are legal and ethical barriers for the implementation of autonomous combat missions by the robots. Taking the man "out of the (lethal) decision loop" obliges such autonomous weapon system (AWS) to follow the rules of humanitarian law on the battlefields. The next part of the paper will evaluate this final threshold from a techno-legal angle.[90]

III. THE LAST TECHNO-LEGAL THRESHOLD FOR AUTONOMY IN COMBAT MISSIONS: THE AUTONOMOUS WEAPON SYSTEMS MUST COMPLY WITH RULES OF HUMANITARIAN LAW

Another great obstacle for deploying autonomous weapon systems (AWS) on the battlefields is the compulsory legal and ethical requirements of the Law of Armed Conflict (LOAC) and Rules of Engagement (ROE).[91]

The LOAC contains two distinct rules of law: weapons law and targeting law. While the former embraces the rules about weapon as being lawful *per se,*[92] the latter includes the prohibited use of the weapons system and will be considered bellow.

1. THE MAIN LEGAL REQUIREMENTS OF TARGETING LAW

The four classic cumulative requirements of targeting law are military necessity, discrimination/distinction, proportionality, and humanity.[93]

a) Military necessity

While military necessity is mentioned in many LOAC treaties, it arises primarily from customary international law. It appears in the re-

---

[90] For a deep analysis about legal implications of autonomous robots *see* Kenneth Anderson & Matthew, C. Waxman, *Law and ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can,* JEAN PERKINS TASK FORCE ON NAT'L SEC. AND L. (2013).

[91] The Rules of Engagement (ROE), prescribe more exactly what is acceptable or not on the battlefield. For a comprehensive discussion of ROE, *see* GARY, D SOLIS, THE LAW OF ARMED CONFLICTS. INTERNATIONAL HUMANITARIAN LAW IN WAR, CAMBRIDGE UNIVERSITY PRESS, 490-512 (2010).

[92] For a complete *discussion see Chapter 20. General Principles on the Use of Weapons,* ICRC, https://www.icrc.org/customary-ihl/eng/docs/v1_cha_chapter20, (last visited May 2, 2015).

[93] Gary, D Solis, *supra* note 91 at 250-85.

quirement that one may target things which are not prohibited by LOAC (military objective such as persons, places, objects, etc.) and which make an effective contribution to military action. The destruction of enemy forces and material would generally meet this test.

b) Distinction/discrimination

The distinction/discrimination is another leading principle of the LOAC. Reflecting customary international law, the distinction requires a military to differentiate between combatants and civilians, as well as between military and civilian objects. This rule is codified in Article 48 of Additional Geneva Protocol I with complementary rules in Articles 51 and 52.[94]

c) Proportionality

The next targeting law precondition/principle is proportionality. Proportionality requires combatants to examine whether the probable collateral damage in the attack would be excessive compared to the expected military advantage. This principle is a custom of international law, and is codified in both Article 51(5)(b) and Article 57(2) (iii) of the Additional Protocol I of Geneva[95] relative to "an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated."[96]

d) Humanity

Customary in nature and codified in Article 57 of Additional Protocol I, the humanity precondition requires an attacker to exercise "con-

---

[94] Article 51(2): The civilian population as such, as well as individual civilians, shall not be the object of attack. Acts or threats of violence the primary purpose of which is to spread terror among the civilian population are prohibited; Article 52(1): Civilian objects shall not be the object of attack or of reprisals. PROTOCOL ADDITIONAL TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949, AND RELATING TO THE PROTECTION OF VICTIMS OF INTERNATIONAL ARMED CONFLICTS (PROTOCOL I), art. 51(2), (1977); PROTOCOL ADDITIONAL TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949, AND RELATING TO THE PROTECTION OF VICTIMS OF INTERNATIONAL ARMED CONFLICTS (PROTOCOL I), art. 52(1), (1977).

[95] *Id.*

[96] At the core of the rule of proportionality lies the standard of 'excessiveness' which is "the product of a case-by-case assessment that is evaluated in terms of its reasonableness given the attendant circumstances". N. Schmitt, Jeffrey S. Thurnher *Out of the loop: autonomous weapon systems and the law of armed conflict*, 4 HARVARD NATIONAL SECURITY JOURNAL 231*,* 255 (on file with author).

stant care…to spare the civilian population, civilians and civilian objects."[97] More exactly, the attacker is required to "do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects, and are not subject to special protection but are military objectives"; to cancel an attack if it becomes apparent that the rule of proportionality will be breached; to provide "effective advance warning" of an attack if it may affect the civilian population "unless circumstances do not permit"; or "[w]hen a choice is possible between several military objectives for obtaining a similar military advantage, [select] that the attack on which may be expected to cause the least danger to civilian lives and to civilian objects"; and to "take all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimizing, incidental loss of civilian life, injury to civilians and damage to civilian objects."[98]

2. THE FIRST TENTATIVE PLAN FOR CREATING MORAL ARTIFICIAL AGENTS ON BATTLEFIELDS AND ITS IMPLICATIONS

a) Pr. Arkin's general architecture

Ronald Arkin, a professor of Robotics at Georgia Tech University, conducted one of the most important researches in this field. In 2009, he published *Governing Lethal Behavior in Autonomous Robots,*[99] a book about designing an artificial moral agent (a robot), capable of matching or surpassing the standard in applying the rules of LOAC/ROE that we expect from human soldiers within all battlefield circumstances or operational environments.

After examining the deontological or Kantian ethical theories, utilitarian theories, cultural relativism, virtue ethics, etc., he finally considered the deontological logic as a primary source for implementing the rules of LOAC/ROE in his robotic system.[100]

The most important part in Arkin's general architecture was the ethical governor which enclosed algorithms for determining if lethal actions are ethical/legal or not. It addressed in its decision flow the issues of military necessity, target discrimination, proportionality, and the application of the Principle of Double Intention (the acting in ways which

---

[97] PROTOCOL I, art. 57(1).

[98] The essence of the verification obligation is the term "feasible" with the meaning of "practicable or practically possible, taking into account all circumstances ruling at the time, including humanitarian and military considerations." Schmitt, *supra* note 96 at 259-261.

[99] *See generally* RONALD C. ARKIN, GOVERNING LETHAL BEHAVIOR IN AUTONOMOUS ROBOTS (2009).

[100] *See generally id.*

would minimize civilian collateral damage).

b) General evaluation of Professor Arkin's project

The thresholds for the actual implementation of such general moral/ethical architecture seem to be tremendous and are located on different levels. In fact, his design was only a blueprint for a system with no direct access to the real world through sensors. Authors, such as Professor Sharkey,[101] considered it as just a "back-end system," based on information received from systems expected to be developed sometime in the future.

According to Professor Arkin himself, the translation of legal principles/rules into algorithm-compatible forms has not been accomplished yet.[102] However, we think that real difficulties lay at another level, the level of *applying* the rules of LOAC by a robot (if they will be, if ever, machine-ready) to particular circumstances.[103] To realize that, a robot must emulate the performance achieved in applying LOAC rules by human soldiers. This application by human soldiers creates a multidimensional challenge that will be analyzed below.

3. COGNITIVE AND MORAL PROCESSES IN HUMAN COMBATANTS APPLYING THE PRINCIPLES OF DISCRIMINATION AND PROPORTIONALITY OF HUMANITARIAN LAW

a) Exploring the multiple difficulties in applying LOAC rules to a given case by humans

According to classic legal theory, the application of (legal) rules to facts implies both the qualification of facts and the interpretation of rules. The problem of qualification of facts occurs when the circumstances of a case are covered by the factual part of rules. It is at this level that the *discrimination/distinction* requirements of LOAC will be applied.

The *problem of interpretation* appears when a rule is obscure and has various meanings. This is almost always the case in law, since legal

---

[101] Noel E. Sharkey, *The Evitability of Autonomous Robot Warfare*, INTERNATIONAL REVIEW OF THE RED CROSS, 790 (2012).

[102] One might question if such effort will ever bear its fruits. See for a discussion Benjamin Kastan, *Autonomous Weapons Systems: A Coming Legal "Singularity?"* 1 J. L. TECH & POL'Y 45 (2013).

[103] See the remarks of Pr.Sharkey: "[the system] has no direct access to the real world through sensors or a vision system…There is neither a method for interpreting how the precepts of the laws of war apply in particular contexts nor is there any method for resolving the ambiguities of conflicting laws in novel situations", Sharkey, *supra* note 101 at 790.

imperatives are made in natural, everyday language. A different problem appears when legal rules contain standards (such as reasonableness, fairness, good faith, etc.) which, as concepts linked to social reality of a given space or time, are highly context dependent.

The rules of LOAC have all of these characteristics, since they are very often obscure and contain many standards. More than that, the rules of LOAC contain *legal balances of proportionality* (as expressed by the preconditions-principles of proportionality and humanity). In general, the reasoning with the *balance of proportionality* is very context dependent and is among the most complex in the legal thinking.[104]

All of the above problems in applying the rules of LOAC to a case are part of a hermeneutical process, which, while applied by the courts (for example, in the *post factum* context of establishing responsibility for infringing humanitarian law), requires the highest legal expertise. The discursive and slow reconstruction, realized by the judges, of the legal and moral reasoning seems different from the quick moral decision-making by the combatants on the battlefield. However, the processes might be similar and the difference exists, most probably, on the level of supporting human cognitive architectures: conceptual rationality vs. high-speed intuitive/emotional abilities[105] (with learned skills acquired during long military drills).[106]

In the following section we will examine the probable architecture and mental processes in human soldiers applying the rules of LOAC on the battlefield.

b) Modular organization of human mind

To understand the implementation efforts needed for AWS in com-

---

[104] In relation to the balance of proportionality in legal *thinking see generally* Remus Titiriga, *The 'Jurisprudence of Interests' (Interessenjurisprudenz) from Germany: History, Accomplishments, Evaluation,* 3.1 INT'L. J. L. LANGUAGE AND DISCOURSE 1 (2013); Robert Alexy, *On Balancing and Subsumption: A Structural Comparison*, 16 RATIO JURIS 433 (2003); Alex Aleinikoff, *Constitutional Law in the Age of Balancing*, 96 YALE L.J. 943 (1987).

[105] Psychological researches have shown that, during real time missions, the highly trained experts, such as firefighters, do not take decisions by weighing alternatives, but by moving forward and trying sequentially different hypotheses. See for further details GARY KLEIN, SOURCES OF POWER: HOW PEOPLE MAKE DECISIONS (1998).

[106] Hauser called this first form *principled reasoning* which is "slow, deliberate, thoughtful, justifiable, requires considerable attention, appears late in development, justifiable, and open to carefully defended and principled counterclaims". He also describe ethical decisions with a different architectural design based upon *intuitions* which are "fast, automatic, involuntary, require little attention, appear early in development, are delivered in the absence of principled reasons, and often appear immune to counter-reasoning". MARC HAUSER, MORAL MINDS: HOW NATURE DESIGNED OUR UNIVERSAL SENSE OF RIGHT AND WRONG (2006).

plying with the rules of LOAC, one must identify, in the most general way, the cognitive processes of human combatants applying these rules on the battlefield. While certain aspects of the "general architecture" of human mind are still mysterious, cognitive scientists are increasingly recognizing its modular structure.[107]

At the most basic level, there are modules that humans share with animals. For example, certain basic forms of perception/belief/desire/planning/motor-control psychology are shared even with invertebrates, such as insects[108]. Besides these common modules there are certain components specifically human. It seems that humans are unique (at least unique in the sophistication of these abilities) in possessing twenty-two special mental capacities.[109]

Relatively little is known about ethical reasoning implied in the production and control of behavior, although there are certain recent advancements. We can try to make an educated guess about the human mind modules needed to achieve these moral/legal abilities by combatants on the battlefield.

Since our purpose is to qualitatively identify the thresholds for creating artificial moral agents able to comply with LOAC, we will examine only the most important requisites—the discrimination and proportionality assessment abilities required of human combatants, fighting on the most complex battlefield (i.e. urban combats against insurgents).

c) The cognitive modules (eventually) implied in discrimination/distinction assessment by humans

Obviously, humans have the visual intelligence to differentiate the elements of the environment, and have deep field awareness. But this capability is just the first level of visual intelligence, the one that humans share with many animals (including certain invertebrates). Perceptual abilities and visual intelligence of insects, bees for example, are quite impressive. They can distinguish between elements of their environment (flowers, members of the same hive, enemies, food, etc.) and their visual intelligence is far greater than that of the most advanced robots (such as Google's car).

However, the discrimination/distinction requirements of LOAC rules are even higher than that. For example, in certain circumstances, one must distinguish between a combatant and a non-combatant (for example, if insurgents pretend to be civilians). Added to this, the military must discriminate between active combatants and wounded ones

---

[107] *See generally,* PETER CARRUTHERS, THE ARCHITECTURE OF THE MIND, OXFORD UNIVERSITY PRESS (2006).

[108] *See generally id.*

[109] *See generally id.*

who are unable to fight or who have surrendered.[110] This level of discrimination/distinction must involve specific human abilities and cognitive modules. The military must understand social interactions, the social cues, and deduce the intention of combatants and non combatants (e.g. the surrender with a white flag).

At this level, two cognitive modules, specifically human, might become essential:

(1) the folk (physics capacity, enabling "deeper" causal reasoning of physical phenomena);[111] and (2) the mind (reading capacity ("theory of mind") which allows humans to attribute mental states to other people, and to predict their probable actions.)[112]

d) The cognitive modules (eventually) implied in proportionality assessment by humans

Proportionality of LOAC requires combatants to examine whether the likely collateral damage of the attack would be excessive compared to anticipated military advantage. As with the distinction, the judgment of proportionality is highly dependent on the environment and the battlefield in which the military are deployed. Combatants need to imagine or represent the alternative future outcomes and choose the one answering the requirements of proportionality. Intuitively speaking, the human mind will use its unique episodic memory for a sort of "time travel," trying to represent different variants of the future and to choose the best one. [113] Some of the following mental capacities/modules are required in proportionality-balancing reasoning: a capacity to entertain suppositions used in counter-factual and hypothetical thinking (and also in mind-reading);[114] within specific circumstances, there might be also a need for capacity to think creatively and generate novel ideas, novel hypotheses, and novel solutions to problems; and a capacity for similarity-based and analogical thinking and reasoning, manifested in human tendency to use one domain as a model for the operations of a less-familiar system.[115]

---

[110] *See generally* PATRICK LIN, GEORGE BEKEY & KEITH ABNEY, AUTONOMOUS MILITARY ROBOTICS: RISK, ETHICS, AND DESIGN (Ethics Emerging Sciences Group at California Polytechnic State University, 1.0.9 ed. 2008).

[111] Carruthers, *supra* note 49 at 156.

[112] *Id.*

[113] For a deep evaluation of the issue *see* THOMAS SUDDENDORF, THE GAP: THE SCIENCE OF WHAT SEPARATES US FROM OTHER ANIMALS (2013).

[114] As described in Carruthers, *supra* note 49 at 156.

[115] That might also require a 10th module: "Normative capacities, containing components specialized for learning the social norms that are operative in one's society, for generating intrinsic (non-instrumental) motivation to comply with those norms, and for generating a range of emotions in case of norm non-compliance (guilt, in one's own case;

4. THE LIMITATIONS FOR THE AUTONOMOUS IMPLEMENTATION OF DISCRIMINATION AND PROPORTIONALITY ASSESSMENT WITHIN ROBOTS: ESCAPE STRATEGIES

An autonomous weapon system complying with LOAC on the battlefield must match the relevant human abilities in relation to the cognitive modules described above. It seems that the implementation of such performances (for discrimination or proportionality evaluations) in autonomous weapon systems is unattainable by actual or by any future AI technology. Therefore, for the foreseeable future, the AWS might comply with LOAC only in very special and narrow circumstances, to be explored bellow.

a) Discrimination

We have seen that, according to LOAC rules, military personnel must distinguish a combatant from a non-combatant, even when, for example, insurgents pretend to be civilians. The military must also need to discriminate between active combatants and wounded ones who are unable to fight or who have surrendered.

In the second part of the paper we discovered that the visual (perceptual) intelligence in state-of-the-art autonomous systems is not capable to distinguish the obstacles even at levels accessible for insects. More importantly, we believe that the artificial emulation of the highest level of human visual awareness (which uses cognitive modules like folk-physics or folk-psychology in understanding complex interactions of the physical and social world), is out of reach, even for AI technologies of the distant future. This problem might be overcome by lowering the threshold of discrimination by "simplifying" or "flattening" the complexity of a battlefield environment.

An early solution has been proposed by John Canning, an engineer at the Naval Surface Warfare Center, who suggested[116] that unmanned systems should target only the enemy weapons and not the enemies themselves. According to him, one would have only the combat of machines (lethal robots) against other machines (robotic or not).

The proposal might work well for weapons such as tanks and other vehicles that are operated only by human combatants. In other cases (individual weapons for example), the proposal might not work since, given the current or foreseeable limitations in AI, robots cannot reliably target just the weapons and not persons, or cannot unfailingly differen-

---

anger and desires to punish, in the case of other people)[.]" *Id.* at 155.

[116] John Canning, A Concept of Operations for Armed Autonomous Systems at the Naval Surface Warfare Center (2006), *available at* http://www.dtic.mil/ndia/2006disruptive_tech/canning.pdf

tiate weapons from non-weapons.[117]

A more radical solution is to eliminate the need for discrimination altogether, by acting autonomously with lethal force, only when civilians are absent of the battlefields. Combat robots might simply operate only in regions of heavy fighting, crowded with valid targets[118] called "kill boxes" or "engagement regions," where the LOAC/ROE requirements are lowered since non-combatants can be realistically presumed to have fled, thus avoiding the issue of discrimination among targets.

b) Proportionality

The proportionality requirements impose an even higher threshold to AWS. In order to become truly autonomous, a weapons system must emulate somewhat of an "imagination" (a representation faculty) that is clearly out of reach of any AI technology. However, as for discrimination, one can remove the proportionality requirement altogether, by acting with autonomous robots only against machines (within the limited circumstances analyzed above) or in the "kill boxes."

GENERAL CONCLUSION

Based on developments from the second part of the paper, one may conclude that during combat operations the autonomous robots would face primarily two challenges: navigation of terrain with the addition of ISR tasks. Cognitively speaking, terrain navigation and obstacle avoidance require pattern recognition and problem solving skills to be implemented autonomously in robots. And even stronger pattern recognition abilities are required during ISR missions.

While autonomous navigation and certain ISR tasks seem to be in the reach of actual UMVs or future UAVs, the autonomous navigation of UGVs seems unfeasible. This was the result of the inverse correlation that we discovered, between attainable autonomy for such tasks and the environment complexity in which robots are deployed. We have also determined that the main technological bottleneck to actual autonomous robots is related to their poor "perceptive and/or visual intelligence."

Based on this threshold, we assessed the technical perspectives for autonomous implementation of lethal missions of military robots. We have concluded that full autonomy of lethal missions is, by now, within the reach of UMVs such as ACTUV (the hunter of diesel submarines).

---

[117] Discussion reproduced in Lin, et. al., *supra* note 110 at 77.

[118] *Id*, at 77.

In the near future we might also see the appearance of UAVs/UCAVs with self-defense abilities or with air-to-air and certain air-to-ground attack abilities.

In the thirds part of the paper we analyzed another challenge, this time only in relation to the autonomous implementation of lethal functions in robots: their "obligation" to fight in compliance with the rules of LOAC/ROE. In order to be legally deployed, an autonomous lethal robot must act on the battlefield like an autonomous moral agent and mimic the high cognitive and moral evaluations of humans.[119]

We determined that the autonomous implementation of such human-mimetic moral abilities is out of reach to any current or far-future AI technologies. This very high threshold will limit, for a long time, the general lawful deployment of AWS, or will restrict their autonomous lethal use.

However, the environmental simplicity might play a role once again. It might allow a legal deployment of AWS with (almost) no discrimination or proportionality assessment abilities, in circumstances where civilians are absent from the battlefield—within the "kill zones," in machine counter machines combats, etc. Such particular simple battlefield environments might be found mostly in sea/undersea combat or air combat.

In this respect, an UMV like ACTUV, provided with antisubmarine attack capabilities (as a submarine destroyer), might be the first fully operational AWS, complying, by default, with rules of LOAC/ ROE.

Other good candidates for implementing full lethal autonomy in robots, without infringing, by default, the LOAC/ ROE rules, are the UCAVs in air-to-air or (specific) air-to-ground combat missions.

We have seen that all other types of autonomous robots, such as UGV, will have very poor navigational abilities. The implementation of ISR or lethal mission in such systems seems to be a technological barrier. Therefore, the problem of such systems complying with LOAC/ROE rules will be, for long time, a purely philosophical question. That will apply even more to UGVs having the shape of android robot-soldiers. The specter of such Terminators coming into being as fighters will haunt, maybe forever, only the nightmares of science fiction fans.

---

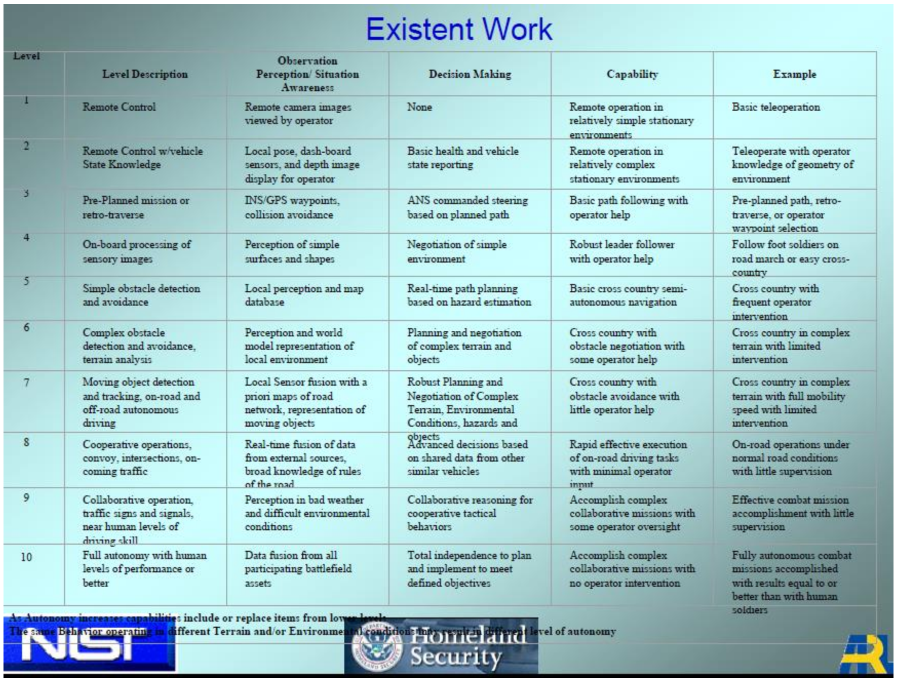
[119] TASK FORCE REPORT, *supra* note 6 at 16.

FIGURES

FIGURE 1: AUTONOMY SPECTRUM AND THE OODA LOOP

Level	Observe	Orient	Decide	Act
0	Flight control sensing & on-board camera.	Telemetered data; remote pilot commands.	None. Off-board pilot.	Control by remote pilot.
5	Local sensors to detect external targets, fused with off-board data.	Group action diagnosis and resource management.	On-board trajectory planning; optimize for current & predicted conditions; collision avoidance.	Group accomplishment of tactical plan as externally assigned; air collision avoidance.
10	Cognizant of all within the battlespace.	Coordinates as necessary.	Capable of total independence.	Requires little guidance of any sort.

FIGURE 2. AUTONOMY FRAME OF THE ARMY SCIENCE BOARD STUDY



Existent Work

Level	Level Description	Observation Perception/ Situation Awareness	Decision Making	Capability	Example
1	Remote Control	Remote camera images viewed by operator	None	Remote operation in relatively simple stationary environments	Basic teleoperation
2	Remote Control w/vehicle State Knowledge	Local pose, dash-board sensors, and depth image display for operator	Basic health and vehicle state reporting	Remote operation in relatively complex stationary environments	Teleoperate with operator knowledge of geometry of environment
3	Pre-Planned mission or retro-traverse	INS/GPS waypoints, collision avoidance	ANS commanded steering based on planned path	Basic path following with operator help	Pre-planned path, retro-traverse, or operator waypoint selection
4	On-board processing of sensory images	Perception of simple surfaces and shapes	Negotiation of simple environment	Robust leader follower with operator help	Follow foot soldiers on road march or easy cross-country
5	Simple obstacle detection and avoidance	Local perception and map database	Real-time path planning based on hazard estimation	Basic cross country semi-autonomous navigation	Cross country with frequent operator intervention
6	Complex obstacle detection and avoidance, terrain analysis	Perception and world model representation of local environment	Planning and negotiation of complex terrain and objects	Cross country with obstacle negotiation with some operator help	Cross country in complex terrain with limited intervention
7	Moving object detection and tracking, on-road and off-road autonomous driving	Local Sensor fusion with a priori maps of road network, representation of moving objects	Robust Planning and Negotiation of Complex Terrain, Environmental Conditions, hazards and objects	Cross country with obstacle avoidance with little operator help	Cross country in complex terrain with full mobility speed with limited intervention
8	Cooperative operations, convoy, intersections, on-coming traffic	Real-time fusion of data from external sources, broad knowledge of rules of the road	Advanced decisions based on shared data from other similar vehicles	Rapid effective execution of on-road driving tasks with minimal operator input	On-road operations under normal road conditions with little supervision
9	Collaborative operation, traffic signs and signals, near human levels of driving skill	Perception in bad weather and difficult environmental conditions	Collaborative reasoning for cooperative tactical behaviors	Accomplish complex collaborative missions with some operator oversight	Effective combat mission accomplishment with little supervision
10	Full autonomy with human levels of performance or better	Data fusion from all participating battlefield assets	Total independence to plan and implement to meet defined objectives	Accomplish complex collaborative missions with no operator intervention	Fully autonomous combat missions accomplished with results equal to or better than with human soldiers

As Autonomy increases capabilities include or replace items from lower levels
The same Behavior operating in different Terrain and/or Environmental conditions may result in different level of autonomy

FIGURE 3. AUTONOMY LEVELS FOR UNMANNED SYSTEMS (ALFUS) MODEL OF AUTONOMY