Working Paper
Research Division
International Security
Stiftung Wissenschaft und Politik
German Institute for
International and Security Affairs

*Marcel Dickow, Anja Dahlmann, Christian Alwardt,
Frank Sauer, Niklas Schörnig*

First Steps towards a Multidimensional Autonomy Risk Assessment (MARA) in Weapons Systems

FG Sicherheitspolitik WP No 05
December 2015
Berlin

SWP
Stiftung Wissenschaft und Politik
German Institute
for International and
Security Affairs

Ludwigkirchplatz 3- 4
10719 Berlin
Phone +49 30 880 07-0
Fax +49 30 880 07-100
www.swp-berlin.org
swp@swp-berlin.org

Dr. Marcel Dickow is the head of the International Security Division of the Stiftung Wissenschaft und Politik – German Institute for International and Security Affairs (SWP).
Anja Dahlmann is a Research Assistant at SWP.
Dr. Christian Alwardt is a Researcher at the Institute for Peace Research and Security Policy Hamburg (IFSH).
Dr. Frank Sauer is a Senior Research Fellow and Lecturer at Bundeswehr University Munich.
Dr. Niklas Schörnig is a Project Director and Senior Researcher at the Peace Research Institute Frankfurt (HSFK).

Table of Contents

The research for this report was financially supported by the Federal Foreign Office. The opinions expressed are those of the authors and do not represent the Foreign Office’s official position.
The paper also has been published as a PRIF Working Paper (HSFK) and as an IFAR Working Paper (IFSH).

Figures

Tables

Abbreviations

CARACaS	Control Architecture for Robotic Agent Command and Sensing
CCW	UN Convention on Certain Conventional Weapons
CIWS	close-in weapons system
LAWS	lethal autonomous weapons systems
MARA	multidimensional autonomy risk assessment
R&D	research and defense
UAV	unmanned aerial vehicle
UCAV	unmanned combat aerial vehicle
UCLASS	Unmanned Carrier-Launched Airborne Surveillance and Strike (system)
UGV	unmanned ground vehicle
UN	United Nations
USV	unmanned surface vehicle

Executive Summary

With recent progress in the fields of robotics and artificial intelligence, so called “lethal autonomous weapons systems” (LAWS), defined by the US Department of Defense as weapons systems capable of selecting and engaging targets without further intervention by a human operator, have become the subject of international debate.

Proponents of LAWS ascribe numerous benefits to them, such as their superior military capabilities and cost-cutting potential as well as the hope for rendering warfare more humane and less atrocious. Critics point to the legal and ethical responsibility gaps created by handing over the kill decision to machines and worry about the proliferation-prone weapon systems’ impact on international security and stability.

Against this background, LAWS have made it onto the agenda of United Nations (UN) arms control diplomacy, particularly with regard to the UN Convention on Certain Conventional Weapons (CCW) in Geneva. No CCW member state has so far stated that it is actively seeking LAWS, but there is disagreement about the need for further action within the CCW.

While some States Parties favor a “wait and see”-approach, others suggest a moratorium, while yet others demand the drawing of a line and to tightly regulate or even completely ban the development and deployment of LAWS – an action that civil society favors as well.

However, there is still a considerable amount of conjecture on all sides of the debate; and it remains unclear what exactly would have to be regulated/banned if the CCW were to take further action. All parties involved would benefit from a more reliable and empirically substantiated base for discussion.

The instrument for a multidimensional autonomy risk assessment (MARA) presented in this study can help remedy this situation and move the debate about autonomy in weapons systems forward in a sober and productive fashion. It is the most sophisticated, rigorous and readily applicable framework for this task to date and promises to be helpful for informing decision-makers and systematizing the issue.

The formula-based instrument MARA allows for generating comprehensive and transparent quantitative descriptions of weapons systems, both deployed and currently under development. MARA is based on fifteen weapons characteristics, so-called vectors, organized in five groups: physical characteristics (V_P), armament characteristics (V_A), human relevance (V_H), information processing/situational awareness (V_{IP}), and exposition (V_E). After scoring all vectors of a given weapons system, the MARA-formula generates the system’s overall MARA-score that can be normalized to the percentage of the maximum value, the MARA%-score. By scoring and ranking a wide variety of weapons systems, MARA can generate a comprehensive, comparative overview against the background of which informed deliberations about a quantitative threshold can subsequently take place. This threshold, a specific MARA%-score, would have to be defined politically as the acceptable maximum of combined autonomy and military capabilities in weapons systems – or, in short, where to “draw the line” for autonomy in weapons systems.

Eleven weapons systems were assessed and compared in this study. They do not add up to a data set robust enough for drawing definitive conclusions. However, we tentatively do suggest a 50%-threshold as a starting point for further discussions. In other words, we argue that any system with a MARA-score of 50% or more at least requires a closer look regarding the prudence of possible regulation.

Only three systems assessed in this study reach a MARA%-score of 50 or more: our extrapolation of the “UCLASS” program, which would be a weaponized, stealthy, autonomous aerial combat vehicle; and our hypothetically derived “Flying Robot Insect” (or “killer bug”), both as a single system and as a swarm. None of these are fielded yet. This is in line with the common understanding that

LAWS as such are not yet in existence, and it underlines the plausibility of using the 50%-threshold as a starting point. Continuing the process from there, MARA can assist policy-makers in coming to an informed decision on the possible establishment of a politically defined maximum of autonomy in weapons systems.

Introduction

The purpose of – and the motivation behind – this study is to move the debate on autonomy in weapons systems ahead by introducing some more conceptual clarity and definitional rigor. To that end, we offer a new instrument for conducting a multidimensional autonomy risk assessment (MARA) in weapons systems. By quantifying and computing key descriptive characteristics ("vectors") of systems to gauge their autonomous and military capabilities, the instrument can be used to generate a comprehensive overview over weapons systems deployed currently and in the near future. This way, it can assist policy-makers in coming to an informed decision on the possible establishment of a politically defined maximum of autonomy in weapons systems.

Our approach considers assistive technologies and implies an overall gradual notion of autonomy – rendering an "either-or-differentiation" between "automatic" and "autonomous" obsolete. While we are certainly not the first researchers to attempt a systematic autonomy assessment in weapons systems (see e.g. Williams/Scharre 2015; Scharre/Horowitz 2015), our approach represents the most sophisticated, transparent and readily applicable to date.

To begin a discussion of autonomy in weapons systems, it is worth pointing out that immobile weapons systems capable of tracking (and even engaging) targets independent from human input, such as Phalanx or PATRIOT for C-RAM or terminal defense purposes, have been in use for decades in militaries around the globe. Also, a growing capacity for independent "decision-making" in mobile, offensive weapons has been under discussion in expert circles since at least the 1980s, following the introduction of modern cruise missiles.

However, more recently – in light of rapid progress in the fields of robotics as well as artificial intelligence and against the background of the ever growing influx of commercially developed hard- and software into weapons systems – this discussion has gained an entirely new level of immediacy and relevance (ICRAC 2009; FLI 2015).

Known as "lethal autonomous weapons systems" (LAWS) in United Nations (UN) parlance – and dubbed "killer robots" by critics (stopkillerrobots.org 2015) –, a new generation of hi-tech weapons systems is expected (feared) to arrive on battlefields in the near future. These LAWS, "once activated, can select and engage targets without further intervention by a human operator", according to the US Department of Defense (US Department of Defense 2012: 13). They are expected to be mobile and capable of roaming freely in open, dynamic, unstructured and uncooperative environments over extended periods of time, making decisions – including the decision to engage targets and to "kill" – without human supervision and only via onboard decision-making algorithms.

It is mainly for applications underwater and in the air – in less complex and comparably open, accessible environments – that the trend toward more such autonomy is currently most apparent. But regardless of the respective domain of use, proponents expect a multitude of benefits from increased autonomy in weapons systems. For sake of brevity, we will only mention three:

(1) Every control and communications link is vulnerable to disruption or capture and may also reveal a system's location. Also, it inevitably creates a delay between the issuing of a command by the responsible human operator and the execution of the command by the unmanned system. The expected benefit of LAWS operating completely independent from human input once activated is that there would be limited need for such a link (arguably even no need at all).

(2) Proponents expect LAWS to deliver superior performance regarding, for instance, endurance and operational range as well as considerable cost-cutting potential, especially due to the reduced need for personnel.

(3) Lastly, as LAWS are immune to fear, stress and overreactions, proponents believe that they offer the prospect of a more humane way of conducting warfare. After all, not only are machines devoid of negative

human emotions; they also lack a self-preservation instinct, so LAWS not only remove the human from the battlefield, thus keeping friendly troops out of harm's way, they could well delay returning fire in extreme cases, "sacrificing" themselves rather than risking the life or well-being of innocents. This, it is argued, could prevent some of the atrocities of war (see e.g. Arkin 2010).

At the same time, there is widespread concern within the pertinent scientific communities as well as the international community at large that these systems could pose a set of new – potentially deeply troubling – political, legal and ethical perils. We will again limit ourselves to only three such objection clusters currently raised by critics:

(1) Roboticists and international lawyers doubt that, at least for the near future, machines can be programmed to abide by international law in the notoriously grey area of decision-making on the battlefield. In close connection to that, it is worth noting that the entire body of international law is based on the premise of human agency; it is therefore unclear who would be legally accountable if human beings – particularly civilians – were unlawfully injured or killed by LAWS (Sharkey/Suchman 2013; McFarland/McCormack 2014; HRW 2015). Also, the Martens' Clause, part of customary international law, holds that in cases not (yet) covered in the regulations adopted in international law, the principles of the laws of humanity and the dictates of the public conscience apply. And the general public appears to in fact harbor serious concerns about LAWS. The findings of a representative survey in the United States shows that a majority (55%) of US citizens is opposed to autonomous weapons on humanitarian grounds, with 40% even "strongly opposed" (Carpenter 2013).

(2) An ethical question raised by LAWS presents another objection. It has been argued that handing the power to decide on the use of force against human beings over to anonymous algorithms is a violation of the basic principles of humanity and human dignity (Asaro 2012).[1]

(3) Finally, the technology driving the trend towards more autonomy in weapons systems is dual-use in nature, can easily be copied (especially regarding pieces of software or algorithms), and is thus prone to quickly proliferate to additional state and non-state actors. Critics thus expect the fielding of LAWS to yield detrimental effects to international security, including increased regional and global instability due to a lowered conflict threshold, arms races, unforeseeable interactions of autonomous systems ("accidental war") as well as an ever increasing speed of battle rendering the retention of human control impossible.

Against the background of these recent developments and controversies, the issue of LAWS has started rising on the agenda of international arms control diplomacy over the last few years, for instance regarding United Nations fora such as the Human Rights Council and the UN General Assembly 1st Committee. Most importantly, however, on 15 November 2013, States Parties to the Convention on Certain Conventional Weapons (CCW) at the UN in Geneva decided to hold a four-day informal "Meeting of Experts" to discuss questions related to LAWS on May 13 to 16 2014. The States Parties discussed a report at the conference in November 2014 and held another informal CCW "Meeting of Experts" from 13 to 17 in April 2015.

A possible outcome of this particular process at the UN CCW could be a decision on a Governmental Panel of Experts for 2016 or even a negotiation mandate that could, further down the road, lead to a CCW Protocol VI banning the development and deployment of LAWS – an outcome that civil society, represented most vocally by the international Campaign to Stop Killer Robots is currently pushing for.

No CCW member state has so far stated that it is actively seeking LAWS. Some have voiced their support for a ban, others (Israel,

---

[1] For a critical reflection on all arguments pro and con see (Sauer 2014a; 2014b; Sauer/Schörnig 2012).

USA) announced to first invest in additional R&D before making a decision. Others yet (France, United Kingdom) have explicitly stated that they will not pursue the development of LAWS further but that they are also not supporting a ban at this point in time. Germany has stated in Geneva that it "will not accept that the decision to use force, in particular the decision over life and death, is taken solely by an autonomous system" and that "the red line leading to weapons systems taking autonomous decisions over life and death without any possibility for a human intervention in the selection and engagement of targets should not be crossed" (Germany 2015). In short, there is some hesitation at the CCW. But as it stands, there is also considerable momentum and the emergence of new regulation (some sort of binding legal instrument, a moratorium, maybe even a ban of LAWS via a CCW Protocol VI) is not entirely ruled out.

However, it remains unclear what exactly would have to be regulated/banned. One very prominent suggestion on how to draw the line between a system that is autonomous (and thus subject to regulation) and one that is not, is the question, if "meaningful human control" (Article 36 2014) is assured to a satisfying degree while the system is in operation. But the debate still lacks a clear idea of what constitutes this meaningful human control. Therefore, the fundamental question remains: What is autonomy in a weapons system, and when is a particular lethal autonomous weapons system "too autonomous"?

The approach presented in this study aims to help with politically defining the threshold above which any system with a specific degree of autonomy has to be considered "too risky to field". We follow the idea that the regulation of autonomous systems should not only take into account the level of remaining human control, but additionally consider other technical characteristics such as its damage output and the situational awareness of the system's artificial intelligence. This holistic view provides a clearer picture of the risk any specific autonomous system represents – with "risk" here primarily defined as a combination of the degree of remaining human control (as operationalized via specific vectors described below) and its overall military potential and thus potential impact on international security and stability at large.

In the upcoming sections we will lay out what can and cannot be expected from our instrument, describe the formula we developed for generating MARA-scores for weapons systems via specific "vectors", and show exemplarily how this approach can be applied to assess both current and future (robotic) weapons systems.

Disclaimer

Our formula-based instrument helps assessing an autonomous weapons system through an operationalization of the concept of autonomy and the addition of supplementary technical characteristics. By scoring a wide variety of weapons systems and comparing their respective MARA-scores, a quantitative threshold can subsequently be defined in a deliberative process. That would be a specific MARA-score, politically defined as the acceptable maximum of combined autonomy and military capabilities – or, in other words, "where to draw the line" regarding the limit beyond which a specific weapons system is considered too autonomous, too powerful, and thus too risky to be used.

It is important to note that while this study is fairly technical in presenting numbers, a formula and also a suggestion on where this line could quite plausibly be drawn, this "precision" should not be misunderstood. First, the values attributed to the weapons systems we scored for our case studies are often estimations (based on thorough research, multiple reviews and intense discussions in a circle of experts, but estimations nevertheless), sometimes derived from comparisons with similar systems. The reason for this is simply that, depending on the system, some or even most parameters were not known to us (because they are classified or at least not in the public domain). However, especially with regard to scoring autonomy-related functions, we aimed for mitigating the resulting fuzziness by assessing the capability of the system on an abstract level, thus reducing the need for detailed information on a system's software, computational power or certain aspects of artificial intelligence. But even given a potentially more precise scoring derived from better data, the MARA-scores generated by us are not absolute and leave some room for interpretation. Lastly, and most importantly, the eventual definition of the threshold is of course open to political discussion and can only be defined via a broad and systematic comparison of different systems and subsequent debate – an endeavor to which our study aims to merely contribute the first step.

In short, we are fully aware that the scores in our case studies are disputable. Consequently, we understand our instrument at this point to primarily offer a framework for peer scholars and practitioners to enter their own data and test and compare existing, evolving or even hypothetical systems. After all, a key feature of the instrument is that it allows for systematically and transparently comparing different weapons systems as well as generations of similar weapons technologies and swarms in contrast to single systems, this way giving a standardized description of technological trends and effects.[2]

To conclude, numbers – even ones more adequate than those we can provide here – are no replacement for political will; and they do not remove the need for negotiation. That said, applying our instrument for quantification will be immensely helpful for moving the current political discussion ahead by informing decision-makers and systematizing the debate.

Vectors and Scales

We base our multidimensional definition of autonomy on fourteen so-called vectors organized in five groups: physical characteristics (V_P), armament characteristics (V_A), human relevance (V_H), information processing/situational awareness (V_{IP}), and exposition (V_E).

The first three vector groups represent physical design elements and characteristics of the platform and its weapon payload as well as the design of the human-machine-interface, while V_{IP} and V_E represent the system's capability for algorithmic decision-making and its communication interfaces.

To keep the categorization as simple as possible, we decided to use a cardinal scale with a range between 1 and 10 for the values of each vector (please find the scales of each vector in detail in the Annex). We are aware that this decision already predetermines normalization and weighs single vectors and thus vector groups over others. To counter additional hidden scaling effects induced by this approach, the weighting of vectors and vector groups will be done identifiably, where necessary, in the final formula. More on that further below.

V_P – Physical Characteristics

The vector group "**physical characteristics**" (V_P) describes the effective range, period of application and speed of the weapons system.

The vector "**effective range**" (V_{P1}) describes the physical range of the weapons platform, while the weapon range is covered by its own vector. The scale starts with immobile systems (1) and ends with systems capable of reaching outer space (10).

Vector V_{P2}, the "**period of application**" scores the time a system can stay active without maintenance. It illustrates for how long the system can – potentially – operate without any human interference, which becomes especially important if it does not fulfil its tasks as expected or the way it is supposed to. The scale ranges from less than one minute (1) to more than ten years (10).

---

[2] In addition to the MARA formula presented below, it is conceivable that sub-formulas could be developed to, for example, specifically highlight the effect of longer ranges, longer periods of application or higher weapon payloads. Currently, we do not include such additional options in our study; but it is possible to add them at a later point in time.

Vector V_{P3} scores the maximum “**travelling speed**” of a system in kilometers per hour. Immobile systems are rated with the lowest score, a one (not zero), the maximum value is the parabolic speed, meaning the speed an object needs to leave the Earth’s orbit.

V_A – Armament Characteristics

The vector group “**armament characteristics**” (V_A) scores the capabilities of the system regarding its weapon payload. With regard to the systems discussed in the case studies below, we acted on the assumption that the respective system carries a standard payload; but the assessment can be adapted to alternative payloads used for specific missions, of course.

The “**weapon range**” (V_{A1}) describes the effective range of the weapon attached to the system. It is of relevance in connection to the effective range (V_{P1}) of the system itself.

Vector V_{A2}, the “**kill radius**”, describes the lethal radius of the weapons attached to the system – or, in cases in which the system is not a weapons platform but a “suicide” system such as Harpy, the system itself. It ranges from an impact limited to one person or a small object to a radius of over 100 km. Of course, this vector is only a rough approximation of the actual weapon’s effects. Nevertheless, it illustrates the damage potential and therefore the risk emanating from targeting errors.

Lastly, vector V_{A3} (“**kill cycles**”) scores the number of weapon uses a system is capable of before having to be rearmed.

V_H – Human Relevance

The vector group “**human relevance**” (V_H) scores the level of human involvement during both the development phase as well as the operational phase of a system. The group currently consists of only one vector, but other elements could complement it in the future.[3]

The “**actual human control**” (V_{H1}) is a fine-tuned description of the familiar “man-in”, “-on-“ or “-out-of-the-loop” concept, that is, the level of human involvement in a system’s command and control process. The lowest score is assigned to an immediately and fully piloted system; but considering the numerous assisting technologies included in most modern weapons systems, most of them will at least reach rank two. The “man-in-the-loop” and in full control is merely and ideal-type concept against this background. The subsequent scores on the scale represent the gradual detachment from human input and decision-making, in other words, the increased outsourcing of tasks to the machine, leaving, in the later stages, only a limited veto power to a human operator, culminating in a system operating with no possible human interference at all.[4]

---

[3] One of these vectors could be “debugging”, which would measure the effort that was put into testing and quality control before the actual fielding of the system. This vector would serve as a proxy for gauging the system’s potential for malfunctions. Again, the value of the vector (in man hours) is inversely proportional to the quality of the debugging, i.e. a very thorough debugging will result in a low score while a superficial one or none at all will score high, the rationale being that good and thorough debugging during development lowers the potential for fielding a system with a high probability of malfunctioning. However, as a more complex system by definition needs a more thorough debugging compared to a simple one, the vector alone is not sufficient but would have to be considered in relation to the system’s overall complexity, scored by the vector group V_{IP}.

1:	1,000,000 h	6:	10,000 h
2:	500,000 h	7:	1,000 h
3:	250,000 h	8:	100 h
4:	100,000 h	9:	10 h
5:	50,000 h	10:	No systematical debugging

[4] The score for human control decreases with increasing values of the system (10 = no human control), which might be counterintuitive. However, this scale follows from our assumption that the lack of human influence means more autonomy and more autonomy can potentially render a system a more risky one to field.

V_{IP} – Information Processing/ Situational Awareness

The vector group “**information processing/situational awareness**” (V_{IP}) consists of four vectors describing a system’s ability to gather data on its environment as well as its ability to interact with its environment and its ability to process a certain level of abstraction in the communication about mission goals with its human operators. The aim of this vector group is to determine the overall quality or sophistication of the information processing chain, from gathering sensor data to the computed outcome. This outcome represents a system’s decision making capability on different layers of abstraction. The quality of hard- and software, the processing power and the effectiveness of a system’s decision-making algorithms determine the results. However, these characteristics cannot be normalized and operationalized due to incomparable hard- and software architectures, lack of open source data and the unquantifiable nature of artificial intelligence. We therefore reduced the scope of vectors in this particular group to two sensor-related (V_{IP1} and V_{IP2}) and two processing-related vectors (V_{IP3} and V_{IP4}).

Vector V_{IP1} is called “**coverage**” and is scaled in decimal fractions of the surface of a complete sphere ($A = 4\pi r^2$). The value 10 represents complete, 3-dimensional sensor coverage of the imagined sphere surface, while 1 represents one-tenth of that surface. Typical forward-looking sensors such as optical and infrared cameras range from 1 to 5, depending on the applied optical lenses with a value of 5 representing a perfect fish-eye view. The coverage is scored cumulatively across all available sensors regardless of their quality or resolution. A spherical camera array could achieve a value of 10, if designed appropriate. A more detailed vector description should be considered when specific data of given sensors is available. However, having in mind that modern platforms use a wide range of sensors and that specifications of military-grade sensors tend to be classified, a much more simplified methodology was deemed necessary. Also, the ability of a weapons system to “recognize” and “understand” its environment depends to a greater extent on its capability for data-processing and -fusion across different types of sensors in real-time rather than the an overall greater amount or data.

Vector V_{IP2} is called “**multispectrality**” and scores the number of dissimilar sensor types relevant to the fulfillment of a system’s mission (with a maximum of ten sensors). This excludes sensors used solely for navigation as these are covered by V_{IP3}. Again, we assumed that sensor data-fusion and -processing can theoretically lead to greater performance. But more sensor data from different sensor types does not automatically add up to more information and thus better “knowledge” or “understanding”. Therefore, this vector refers to the availability of different types of sensor data only.

The third vector in this group is the “**ability to interact with the environment**” (V_{IP3}) with a focus solely on navigation. Interaction with the environment is a complex task depending on multiple functionalities of the system, i.e. pattern recognition, navigation, subject and object classification, etc. Instead of adding up single capabilities, which differ across systems, we decided to describe an overall functionality with direct reference to the action in a given environment. While, for example, a very simple system is unable to maneuver by itself or even move at all (vector value = 1), a more advanced system is able to move and detect and avoid obstacles (vector value = 4). Even more sophisticated systems can follow rules and abstract instructions (e.g. the Google Car in civilian traffic; vector value = 6) or learn and teach themselves new rules in a dynamic fashion (vector value = 9).

Vector V_{IP4} scores the “**level of abstraction regarding mission goals**” in analogy to human communication when tasking missions. For example, a human weapons system operator understands the order to “identify and destroy mission-relevant targets” because he or she can identify, classify and prioritize objects and subjects against the background of knowledge and experiences regarding the environment and the mission at hand. The capability to perform “reasoning” at such a high level of abstraction in this example would generate high vector values. The greater the level of detail in which a human operator needs to predefine a target for an autonomous system on the other hand, the less the system’s capability of abstraction and thus the lower the value of the vector. The vector therefore describes the level of abstraction possible during a human-machine communication process. It offers a soft- and hardware-independent way to gauge a weapon system’s level of ”artificial intelligence” (with “AI” really becoming relevant at only higher vector values). The attribution of a score depends on assessing the ability of a system to classify a specific target (the tactical and strategic relevance of an identified object or subject to the mission). Very crude systems will either engage indiscriminately (value = 1) or attack only targets clearly marked by humans beforehand, for instance with a laser (value = 2), while more complex systems differentiate between classes of targets (i.e. blue vs. red force, simple categories of ships, planes, tanks etc.; value 4/5/6). Even more sophisticated systems can differentiate targets according to their behavior in real-time (8) and prioritize on a tactical (9) or even strategic (10) level.

V_E – Exposition

The vector group “**exposition**” (V_E) is a set of generic vectors which describes the vulnerability of the system against external hacking and manipulation. It does not cover vulnerability against physical attacks. Instead, the idea is that the more vulnerable a system, the greater a threat it poses to the user, either because of potential direct damage when the system is turned against one’s own troops or indirect or political damage when the system is spoofed into attacking, for example, noncombatants. The risk of take over for a system through an attack on its communication links depends on at least two factors: first, the degree of exposition to the environment due to its physical communication interfaces, and second, the likelihood of systematic or stochastic flaws (bugs) in its software (firmware, operating system). The latter cannot be determined without an in-depth code review is thus not considered here (but the software error probability might at least be approximated via a vector V_{H2} “debugging”, as described in footnote 3). As it is difficult to operationalize a system’s exposition to manipulation based on its physical interfaces, we decided to focus on three generic vectors approximating physical conditions to interfere with communication links.

The “**interface range**” (V_{E1}) is the maximum distance from which a particular interface of a system can be attacked. It relates to the transmission range of its communication system. If a system has, for example, no transmitting interface and all communication with the system is wired, the value of this vector is 1. Near-field communication like Bluetooth would score higher, and radio- or satellite-communication even more so.

Vector V_{E2} is called “**directivity**”. It represents the directional characteristics of the transmitting and receiving communication link. Obviously, a point-to-point communication with laser light is less vulnerable to interference (in the sense of less likely to be compromised) than an omnidirectional radio link. The vector does not need the granularity of ten steps, leaving some in-between steps blank.

Vector V_{E3} assesses the quality of “en**cryption**” used when data is transmitted to and from the weapons system. The value of the vector is inversely proportional to the quality of the encryption, i.e. a very sophisticated encryption will get a low score while a bad one or none at all will score high, as good encryption lowers the overall exposition of the system. As of now, we deemed only three values of encryption quality necessary: The strategic encryption (value = 1) assumes state-of-the-art algorithms with proper implementation, perfect random number generators, highly secured key exchange mechanisms and session-based security features like “Perfect Forward Security”.

Secrecy, integrity and authenticity will be guaranteed for decades or centuries with today’s computing power. The tactical encryption (value = 5) favors simplicity and speed over security, thus satisfying only temporary security needs and waiving elaborate encryption mechanisms. The highest score implies no encryption at all.

As modern weapons systems tend to have multiple interfaces, each interface – based on the three vectors described – is scored, but only the interface with the highest score is included in the calculation to represent the minimum exposition of a system to external manipulation.

The MARA-Formula and a Possible Threshold

The MARA-Formula

The basic idea of the MARA-formula is to come up with one value for any particular weapons system summarizing the "risk" deriving from the individual system due to its level of autonomy and military capabilities. For the actual formula, we use a simple additive approach including all vectors described above. However, to give the crucial aspect of "system autonomy" more weight, we added the sum of the most important autonomy-related vectors (V_H: actual human control; V_{IP3}: ability to interact with the environment; V_{IP4}: mission tasking capability) with a weighting factor n. This way, the final MARA-formula looks like this:

MARA

$$= V_P+V_A+V_E + (1+n)^* V_H+V_{IP1}+V_{IP2}+(1+n)^* V_{IP3}+(1+n)^* V_{IP4} \qquad (1)$$

$$= \underbrace{V_P+V_A+V_H+V_{IP}+V_E}_{\text{sum of all vector groups}} + \underbrace{n^*(V_H+V_{IP3}+V_{IP4})}_{\text{additional weight of crucial aspects of system autonomy}} \qquad (2)$$

(with $n = 1, 2, 3, \ldots$)

Depending on n, the maximum MARA-score for a particular weapons system (MARAmax) would be 170 (for n = 1) or 200 (for n = 2) respectively. In order to compare different weightings, MARA-scores for individual weapons systems can be normalized to their percentage of the maximum value. A MARA%-score of 50 would therefore mean that the particular systems scored 90 or 105 in absolute terms. We indicate MARA% as a normalized value in contrast to the absolute MARA-score. Concerning all following considerations, we suggest the weighting factor n to be kept at 1, with all autonomy-relevant vectors having twice the weight in the end-result. However, in further studies the weighting factor n may be changed as a helpful way to rebalance the relative spacing of different systems depending on their level of autonomy.

Threshold

Having defined the MARA-formula and its potential range leads to the obvious question of when a system scores "too high", meaning when it can be considered too autonomous and powerful and thus too risky to field. In the introduction, we discussed that this is a political, rather than an analytical question and cannot be answered ex-ante.

But: To offer a guideline for interpretation and how to make use of MARA, we suggest a 50% threshold that is as simple as it is arbitrary and supposed to work as a starting point for further discussion. Any system with a MARA-score of at least 50% or more bears, in this line of thought, enough risk to warrant a closer look if regulation might be prudent.

As we will show in the section on results, only a few systems reach a MARA% of 50 or more, and none of these are fielded yet. This is in line with the common understanding that LAWS as such are not yet in existence, and it underlines the plausibility of using the 50%-threshold as a starting point for now. Nevertheless, we strongly suggest adding more systems where detailed data is available in order to come to a more comprehensive MARA-based review and possibly adjust the cut-off point. As it stands, our universe of eleven cases based only on publicly available data is not a data set robust enough for drawing definitive conclusions for the current debate.

Case Selection

Our case selection gives a first impression of the scope of assessments possible with our instrument. We strove for a wide range of different (autonomous) weapons systems. We chose well-known and established systems first and added a few evolving and hypothetical systems later on. Due to their prevalence, the focus lies on aerial vehicles; but we also included sea- and land- based systems.

To test the descriptive power of the instrument we also assessed as a synthetic testbed two systems that are outside the scope of the debate around autonomous weapons systems: the antipersonnel mine and the Eurofighter. Next to these two "outliers", we limited the case selection process to systems with enough open data available about their technical specification and grade of autonomous functions (although the latter information is usually not documented in detail). Lastly, two hypothetical case studies (UCLASS, flying robot insect) were conducted as plausible extrapolations from of already existing technologies, prototypes or ongoing research projects. Their features have been demonstrated generally, but they do not exist as weapons systems as of yet. We assumed their level of mission-relevant artificial intelligence according to the latest AI research results, plausible future projections and the requirements to be met by future weapons system according to the newest military roadmaps.

Harpy

Harpy is a weapons system developed by Israel Aerospace Industries in the 1990s, designed to detect and destroy radar systems by slamming itself into the emitter. In that sense it blurs the lines between a cruise missile and an unmanned aerial system (UAV); it is a "fire-and-forget" weapons system, which can be launched from a ground vehicle or a ship and loiters in the air until detecting radar emissions. Harpy has a range of 500 km, a maximum speed of 185 km/h and a standard armament of one 32 kg high-explosive warhead. Harpy is a good example for a far-reaching weapons system, with a certain degree of autonomy but at the same time very limited military capabilities. It is designed to be used in traditional symmetric warfare to counter aerial access-denial strategies of a given adversary and to engage against active radar-targeting weapons systems.

MQ-9 Reaper

The MQ-9 Reaper was developed by the US Company General Atomics, taking its first flight in 2001. Its predecessor-UAVs (such as the MQ-1 Predator) were originally designed for long-endurance, high-altitude surveillance tasks and were armed only later on. The main purpose of MQ-9 UCAVs is close air support and targeted killings in asymmetric warfare scenarios. It has a range of 3,065 km, a maximum speed of 482 km/h and can stay in up in the air for 30 hours. The platforms' standard weapons payload consists of four AGM-114 Hellfire and two AGM-176 Griffin, or GBU-38 Joint Direct Attack Munition. It can be outfitted with air-to-air missiles as well.

UCLASS

The Unmanned Carrier-Launched Airborne Surveillance and Strike (UCLASS) system is a program currently run by the US Navy. We use the abbreviation UCLASS as a synonym for a future unmanned, carrier-based fighter aircraft with advanced autonomous functions. It will fly much faster than existing surveillance UAVs and will be stealthy, weaponized and able to operate in contested airspace. Some of the capabilities of our assumed UCLASS system have already been tested with the Northrop Grumman X-47B technology demonstrator, including autonomous take-off and landing on a carrier ship, autonomous aerial refueling and stealth.

The extrapolated UCLASS will be able to autonomously identify and engage targets based on relatively abstract mission objectives. Remote piloting will no longer be needed and although communication links exist, they do not need to be used throughout the entire operation if a reduction of the system's electronic signature is required. Additionally, in our extrapolation of the X-47B and the

UCLASS program, the future weapons system relies on a variety of sensors including optical, infrared laser and radar sensors. It will be able to fly and navigate autonomously and will take tactical decisions by making use of sophisticated on-board situational awareness capabilities, i.e. it will prioritize target engagements or reassess target distribution. Due to higher cruising speeds and stealth, its weapons are carried in a weapons bay, thus limiting the available payload. In our definition, UCLASS has air-to-air combat (i.e. dogfighting) capabilities and will also conduct precision strikes on the ground. It therefore is multi-role capable.

The (Hypothetical) Flying Robot Insect

The hypothetical robot insect (or "killer bug") would be a very small, highly autonomous aerial vehicle mimicking a flying insect.5 It could work alone or in a swarm. The robot insect would enter secluded areas that soldiers or taller-sized robots would not be able to reach in order to surveil or kill human targets. It would poison its targets, so one bug could kill or stun only a few – probably just one – human being at a time.

To fulfill these tasks, it has to be hard to detect, requiring a very small, silent, probably fast machine. It also needs several sensors to navigate and detect, although the latter ones could also be distributed within a swarm. The robot insect would be able to carry small cameras and microphones or sensors to detect body heat or chemicals depending on the mission.

Since the robot insect would be able to work in isolated areas like deeply buried bunkers, it would be able to navigate swiftly through narrow terrain and possibly act within a swarm. Outside communication and control would be very limited: Once started and introduced to the mission parameters, the robot insect acts autonomously (for instance without having to rely on GPS signals), according to our assumptions. It might be able, however, to receive, within limited parameters, electro-magnetic (radio, light) or acoustic mission abortion signals. Within the swarm, the robot insect would make use of biomimetic techniques, such as chemicals to establish mesh-network coordination comparable to bees or ants. Its energy supply would be limited but might be complemented with energy-harvesting capabilities giving the insect a longer endurance.

Guardium

The Guardium is an unmanned ground vehicle (UGV) developed by G-NIUS, a joint venture by Israel Aerospace Industries and Elbit Systems. It entered operational service in the Israel Defense Forces in 2008. The vehicle can be remotely controlled or used in "autonomous mode", which in this case appears to primarily mean the ability to drive along preprogrammed routes. As we have no further details on the specific autonomous functions of GUARDIUM, we assumed only very limited capabilities. For our calculations, we assumed the system to be operated remotely and not able to engage targets without direct human control; in particular, being unable to identify, select or prioritize targets for attack. Its maximum speed is 80 km/h; it can drive for up to 103 consecutive hours and carry lethal and non-lethal weapons.

Phalanx (Close-In Weapons System, CIWS)

Phalanx (originally General Dynamics, now Raytheon) is a ship-based, tactical air defense system against incoming rockets, anti-ship missiles and artillery fire developed in the 1970s and further upgraded (actual Block 1B) since then. It features a 20 mm gun and a radar detection system (a surveillance and a tracking radar) for target identification, tracking and engagement. It is the last line of naval air defense, acting only on very short ranges. Phalanx cannot move laterally but is able to turn by 360° and elevate from -25° to +85°. Phalanx is designed to engage quickly and on its own once turned on and set to

---

[5] Our outline of this "killer bug" is mostly based on a tender by the Defense Advanced Research Projects Agency (DARPA), an agency of the US Department of Defense, calling for ideas regarding a "Fast Lightweight Autonomy Program" (Scola 2014; Tucker 2014).

automatic mode. It cannot identify friend or foe but classifies potential targets from basic patterns of their behavior including approaching velocities and course. Phalanx is the standard CIWS on US Navy ships and well established also with US allies.

CARACaS

CARACaS (Control Architecture for Robotic Agent Command and Sensing), developed by the US Office of Naval Research and demonstrated for the first time in 2014 (US Office of Naval Research 2014), is a module to turn "almost any boat" (Smalley 2014) into an unmanned surface vehicle (USV). In single mode, CARACaS allows to control a boat remotely. In swarm mode, the system coordinates client boats from a central operating point. CARACaS provides only command, control and sensing, it is not the actual weapons platform. For our case study we assumed the CARACaS system as being applied to a small, lightly armed platform (a rigid-hulled inflatable boat or a swarm of such boats respectively), i.e. vessels carrying a machine gun and travelling at high velocities. The aim of the swarm is to protect friendly ships and fleets from attack by adversary boats. According to project documents, CARACaS is able to coordinate the boat swarm autonomously for regular cruising; when attacked, it engages the adversary. As more detailed information about the grade of autonomous function is unavailable, we assumed some autonomous functioning in situational awareness and coordination of movements. At the same time, we assumed the autonomous attacking capabilities are very limited.

Antipersonnel Mine

We use the antipersonnel landmine as a synthetic test case for the validation of our instrument. We assume that the mine is completely disconnected from any human control once delivered to the theater and activated. The mine only has one sensor, sensing physical pressure from above. It has no analogue or digital data interface and no inbuilt deactivation mechanism.

The low MARA-score does not imply that the antipersonnel mine is harmless – it has been banned by international law for good reasons.

Eurofighter

The Eurofighter is our second synthetic test case. It is a fifth-generation, multi-role capable, manned fighter aircraft. We assumed standard weaponry in the ground strike configuration. The Eurofighter's "Attack and Identification System" fuses different sensor data, processes tactical information and is able to prioritize targets, threats and actions. The Eurofighter is a good outlier case because it features highly advanced assistance systems providing autonomous functions within a piloted weapons platform. We use the example to demonstrate that our instrument can assess system capabilities also with regard to immediately human-controlled platforms. But obviously the MARA calculation for the Eurofighter produces a synthetic (meaning non-significant) value as the platform is piloted and has an onboard human-machine-interface.

Results

MARA-Scores: Overview

Table 1 shows all scores attributed to the eleven weapons systems assessed for this study. It bears repeating that we do not consider these scores as definite results but hope for them to be improved upon or adjusted as soon as more and better data on the systems becomes available. For now, those are assumed values.

Applying the MARA-formula with a weighting factor of n = 1 (as discussed in the section on the MARA-formula) to the cases above results in one risk assessment score per system. For absolute MARA-scores see Figure 1 and for MARA%-scores see Figure 2 below. The dotted line indicates the 50%-threshold (see section on MARA-formula). The "dumb" landmine and the piloted Eurofighter are test cases, calculated to show that even with these artificial examples the overall relation of results is plausible.

System	V_{P1}	V_{P2}	V_{P3}	V_{A1}	V_{A2}	V_{A3}	V_{H1}	V_{IP1}	V_{IP2}	V_{IP3}	V_{IP4}	V_{E1}	V_{E2}	V_{E3}
Harpy	5	3	5	5	5	1	9	4	2	3	3	7	8	5
Reaper	7	4	6	4	6	3	2	5	4	3	2	10	4	5
UCLASS	8	4	7	5	6	3	8	9	6	7	5	10	8	1
Robot Insect	3	3	3	1	1	1	9	8	5	8	6	4	10	5
Robot Insect (50)	3	3	3	1	1	6	9	10	5	9	7	4	10	5
Guardium	5	4	4	3	3	5	3	4	3	4	3	6	8	5
Phalanx	3	8	1	3	1	2	9	4	1	1	3	1	1	10
CARACaS	4	4	4	3	2	1	4	4	4	4	3	5	8	5
CARACaS (10)	4	4	4	3	1	4	8	4	4	6	3	6	8	5
Antipersonnel Mine	1	10	1	1	2	1	10	1	1	1	1	0	0	0
Eurofighter	7	4	8	5	6	4	1	7	5	4	2	10	8	5

Table 1: Assumed vector values for the selected cases (weapons systems)

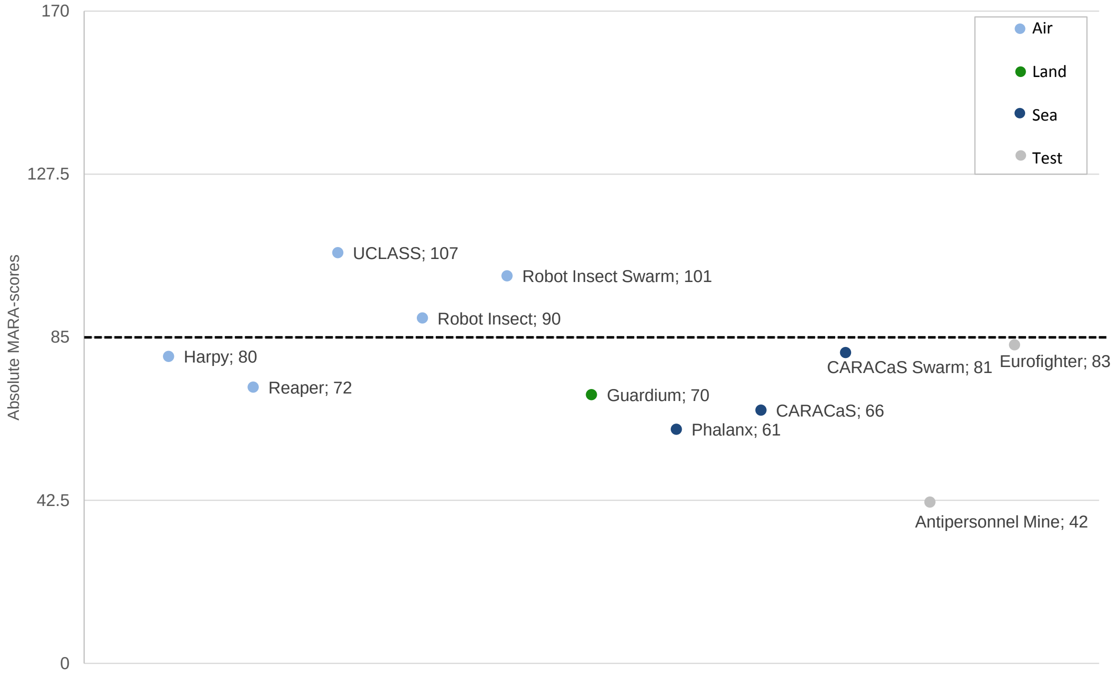

Figure 1: Absolute MARA-scores calculated for each of the eleven cases with a weighting factor of n = 1

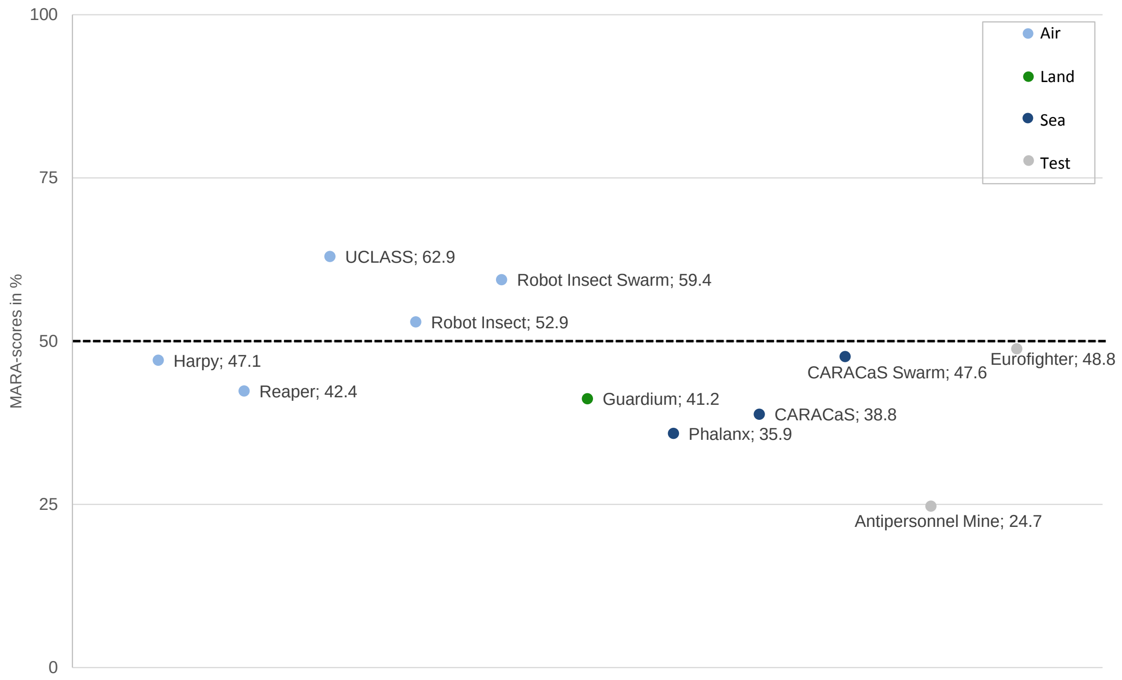

Figure 2: MARA-scores in percentage of maximum value (170) calculated for each of the eleven cases with a weighting factor of n = 1

System	MARA	MARA%	Ranking
UCLASS	107	**62.94**	1
Robot Insect (10)	101	**59.41**	2
Robot Insect	90	**52.94**	3
Eurofighter	83	48.82	4
CARACaS (50)	81	47.65	5
Harpy	80	47.06	6
Reaper	72	42.35	7
Guardium	70	41.18	8
CARACaS	66	38.82	9
Phalanx	61	35.88	10
Antipersonnel Mine	42	27.71	11

Table 2: Summary of the MARA results (n = 1)

Table 2 summarizes Figure 1 and Figure 2 and ranks the eleven systems according to their MARA-scores.

Several aspects are worth noting. First, the resulting ranking is consistent with what one would expect given the overall impression of the technical sophistication of the individual systems. This shows that the overall approach is capable of producing plausible results, as our test run did not lead to totally unexpected or counter-intuitive outcomes. Second, systems with a swarming ability are assessed as riskier when actually used in a swarm rather than individually. Third, the systems with the highest scores are not fielded yet but are either on the drawing board (UCLASS) or currently contemplated by technicians and defense policy-makers (robot insect). Finally, the fact that the piloted Eurofighter as one of the artificial test cases ends up with a rather high MARA-score (the fourth highest of all the systems in this study) underlines the fact that many manned systems in use today already feature a rather high degree of technical sophistication and autonomous functionality (as defined within the scope of our study). While the Eurofighter is still below the 50% threshold, it would rise far beyond that when (hypothetically) being untethered from constant human control (due to the resulting increase in vector values V_{IP3}, V_{IP4} and V_H).

Case Studies

To get a better understanding of MARA and the interpretation of the results our instrument produces, we will discuss three examples in more detail below. The Guardium exemplifies a single, land-based system, the Reaper and UCLASS are used for a comparison of two generations of unmanned aerial combat vehicles, and CARACaS illustrates the difference between the use of single systems and a swarm.

Example 1: Guardium

System	V_{P1}	V_{P2}	V_{P3}	V_{A1}	V_{A2}	V_{A3}	V_{H1}	V_{IP1}	V_{IP2}	V_{IP3}	V_{IP4}	V_{E1}	V_{E2}	V_{E3}	Weight (n=1)	MARA	MARA %
Guardium	5	4	4	3	3	5	3	4	3	4	3	6	8	5	+10	70	41.18
Sum of Vector Group			13			11	3				14			19			

Table 3: Vector values and MARA results of the Guardium System

The Israeli Guardium system is advertised by its manufacturer as a "[f]ully-autonomous unmanned ground vehicle for complex combat missions" (G-NIUS 2008). However, with a MARA% score of only 41.18 it scores below the 50% threshold in our assessment. This is because we assumed that despite the boosting rhetoric in the advertisement there is still a significant amount of human control over the system when it comes to the use of its weapons ($V_H = 3$). In addition, we assumed that the interaction with its environment (V_{IP3}) and its mission tasking capabilities (V_{IP4}) are rather basic, leading to scores of 4 and 3 respectively. Should the autonomy of the system improve in the future, though, a significantly higher MARA-score due to the weighting factor n would result.

Example 2: Reaper vs. UCLASS

System	V_{P1}	V_{P2}	V_{P3}	V_{A1}	V_{A2}	V_{A3}	V_{H1}	V_{IP1}	V_{IP2}	V_{IP3}	V_{IP4}	V_{E1}	V_{E2}	V_{E3}	Weight (n=1)	MARA	MARA %
Reaper	7	4	6	4	6	3	2	5	4	3	2	10	4	5	+7	72	42.35
UCLAS	8	4	7	5	6	3	8	9	6	7	5	10	8	1	+20	107	62.94

Table 4: Vector values and MARA results of the Reaper and UCLASS System

The comparison of the already fielded Reaper with a future UCLASS system reveals how the 50% threshold can be crossed due to technological developments in unmanned aerial vehicles. While we do not assume that the UCLASS's physical characteristics will be markedly different from the Reaper (V_P), the grade of human control will be significantly different due to the difference in mission specification. In the case of this comparison, this is the change from remotely piloted close air support in an uncontested airspace to a dogfighting capability in contested airspace necessitating limited human control ($V_{H1}^{R} = 2$ vs. $V_{H1}^{UC} = 8$). In result, the UCLASS scores higher throughout the entire vector group "information processing" ($V_{IP}^{R} = 14$ vs. $V_{IP}^{UC} = 27$). Given the additional impact of the weighting factor, this difference in information processing capabilities as well as the lack of immediate human control moves the UCLASS beyond the Reaper by a significant margin on the MARA-scale.

Example 3: CARACaS vs. CARACaS Swarm

System	V_{P1}	V_{P2}	V_{P3}	V_{A1}	V_{A2}	V_{A3}	V_{H1}	V_{IP1}	V_{IP2}	V_{IP3}	V_{IP4}	V_{E1}	V_{E2}	V_{E3}	Weight (n=1)	MARA	MARA %
CARACaS	4	4	4	3	1	1	4	4	4	4	3	5	8	5	+11	65	38.24
CARACaS Swarm (10)	4	4	4	3	1	4	8	4	4	6	3	6	8	5	+17	81	47.65

Table 5: Vector values and MARA results of the CARACaS System; single and swarm

When comparing a single system with a swarm composed of various units of that same system, many vector values stay the same as the central physical characteristics of the system remain constant. However, we assume that, first, human control will decrease for the swarm, as it is a central characteristic of a swarm that it co-ordinates itself according to a certain "swarm intelligence" without external interference (V_{H1}). Second, we assume that a swarm will have enhanced capabilities to interact with its environment in order to function as a swarm (V_{IP3}). Finally, a swarm can be more capable in terms of damage output as each member of the swarm can attack an individual target, thereby increasing the "kill cycle" of the overall system significantly (V_{A3}). As the ability to act as a swarm is closely linked to an increased level of autonomy, the weighting factor comes into play as well in this case. In result, a swarm of the same system ends up with a significantly higher MARA-score than each individual system would on its own, underlining the fact that swarms of the same system need to be assessed as more risky than a single unit of that same type of system.

Conclusion

The currently ongoing debate between proponents and critics of LAWS can – somewhat provocatively – be summed up like this: Proponents charge critics with being luddites or at the very least ignorant towards the plethora of potential benefits attainable by developing and fielding LAWS. In turn, critics judge proponents and their expectations as either naïve (regarding the hope for a more humane way of warfare) or as blind towards the fact that short-term benefits are outweighed by long-term risks, such as an ensuing arms race, regional and global instability, the erosion of the laws of war, and the ethical dilemma conjured up by "outsourcing" the decision to killer other human beings to a machine.

There is still a considerable amount of conjecture on both sides of the debate, however. The political process that is underway at a global level in general and within United Nations fora in particular requires a more reliable and empirically substantiated base for discussion. All parties involved are in need of a clearer notion of how autonomy in weapons systems can be grasped – and what the process of reaching such a common understanding implies for possibly regulating or banning lethal autonomy in weapons systems in turn.

We are confident that the instrument for multidimensional autonomy risk assessment presented in this study can help remedy this situation through the development of comprehensive and transparent quantitative descriptions of weapons systems, both deployed and currently under development. We have shown in our study that it is, in general, possible to develop such instruments to better assess lethal autonomous weapons systems. The instrument we presented proves that a well-designed set of indicators can lead to plausible results when calculating an overall index, the MARA in our case, thoroughly. This, we believe, has the potential to move the debate about regulating autonomy in weapons systems forward in a sober and productive fashion.

Bibliography

Arkin, Ronald C. 2010: The Case for Ethical Autonomy in Unmanned Systems, in: Journal of Military Ethics 9 (4): 332-341.

Article 36 2014: Key Areas for Debate on Autonomous Weapons Systems. Retrieved 15/10/2015 from http://www.article36.org/wp-content/uploads/2014/05/A36-CCW-May-2014.pdf.

Asaro, Peter 2012: On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making, in: International Review of the Red Cross 94 (886): 687-709.

Carpenter, R. Charli 2013: Beware the Killer Robots: Inside the Debate over Autonomous Weapons, Foreign Affairs. Retrieved 15/10/2015 from http://www.foreignaffairs.com/articles/139554/charli-carpenter/beware-the-killer-robots#cid=soc-twitter-at-snapshot-beware_the_killer_robots-000000.

FLI 2015: Autonomous Weapons: an Open Letter from AI & Robotics Researchers. The Future of Life Institute. Retrieved 31/08/2015 from http://futureoflife.org/AI/open_letter_autonomous_weapons#signatories.

Germany 2015: General Statement by Germany, CCW Expert Meeting, Lethal Autonomous Weapons Systems, Geneva 13-17 April 2015.

G-NIUS Unmanned Ground Systems 2008: Guardium MK II. Retrieved 15/10/2015 from http://g-nius.co.il/unmanned-ground-systems/guardium-mk-iii.html.

HRW 2015: Mind the Gap: The Lack of Accountability for Killer Robots, Human Rights Watch. Retrieved 15/10/2015 from https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots.

ICRAC 2009: Statements by the International Committee for Robot Arms Control. Retrieved 31/08/2015 from http://icrac.net/statements/.

McFarland, Tim/McCormack, Tim 2014: Mind the Gap: Can Developers of Autonomous Weapons Systems be Liable for War Crimes?, in: International Law Studies 90: 361-385.

Sauer, Frank 2014a: Autonomous Weapons Systems. Humanising or Dehumanising Warfare?, in: Global Governance Spotlight 4.

Sauer, Frank 2014b: Einstiegsdrohnen: Zur deutschen Diskussion um bewaffnete unbemannte Luftfahrzeuge, in: Zeitschrift für Außen- und Sicherheitspolitik 7 (3): 343-363.

Sauer, Frank/Schörnig, Niklas 2012: Killer Drones - The Silver Bullet of Democratic Warfare?, in: Security Dialogue 43 (4): 363-380.

Scharre, Paul D./Horowitz, Michael C. 2015: An Introduction to Autonomy in Weapon Systems, CNAS Working Paper, Center for a New American Security. Retrieved 15/10/2015 from http://www.cnas.org/sites/default/files/publications-pdf/Ethical%20Autonomy%20Working%20Paper_021015_v02.pdf.

Scola, Nancy 2014: DOD wants to build drones that can buzz into bad guys' doorways, Washington Post. Retrieved 15/10/2015 from https://www.washingtonpost.com/news/the-switch/wp/2014/12/30/dod-wants-to-build-drones-that-can-buzz-into-bad-guys-doorways/.

Sharkey, Noel/Suchman, Lucy 2013: Wishful Mnemonics and Autonomous Killing Machines, in: Artificial Intelligence and the Simulation of Behaviour (AISB) Quarterly (136): 14-22.

Smalley, David 2014: The Future Is Now: Navy's Autonomous Swarmboats Can Overwhelm Adversaries, Office of Naval Research. Retrieved 15/10/2015 from http://www.onr.navy.mil/en/Media-Center/Press-Releases/2014/autonomous-swarm-boat-unmanned-caracas.aspx.

stopkillerrobots.org 2015: Campaign to Stop Killer Robots Website. Retrieved 15/10/2015 from http://www.stopkillerrobots.org/.

Tucker, Patrick 2014: The Military Wants Smarter Insect Spy Drones, Defense One. Retrieved 15/10/2015 from http://www.defenseone.com/technology/2014/12/military-wants-smarter-insect-spy-drones/101970/.

US Department of Defense 2012: Directive: Autonomy in Weapon Systems, Department of Defense, Washington, D.C. Retrieved 15/10/2015 from http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf.

US Navy Research 2014: Autonomous Swarm. Retrieved 15/10/2015 from https://www.youtube.com/watch?v=ITTvgkO2Xw4.

Williams, Andrew P./Scharre, Paul D. (Eds.) 2015: Autonomous Systems: Issues for Defense Policymakers, NATO, Capability Engineering and Innovation Division, Headquarters Supreme Allied Commander Transformation, Norfolk, VA.

Annex

Table 6: Multidimensional Autonomy Risk Assessment: Vectors & Scales

Vector Group / Vector	Scale			
V_P) Physical Characteristics				
V_{P1}) Effective Range	1	Locally bound	6	500-1,000 km
	2	0-10 m	7	1,000-5,500 km
	3	10 m -1 km	8	> 5,500 km
	4	1-10 km	9	Global
	5	10-500 km	10	Space
V_{P2}) Period of Application (incl. Stand-by) resp. maintenance cycles	1	< 1 min	6	½-1 week
	2	1-10 min	7	1-4 week
	3	10-60 min	8	1-12 months
	4	1-24 h	9	1-10 years
	5	24-72 h	10	< 10 years
V_{P3}) Speed	1	immobile	6	200-800 km/h
	2	< 10 km/h	7	801 km/h - sound
	3	10-30 km/h	8	1,079.3 km/h (sound at -50°C through air)
	4	30-80 km/h	9	> Mach 3
	5	80-200 km/h	10	7-8 km/s (parabolic speed from Earth)
V_A) Armament Characteristics (standard armament)				
V_{A1}) Weapon Range	1	Locally bound	6	501-1,000 km
	2	0-10 m	7	1,000-5,500 km
	3	10 m -1 km	8	> 5,500 km
	4	1-10 km	9	Global
	5	10-500 km	10	Space
V_{A2}) Kill Radius	1	Impacts only the target	6	50 m
	2	Impacts target and surrounding environment	7	100 m
	3	5 m	8	1 km
	4	10 m	9	10 km
	5	20 m	10	> 100 km

V_{A3}) Kill Cycles (shot x probability of success)	1	1	6	32-63
	2	2-3	7	64-127
	3	4-7	8	128-255
	4	8-15	9	>=256
	5	16-32	10	infinite (laser)
V_{H}) Human Relevance				
V_{H1}) Actual Human Control	1	Remote cockpit	6	Machine recognizes and chooses target; Human prioritizes and fights target
	2	Joystick and assisting systems / keyboard and mouse	7	Machine recognizes, chooses and prioritizes target; Human fights target
	3	Point & click: lines of movement	8	Machine recognizes, chooses, prioritizes and fights target; Human has veto power
	4	Point & click: areas of movement	9	Human can only abort mission as a whole
	5	Machine recognizes target; Human chooses, prioritizes and fights target	10	No human control
V_{IP}) Information Processing/ Situational Awareness				
V_{IP1}) Coverage (parts of a sphere)	1	1/10	6	6/10
	2	2/10	7	7/10
	3	3/10	8	8/10
	4	4/10	9	9/10
	5	5/10	10	10/10
V_{IP2}) Multispectrality (kinds of sensors)	1	1	6	6
Visible light, infrared, magnetic field, radar, pressure, GPS, gravitation, …	2	2	7	7
	3	3	8	8
	4	4	9	9
	5	5	10	10 or more

V_{IP3}) Ability to interact with the environment	1	Cannot interact but change its orientation	6	Is able to implement additional rules and instructions for its movement and navigation (Google Car)
	2	Is able to change orientation and position	7	Is able to classify obstacles and change its behavior dynamically
	3	Is able to safeguard its movement	8	Is able to recognize, distinguish and classify objects and subjects in its environment
	4	Is able to recognize obstacles and circumvent them (take-off, landing, home coming)	9	Is able to dynamically adapt rules by machine learning
	5	Is able to plan and chose tracks (take-off and landing on air craft carriers)	10	Is able to use strong Artificial Intelligence to interact, i.e. by taking strategic decisions based on complex intentions
V_{IP4}) Mission Tasking Capability (= destruction of objects, killing of humans)	1	Is not able to differentiate targets (indiscriminate action)	6	Is able to engage targets by distinction of object classes through advanced detection and detailed classification
	2	Is able only to engage distinct targets by direct preselection (laser illumination, geographical coordination)	7	Is able to engage targets by distinction of object classes through extended detection and complex and robust classification, i.e. even when camouflage is applied
	3	Is able to engage targets through recognition of sources of emission of electromagnetic signatures (radar, infrared, detection of movement)	8	Is able to identify targets by functions and behavior, i.e. Blue Force Identification by observed patterns of behavior (cognitive concept)
	4	Is able to engage by distinction of friend and foe (Blue Force Recognition)	9	Is able to identify targets by functions and prioritized target selection by observed pattern of behavior on a tactical level
	5	Is able to engage targets by distinction of object classes through simple detection and basic classification	10	Is able to identify targets by functions and prioritized target selection by observed pattern of behavior on a strategic level.

V_E) Exposition				
V_{E1}) Interface Range	1	0 m (hard-wired communication interfaces)	6	5-50 km
	2	1-10 m (near-field communication)	7	50-100 km
	3	10-100 m	8	200-1,000 km
	4	100-1,000 m	9	1,000-10,000 km
	5	1-5 km	10	36.000km (communication via geo-stationary satellites, single way)
V_{E2}) Directionality	1	Hard-wired interface	6	Directional characteristic (cardioid lobe)
	2	Directed Point-to-point	7	-
	3	-	8	Hemisphere
	4	Strong directional characteristic (hyper cardioid lobe)	9	-
	5	-	10	Omnidirectional
V_{E3}) Encryption	1	Strategically	6	-
	2	-	7	-
	3	-	8	-
	4	-	9	-
	5	Tactically	10	Clear text