Briefing Paper

CCW Informal Meeting of Experts on Lethal Autonomous Weapons,
Geneva, 12 April 2016, Panel ‘Towards a Working Definition of LAWS’: Autonomy

Situational awareness and adherence to the principle of distinction as a necessary condition for lawful autonomy

Lucy Suchman, Professor of Anthropology of Science and Technology,
Lancaster University, UK

Summary

As a contribution to the CCW's third informal meeting of experts on lethal autonomous weapon systems (LAWS), this briefing paper focuses on the implications of the requirement of situational awareness for autonomous action – whether by humans, machines or complex human-machine systems. For the purposes of this paper, 'autonomy' refers to self-directed action, and more specifically the action-according-to-rule that comprises military discipline. Unlike the algorithmic sense of a rule as that term is used in Artificial Intelligence (AI), military rules always require interpretation in relation to a specific situation, or situational awareness. Focusing on the principle of distinction, I argue that International Humanitarian Law (IHL) presupposes capacities of situational awareness that it does not, and cannot, fully specify. At the same time, autonomy or 'self-direction' in the case of machines requires the adequate specification (by human designers) of the conditions under which associated actions should be taken. This requirement for unambiguous specification of condition/action rules marks a crucial difference between autonomy as a legally accountable human capacity, and machine autonomy. The requirement for situational awareness in the context of combat, as a prerequisite for action that adheres to IHL, raises serious doubts regarding the feasibility of lawful autonomy in weapon systems.

The questions surrounding lethal autonomous weapon systems (LAWS) are being addressed by the Convention on Certain Conventional Weapons (CCW) along multiple lines of analysis. This briefing paper is meant as a contribution to discussions regarding the concept of *autonomy*, on the basis of which I present an argument questioning the feasibility of LAWS that would comply with International Humanitarian Law (IHL).[1] This argument is based not on principle, but rather on empirical evidence regarding the interpretive capacities that legal frameworks like IHL presuppose for their application in a specific situation. These capacities make up what in military terms is named *situational awareness*.[2] Despite other areas of

[1] International Humanitarian Law ‘is a set of rules which seek, for humanitarian reasons, to limit the effects of armed conflict. It protects persons who are not or are no longer participating in the hostilities and restricts the means and methods of warfare.’ See https://www.icrc.org/eng/assets/files/other/what_is_ihl.pdf (accessed 23 March 2016).

[2] Situational awareness can be defined as ‘understanding of the operational environment in all of its dimensions – political, cultural, economic, demographic, as well as military factors.’ Dostal, Major Brad C. (2001). *Enhancing situational understanding through the employment of unmanned aerial vehicles*. Center for Army Lessons Learned. Retrieved from

progress in artificial intelligence (AI) and robotics, it is my assessment that none has been made in the operationalization of situational awareness in an indeterminate environment of action. More specifically for the question of LAWS, situational awareness as a prerequisite for the identification and selection of legitimate targets – what has been named the Principle of Distinction – is not translatable into machine executable code. Yet situational awareness is essential for adherence to IHL or any other form of legally accountable rules of conduct in armed conflict.

This assessment is based on my position as an anthropologist engaged for over three decades with the fields of AI and human-machine interaction.[3] A central aim of my remarks (and of my larger body of research) is demystification of the field of AI, particularly with respect to clarification of the differences between human and machine capabilities. My work in tracking developments in AI and robotics involves taking seriously the claims that are made for intelligent machines and comparing them to extensive studies of the competencies – perceptual, and also crucially social and interactional – that are the basis for associated human activities. My focus on situational awareness in the context of this panel arises not only from the fact that it is a prerequisite for lawful action within the framework of IHL, but also because this is an area in which I hope that my particular perspective can contribute to greater clarity on the key concept of autonomy.

LAWS and The Principle of Distinction

The elements of situational awareness that I believe are most relevant to the question of whether LAWS can be adherent to IHL are those that inform *the requirement of distinction* in the use of lethal force; that is, discrimination between legitimate and non-legitimate targets.[4] I recognize that the requirements of distinction and proportionality are closely linked, but insofar as proportionality judgments presuppose that distinction has been made, I focus on distinction here. In the case of autonomous weapons, adherence to the Principle of Distinction would require that robots have adequate vision or other sensory processing systems, and associated algorithms, for separating combatants from civilians and for reliably differentiating

http://www.globalsecurity.org/military/library/report/call/call_01-18_ch6.htm (accessed 23 March 2016).

[3] Before taking up my present Chair at Lancaster University, I was a Principal Scientist at Xerox's Palo Alto Research Center, where I spent twenty years as a researcher. I have written for both social and information sciences audiences, including two books, *Human-Machine Reconfigurations* (2007) and *Plans and Situated Actions: the problem of human-machine communication* (1987), both published by Cambridge University Press. In 2002 I received the Benjamin Franklin Medal in Computer and Cognitive Sciences, and in 2010 the ACM SIGCHI Lifetime Research Award. In 1983 I was a founding member of Computer Professionals for Social Responsibility, an organisation formed to address the increasing reliance on computing in the control of nuclear weapon systems; I am now a member of the International Committee for Robot Arms Control, and on the Executive Board of the Foundation for Responsible Robotics.

[4] On the Principle of Distinction see http://www.icrc.org/customaryihl/eng/docs/v1_cha_chapter1_rule1 (accessed 23 March 2016). Denise Garcia (2016: 96) has recently called for what she terms 'preventive security governance,' focused on the three areas of a) preventing future harm to civilians, b) responsibility and accountability, and c) the question of what constitutes a legal and legitimate attack.

wounded or surrendering combatants from those who pose an imminent threat. Existing machine sensors such as image processing cameras, infrared temperature sensors, and the like may be able to identify something as a human, but they cannot make the discriminations among persons that are required by the Principle of Distinction.[5] Even if machines had adequate sensing mechanisms to detect the difference between civilians and uniform-wearing military, they would fail under situations of contemporary warfare where combatants are most often not in uniform.[6] And more sophisticated technologies such as facial or gait recognition are still reliant on the existence either of a pre-established database of templates, against which a match can be run, or profiles, which are inherently vulnerable to false positives and other forms of inaccurate categorization.[7]

I take as a working definition of LAWS, weapon systems in which the identification and selection of human targets and the initiation of violent force is carried out under machine control; that is, these capacities are delegated to the system in ways that preclude deliberative and accountable human intervention or what, in the current discussion, has been characterized as 'meaningful human control'.[8] This definition follows that adopted by UN Special Rapporteur on Extrajudicial, Summary or

[5] Some opponents of a ban on LAWS imagine scenarios in which the mere presence of a human body is an adequate criterion for the identification of that person as a legitimate target. But that requirement is counter to the direction in which conflict is moving, as the boundaries that designate geographic zones of combat are increasingly fluid and contested.

[6] On the increasing complexity of the combatant/civilian distinction, Christiane Wilke (2014) observes that 'the rise of the figure of the "unlawful combatant" . . . is accompanied by a corresponding rise of the figure of the illegitimate, noninnocent, suspicious civilian.' C. Wilke, 'Civilians, combatants and histories of international law', 28 July 2014, available at http://criticallegalthinking.com/2014/07/28/civilians-combatants- histories-international-law/.

[7] With respect to the development of algorithmic templates for the identification of legitimate targets, Susan Schuppli (2015) observes that 'the recently terminated practice of "signature strikes" in which data-analytics were used to determine emblematic "terrorist" behaviour and match these patterns to potential targets on the ground already points to a future in which intelligence gathering, assessment, and military action, including the calculation of who can legally be killed, will largely be performed by machines based upon an ever expanding database of aggregated information.' S. Schuppli, Deadly Algorithms, *Continent*, Issue 4.4, pp. 20-27, available at http://www.continentcontinent.cc/index.php/continent/article/view/212. The concern here is with an increasing push towards reliance on *a priori* stereotyping, rather than systematic intelligence gathering; it is the unreliability of stereotyping that has discredited this practice. On the exacerbation of the problem of targeted killing by LAWS see also Heyns, UN Doc. A/HRC/23/47, 'Report of the special rapporteur'.

[8] For the minimum necessary conditions for meaningful human control see Article 36 (2013); Article 36 (2014); Frank Sauer, *ICRAC Statement on Technical Issues to the 2014 UN CCW Expert Meeting*, ICRAC (May 14, 2014), http://icrac.net/2014/05/icrac-statement-on-technical-issues-to-the-un-ccw-expert-meeting. The word 'meaningful' here is meant to anticipate and reject the proposition that any form of oversight over automated target identification constitutes adequate human control. Horowitz and Scharre (2015: 4) propose that in its emerging usage, 'meaningful human control has three essential components: Human operators are making informed, conscious decisions about the use of weapons; human operators have sufficient information to ensure the lawfulness of the action they are taking, given what they know about the target, the weapon, and the context for action; and the weapon is designed and tested, and human operators are properly trained, to ensure effective control over the use of the weapon.'

Arbitrary Executions, Christof Heyns (2013), who defines LAWS as 'robotic weapons systems that once activated, can select and engage targets without further intervention by a human operator.'[9] The emphasis in this discussion is specifically on *human* targets; that is, the identification of humans or human-inhabited objects (buildings, vehicles) as lawful targets for engagement. I am bracketing, in other words, defensive weapon systems that operate on the basis of unambiguous signals from another (unmanned or uninhabited) device that comprises an imminent threat.

The fundamental problem in meeting the requirements of the Principle of Distinction is that we do not have a definition of a civilian that can be translated into a recognition algorithm. Nor can we get one from IHL.[10] The 1949 Geneva Convention requires the use of 'common sense,' while the 1977 Protocol I essentially defines a civilian in the negative sense, as someone who is not a combatant.[11] While robotics may achieve effective sensory and visual discrimination in certain narrowly constrained circumstances, human level discrimination with adequate common sense reasoning for situational awareness would appear to be computationally intractable.[12] At this point, at least, there is no evidence or research result to suggest otherwise.

Relations of automation and autonomy

While drawing a line between automation and autonomy is necessary in the context of the CCW's deliberations, this does not imply that autonomous systems are not automated. The crucial question, rather, is whether or not an automated system is subject to meaningful human control.[13] We could, in other words, define autonomous systems precisely as those in which the identification, selection and engagement of targets has been fully automated – this definition still provides a clear distinction between automated systems under human control and those that are not (i.e., weapons systems that are acting autonomously).

It is also the case that the question of autonomy with respect to LAWS needs to be

[9] See also Sharkey, 2012. Scharre and Horowitz (2015: 16) offer a closely related definition, but one focused more specifically on the question of targeting, viz.: 'An autonomous weapon system is a weapon system that, once activated, is intended to select and engage targets where a human has not decided those specific targets are to be engaged.' Addressing the key phrase 'select and engage', Gubrud (2014) observes that 'selection' or targeting is complicated by the fact that 'the status of an object as the target of a weapon is an attribute of the weapon system or persons controlling and commanding it, not of the object itself.' Target selection, Gubrud argues, is where the crucial questions and indeterminacies lie, and the operator, 'the final human in the so called "kill chain" or "loop"' should be the final decision point.

[10] Asaro (2009) reminds us that IHL comprises a diverse body of international laws and agreements (such as the Geneva Conventions), treaties, and domestic laws. These are far from algorithmic specifications for decision-making and action.

[11] Art 50(1) of the Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, 8 June 1977.

[12] See Sharkey, N. (2008) Grounds for Discrimination: Autonomous Robot Weapons, in RUSI Defence Systems, Vol. 11, No. 2, pp. 86-89.

[13] In a distinction consistent with the definition adopted here, Scharre and Horowitz (2015: 17) propose the crucial difference as that between a weapon that is selecting targets without human decision ('self-targeting' or autonomous) and a weapon engaging a human-selected target. See also Mark Gubrud, *Autonomy Without Mystery: Where* Do *You Draw the Line?*, 1.0 HUM. (May 9, 2014), http://gubrud.net/?p=272.

considered within a longer history of the intensifying automation of warfare. This is a trajectory justified as a necessary response to the demand for increasingly rapid engagement, along with the vulnerabilities incurred through reliance on complex information and communications networks; a problem that greater automation and system complexity further exacerbates.

We have seen these dynamics before in the case of launch on warning in nuclear weapon systems, and some of the questions currently under debate were addressed, and arguably resolved, in the work of computer scientists like David Parnas in the 1980s.[14] In the context of the US Strategic Defense Initiative, Parnas made the crucial distinction between a computational system's verifiable execution of its specifications on one hand (what is commonly referred to as the software's 'correctness,' or reliability in the narrow sense described in Asaro 2015: 90), and the system's ability to assess the conditions in which those specifications apply on the other (necessary for its reliability in operation). Simulated testing of automated weapon systems can assess correctness, but it can never definitively assure reliability under actual conditions of use. The only way to achieve the latter is through practical methods of iterative development based on repeated trials under conditions that closely match those of intended deployment, or informed by experience of the system in use, neither of which is possible in the case of complex weapon systems with deadly consequences. It was for this reason, among others, that the Strategic Defense Initiative was finally abandoned.

The US Department of Defense *Unmanned Systems Integrated Roadmap 2011–2036* distinguishes automatic from autonomous systems in this passage:

> Dramatic progress in supporting technologies suggests that unprecedented levels of autonomy can be introduced into current and future unmanned systems. . . Automatic systems are fully preprogrammed and act repeatedly and independently of external influence or control. . . However, the automatic system is not able to define the path according to some given goal or to choose the goal dictating its path. By contrast, autonomous systems are self-directed toward a goal in that they do not require outside control*, but rather are governed by laws and strategies that direct their behavior. . . The special feature of an autonomous system is its ability to be goal-directed in unpredictable situations*. This ability is a significant improvement in capability compared to the capabilities of automatic systems.[15]

The key phrase here is 'governed by laws and strategies that direct their behaviour … in unpredictable situations.' As I have stated above, 'laws and strategies' are not translatable to executable code. In assessing the feasibility of the system posited in this passage, it is crucial to keep in mind that autonomy or 'self-direction' in the case of machines presupposes the unambiguous specification (by human designers) of the conditions under which associated actions should be taken. And this requirement for unambiguous specification of condition/action rules marks a crucial difference

[14] See Parnas D.L. (December 1985), Software aspects of strategic defense systems. *Comm ACM* 28 (12): 1326–35; see also Smith, B.C. (December 1984), The Limits of Correctness. *ACM SIGCAS*, Computers and Society: Volume 14,15 Issue 1,2,3,4, Jan 1 1985.

[15] US DOD *Unmanned Systems Integrated Roadmap FY2011-2036*, p. 43, my emphasis.

between autonomy as a human capacity, and machine autonomy. As I have argued in previous writing, autonomy as we understand it in the context of human action means self-direction under conditions that are not, and cannot be, fully specified by rule.[16] This in turn accounts for what we might call the *strategic vagueness* of any kind of rule or directive for action; that is, the assumption that the exercise of the rule, or the execution of the directive or plan, will involve *in situ* judgment regarding the rule's application. Where the requisite competencies are in place, this openness – far from being a problem – is what enables the effectiveness of a general plan or rule as a referent for situated action.

Limits to information processing as a model of situational awareness

To make this more concrete we can take the case of the human action-according-to-rules which defines military discipline, and most pertinent to this discussion IHL and the rule of distinction. Because the precise conditions of action in combat cannot be known in advance, the directives for action in the case of military operations presuppose competencies for their accurate 'execution' that the directive as such does not and cannot fully specify. Thus the requirement for situational awareness, as necessary to effective and, most importantly for our purposes, legally accountable warfare.

Approaches to AI-based robotics share the common requirement that a machine can engage in a sequence of 'sense, think and act'.[17] It is crucial in this context, however, to be wary of the use of evocative terms that imply the functionality of programs, rather than providing technical descriptions of actual capabilities.[18] Does 'sense, think and act' refer to an assembly line robot that performs an action at a fixed location, in relation to an environment carefully engineered to match its sensing capacities, and where the consequences of failure are non-lethal? Or does it invoke sensing and perception as dynamic and contingent capacities, in open-ended fields of (inter)action, with potentially lethal consequences? In the case of human combatants, the ability to be goal-directed in unpredictable situations presupposes capacities of situational assessment and judgment, in circumstances where the range of those capacities is necessarily open ended. Combat situations, moreover, frequently involve opponents who work hard and ingeniously to identify and defeat any prior assumptions about how they will behave. This is in marked contrast to the situations in which AI techniques, and automation more generally, have been successfully applied. In any case, the burden of proof here must rest with proponents, and require a higher standard than general assertions of progress in artificial intelligence, which is debatable other than in certain, limited technical areas that do not yet begin to address problems of reliable discrimination between legitimate and illegitimate human

[16] See Suchman, L. (1987) *Plans and Situated Actions: the problem of human-machine communication*. Cambridge University Press; and (2007) *Human-Machine Reconfigurations*. Cambridge University Press. This problem is not resolved by the promises of 'machine learning' to enable derivations from data external to the specified rule, insofar as learning algorithms continue to rely upon the availability of specified data formats rather than open-ended horizons of input (see Asaro 2015).

[17] See Suchman and Weber 2016, 9-10.

[18] See Sharkey, N. and Suchman, L. (2013) Wishful Mnemonics and Autonomous Killing Machines. *AISBQ Quarterly, the Newsletter of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour, 136*, 14-22.

targets.[19]

A final note is that autonomy is best understood not as an individual capacity – whether human or machine – but rather as a capacity enabled by particular configurations of people and technologies. Different configurations make different capacities for action possible. In thinking about life critical technical systems, it is the question of what conditions a particular configuration affords for human responsibility and accountability that is key. This is where the concept of meaningful human control becomes crucial: What is required to ensure that delegations of agency to machines allow the preservation of human responsibility and accountability? In a report issued in February of this year,[20] UN Special Rapporteurs Maina Kiai and Christof Heyns wrote that 'Autonomous weapons systems that require no meaningful human control should be prohibited.' I would add that it is not only the case that autonomous weapon systems might circumvent meaningful human control; the greater concern is that they could render it impossible. The judgment required for effective and legal action-according-to-rule requires time for the assessment of a current situation, and decreasing timeframes due to increasing automation close down the time available for assessment. The proposed solution of 'human-machine teaming,' moreover, is only effective to the extent that system designs maintain the system dynamics (more colloquially, the time) required to allow for meaningful human control.[21] This requirement, in turn, poses further limits to weapon system automation.

Implications for lawful weapon system autonomy

Conceptual clarity regarding the capacities that enable situational awareness in the case of human combatants, with a particular focus on the Principle of Distinction, clarifies in turn the requirements for autonomous technologies, and more specifically for LAWS. Citing Article 48 of the First Additional Protocol to the Geneva Conventions, Crootof (2015: 1873) observes that one implication of the Principle of Distinction is that:

> parties are prohibited from using inherently indiscriminate weapons, which are usually defined either as weapons that cannot be directed at lawful targets or as weapons whose effects cannot be controlled. Additionally, any given attack in an armed conflict cannot be indiscriminate: it must be directed at a lawful target and cannot utilize indiscriminate weapons or methods of warfare.

The defining question for LAWS is whether the discriminatory capacities that are the precondition for legal killing can be reliably and unambiguously encoded in weapon

[19] Assertions that 'technology may evolve and meet the requirements [for human target identification] in the future' (cited in Report of the 2015 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), CCW/MSP/2/015/3, 2 June 2015, p. 14), or 'Autonomous technologies *could* lead to more discriminating weapons systems' (ibid., p. 15, my emphasis) do not comprise evidence-based statements of fact.

[20] See Kiai and Heyns (2016) Joint report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association and the Special Rapporteur on extrajudicial, summary or arbitrary executions on the proper management of assemblies, 4 February 2016.

[21] See Scharre, P (2016) 'Autonomous Weapons and Operational Risk,' Center for A New American Security, http://www.cnas.org/autonomous-weapons-and-operational-risk.

systems. As noted above, this judgment has becoming increasingly difficult for human warfighters, for several reasons. First, the conditions of so-called irregular warfare have removed traditional designations of battle zones and combatants, requiring much more subtle and uncertain readings of the presence of an imminent threat.[22] Second, because military systems involve increasingly complex, distributed, real-time networks of information and communication, the possibilities have amplified not only for strategic accuracy, but also for noise.[23] And finally, the intensifying automation of warfare has effected a progressive narrowing of timeframes for situational assessment.

Lawand (2013) proposes that 'A truly autonomous weapon system would be capable of searching for, identifying and applying lethal force to a target, including a human target (enemy combatants), without any human intervention or control.'[24] But in the parenthetical 'enemy combatants' lies the crux of the problem: how is the identification of 'human target' with 'enemy combatant' confirmed? And what uncertainties characterise the category of 'enemy combatant' that confound, rather than clarify, the problem of legitimate target identification in contemporary warfare? Autonomous systems can be made reliable only to the extent that their environments, the conditions of their operation, can be specified, engineered and stabilized; these requirements do not hold in situations of combat.[25] All of the evidence to date indicates that this is at best an unsolved problem for machine autonomy, and at worst (and this is my position, for the reasons set out above) an unsolvable one.

In sum:

1. We take as our definition of lethal autonomous weapons, robotic weapons systems that once activated, can select and engage *human* targets without further intervention by a human operator.

2. Autonomy in human or machine systems implies self-directed action, including crucially in the case of military operations, action-according-to-rules.

[22] Melissa L. Flagg, Deputy Assistant Secretary of Defense at the U.S. Office of the Undersecretary of Defense for Acquisition, Technology and Logistics' research directorate, imagines a situation in which 'a robotic system is in a battle zone, knows the mission, has been thoroughly tested, has the kinetic option, and its communications links have been cut off,' and then asks whether that machine should then make the decision to deploy a weapon independently. But it is precisely this clarity that is absent in actual situations of war fighting. (Autonomous, Lethal Robot Concepts Must Be 'On the Table,' DoD Official Says,' Stew Magnuson, *National Defense Magazine* http://www.nationaldefensemagazine.org/blog/Lists/Posts/Post.aspx?ID=2110).

[23] On the intransigence of this problem see for example Patrick Cronin (2008) *The impenetrable fog of war: reflections on modern warfare and strategic surprise*. Westport, CT: Praeger Security International.

[24] Kathleen Lawand, *Fully Autonomous Weapon Systems*, INT'L COMMITTEE RED CROSS (Nov. 25, 2013), http://www.icrc.org/eng/resources/documents/statement/2013/09-03-autonomous-weapons.htm.

[25] It is widely recognized that 'as the behavior of automated systems becomes more complex, and more dependent on inputs from environmental sensors and external data sources, the less predictable they become' (Asaro, 2015). And as the last expert's panel observed 'Deploying a weapon system with unpredictable effects creates a significant risk of a breach of International Humanitarian Law' (Report, 2015: 15).

3. Action-according-to-rules in the case of human action presupposes capabilities that the rules cannot fully specify; in particular, those competencies that are required to map the conditions assumed by the rule to actually occurring situations.

5. In the case of LAWS that would be adherent to IHL, machine autonomy requires reliable, unambiguous translation of rules for situational awareness, particularly for the identification of legitimate human targets (the Principle of Distinction), into machine executable code.

6. Contrary to assertions regarding the rapid advance of artificial intelligence and robotics, there is no empirical evidence of progress in operationalizing the capacities of situational awareness that are required for adherence the Principle of Distinction.

7. This raises serious doubts regarding the feasibility of lethal autonomous weapons adherent to IHL.

To conclude, conceptual clarity regarding the capacities that enable situational awareness in the case of human combatants, with a particular focus on the Principle of Distinction, clarifies in turn the requirements for lethal autonomous weapon systems. The defining question for autonomous weapons is whether the discriminatory capacities that are the precondition for legal killing can be reliably and unambiguously encoded. My argument is that they cannot, and that as a consequence lethal autonomous weapons are in violation of IHL, and should be prohibited.

Appendix: Response to questions from Ambassador Biontino's 'Food for Thought' paper, distributed to panel participants:

a) Can autonomy in LAWS be best understood or defined in relation to the critical functions of a weapon (i.e. the selection, engagement and tracking of a target)?

> I would say that the answer is 'yes,' and that these are indeed the critical functions. The key term here is 'selection,' as it's there that the question of distinction is most salient and, correspondingly, the requirement for meaningful human control.

In how far could such an approach be operationalized for the purposes of developing regulations and policies on LAWS?

> I would assume that it would be possible to assess the key operation of 'selection' in ways that would make a prohibition on LAWS practicable.

b) Can autonomy be defined objectively?

> I would say yes, if by objective we mean in a way that is accountable to agreed upon criteria.

c) Is 'predictability' a useful indicator to measure the level of autonomy of a system? In how far can the notion of predictability be operationalized in a definition?

> Predictability becomes problematic as systems (including automated systems with human involvement) increase in speed and complexity. To the extent that a human-machine system is unpredictable, it requires the building in of adequate opportunities for review and deliberation. It's in that respect that autonomy introduces a new set of problems. (Cf. the case of launch on warning.)

d) Can the level of a weapon system's autonomy be assessed by a set of indicators? For example, physical characteristics, etc. alongside the level of human control? What are the advantages and disadvantages of such an approach?

> While levels or degrees of system automation can certainly be assessed, human control defines the discontinuity that is relevant to the differentiation of automated (including so-called 'semi-autonomous') from autonomous systems. More specifically, there is a discontinuity at the point where target selection and engagement become automated to the extent that human control is not an in-built system requirement. In this sense 'autonomy' is a discrete system indicator rather than a matter of level or degree.

e) Does the level of human control of a weapon system assist in identifying what is a LAW? How could the required level of human control be defined?

> Yes, the level, or more specifically kind, of human control is key and must be defined in terms of the possibility of making a deliberative judgment regarding the lawfulness of targets.

f) What is “meaningful human control” of a weapon system?

See (e)

g) What is the role of human judgment in the targeting process? How is ‘human judgment’ put into effect over a weapons system? Does the level of human judgment in the targeting process of a weapon system assist with identifying what is a LAW?

Human judgment is key, and irreplaceable, in the identification of lawful human targets. This is the core of my argument, and establishes the limiting factor on the legality of LAWS.

h) Are there other approaches to defining LAWS?

I think the current working definition is a sound basis on which to develop requisite legal frameworks.

References cited

Article 36 (2013) Killer Robots: UK Government Policy on Fully Autonomous Weapons. http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf

Article 36 (2014) Key areas for debate on autonomous weapon systems: memorandum for delegates at the Convention on Certain Conventional Weapons', paper presented at the Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, 13–16 May 2014, available at www.article36.org/wp-content/uploads/2014/05/A36-CCW-May-2014.pdf

Asaro, P. (2009) How Just Could a Robot War Be? In P. Brey, A. Briggle, & K. Waelbers (Eds.), *Current Issues in Computing And Philosophy* (pp. 50-64). Amsterdam: IOS Press.

Asaro, P. (2015) Roberto Cordeschi on Cybernetics and Autonomous Weapons: Reflections and Responses, *Paradigmi: Rivista di critica filosofica*, Anno XXXIII, no. 3, Settembre-Dicembre, 2015, pp. 83-107.

Patrick Cronin (2008) *The impenetrable fog of war: reflections on modern warfare and strategic surprise*. Westport, CT: Praeger Security International.

Crootof, A. (2015) The Killer Robots Are Here: Legal and policy implications. Cardozo Law Review, Vol. 36, pp. 1837-1915.

Dostal, Major Brad C. (2001) *Enhancing situational understanding through the employment of unmanned aerial vehicles*. Center for Army Lessons Learned. Retrieved from http://www.globalsecurity.org/military/library/report/call/call_01-18_ch6.htm

Garcia, D. (2016) Future arms, technologies, and international law: Preventive security governance. *European Journal of International Security, 1*(1), 94-111.

Gubrud, M. (2014) Autonomy without mystery: where do you draw the line?, 9 May 2014, available at http://gubrud.net/?p=272.

Heyns, C. (2013) UN Doc. A/HRC/23/47, Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions', United Nations (9 April 2013).

Horowitz, M. and Scharre, P. (2015) Meaningful Human Control in Weapon Systems: A Primer. *Center for a New American Security Working Paper*, March, 2015.

Kiai, M. and Heyns, C. (2016) Joint report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association and the Special Rapporteur on extrajudicial, summary or arbitrary executions on the proper management of assemblies, 4 February 2016. A/HRC/31/66, available at http://freeassembly.net/reports/managing-assemblies/.

Lawand, K. (3013) *Fully Autonomous Weapon Systems*, International Committee of the Red Cross (Nov. 25, 2013), available at http://www.icrc.org/eng/resources/documents/statement/2013/09-03-autonomous-weapons.htm.

Parnas D.L. (December 1985), Software aspects of strategic defense systems. *Comm ACM* 28 (12): 1326–35.

Report of the 2015 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), CCW/MSP/2/015/3, 2 June 2015.

Sauer, F. (2014) *ICRAC Statement on Technical Issues to the 2014 UN CCW Expert Meeting*, ICRAC (May 14, 2014), http://icrac.net/2014/05/icrac-statement-on-technical-issues-to-the-un-ccw-expert-meeting.

Scharre, P (2016) 'Autonomous Weapons and Operational Risk,' Center for A New American Security, http://www.cnas.org/autonomous-weapons-and-operational-risk.

Scharre, P. and Horowitz, M. (2015) Autonomy in Weapon Systems. *Center for a New American Security Working Paper*, February, 2015.

Schuppli, S. (2015) Deadly Algorithms, *Continent*, Issue 4.4, pp. 20-27, available at http://www.continentcontinent.cc/index.php/continent/article/view/212

Sharkey, N. (2008) Grounds for Discrimination: Autonomous Robot Weapons, in RUSI Defence Systems, Vol. 11, No. 2, 2008, pp. 86-89.

Sharkey, N. (2012) Automating warfare: lessons learned from the drones, *Journal of Law, Information and Science*, vol. 21, no. 2.

Sharkey, N. and Suchman, L. (2013) Wishful Mnemonics and Autonomous Killing Machines. *AISBQ Quarterly, the Newsletter of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour, 136*, 14-22

Smith, B.C. (December 1984), The Limits of Correctness. *ACM SIGCAS*, Computers and Society: Volume 14,15 Issue 1,2,3,4, Jan 1 1985.

Suchman, L. (1987). *Plans and Situated Actions: the problem of human-machine communication*. New York: Cambridge University Press.

Suchman, L. (2007). *Human-Machine Reconfigurations.* New York: Cambridge.

Suchman, L. and Weber, J. (2016) Human-Machine Autonomies. In N. Bhuta, S. Beck, R. Geis, H-Y Liu, and C. Kreis (eds.) *Autonomous Weapons Systems*. Cambridge, UK: Cambridge University Press, pp. 75-102.

US Department of Defense (DoD) (2012) Directive 3000.09, 'Autonomy in weapon systems', 21 November 2012, available at www.dtic.mil/whs/directives/corres/pdf/300009p.pdf

C. Wilke, 'Civilians, combatants and histories of international law', 28 July 2014, available at http://criticallegalthinking.com/2014/07/28/civilians-combatants-histories-international-law/