The Distraction of Full Autonomy and the Need to Refocus the CCW Laws Discussion on Critical Functions

Chris Jenks*

Introduction

“During the whole debate on technical issues of lethal autonomous weapons systems, the notion of autonomy and its definition was at the center of interest. It became quite obvious that there is no ready-made, generally accepted definition of what is an ‘autonomous system’ and as to where to draw the line between ‘autonomous’ and ‘automatic’ or ‘automated’.”

That apt description summarises the results of the first meeting of experts on lethal autonomous weapons systems (LAWS), held in Geneva, Switzerland, in May 2014,

* Assistant Professor of Law, SMU Dedman School of Law, Dallas, Texas. This article supplements a presentation delivered as part of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons as part of the Convention on Certain Conventional Weapons treaty.

under the auspices of the United Nations’ Convention on Certain Conventional Weapons (CCW). The international community held the third such meeting in May 2016, at which States Parties agreed to recommend to the upcoming CCW Fifth Review Conference that a group of governmental experts (GGE) meet in 2017-2018. With the amount of time and effort already expended, and likely with (at least) two more years of meetings to come, it seems appropriate to take stock of the LAWS discussions to date.

Bluntly stated, the LAWS discussions have been confused, not constructive, and largely for the same definitional reasons identified two years ago. This section attempts to address the question why the dialogue has proceeded as it has, and proposes how it should proceed at the GGE meetings.

Attention and concern on LAWS has rapidly gained momentum, leading to CCW States Parties agreeing to convene the informal discussions of the last three years. Yet, despite those meetings, there still is not even consensus about what is being discussed, due, at least in part, to the varied understandings and meanings of autonomy. This section focuses on the problems created by framing the LAWS discussion in terms of full autonomy and suggests that CCW States Parties refocus on the critical functions of selecting and engaging targets.

The results of the LAWS discussions to date are akin to a car racing into a cul-de-sac. That the car has reached the cul-de-sac quickly is meaningless. And once there, the options are limited and unproductive – stop, withdrawal, or
drive in circles. In order for that to change, the LAWS discussion must move beyond and away from the conceptual framework of full autonomy. That framework both confuses and distracts – and has driven the dialogue in an unproductive circular direction. Focusing on the critical functions of selecting and engaging of targets may facilitate progress, and is consistent with CCW’s purpose “to ban or restrict the use of specific types of weapons that are considered to cause unnecessary or unjustifiable suffering to combatants or to affect civilians indiscriminately”.

To explain why framing the discussion in terms of full autonomy dooms the LAWS discussion to be perpetually circular in nature, Part I of this paper explores the challenges in trying to define and otherwise categorise system autonomy. This includes why distinguishing systems based on whether a human is in, on, or out of the loop, or whether a system is automated, automatic, or autonomous, is only of minimal, descriptive, utility. Part II then applies the concept of autonomy to weapons systems, focusing first on the reality that weapons systems capable of selecting and engaging targets without further human intervention have existed for decades, and then on explaining why full autonomy is a distractor. Part III then emphasises the need to refocus the LAWS conversation within the purpose of the CCW.

____________________

Autonomy

As noted at the outset, a threshold challenge in discussing LAWS is that there are wildly varied understandings of what is meant by autonomy in general, let alone as applied to weapons systems. Given autonomy’s complex nature, this should not be surprising. But these different understandings set the conditions for a dialogue bordering on incoherence. So much so that it would be tremendous progress for the international community if there was a complete and utter lack of consensus regarding whether to develop and employ LAWS, but agreement as to what was meant by LAWS.3 But as of now, we cannot even agree on what we are discussing.

The United Nations Institute for Disarmament Research succinctly explained how the discussion of LAWS presently lacks focus, tacking between things (for example, drones, robots and systems), a characteristic (autonomy) and uses (defensive measures? Targeting? Kill Decisions?) in an inconsistent and often confusing way. One of the reasons there are so many different terms being proposed as the object of discussion is that some actors are trying to capture a mix of variables of concern (such as lethality or degrees of human control), while others are talking about more general categories of objects.4

A constructive LAWS dialogue requires a shared and coherent understanding of machine or system autonomy. Thus far, the international community has neither, and perhaps even worse, continues to engage in overly broad and conceptually confusing inquiries. In order to recognise the inherent futility of these inquiries, the discussion could draw from the U.S. militaries’ failed efforts in conceptualising and explaining autonomy.

In 2012, the U.S. Defense Science Board issued a report, “The Role of Autonomy in [Department of Defense] Sys-

tems”. As the report explains, what had been occurring was that different U.S. military services (Army, Navy, and Air Force) were “making significant investments of time and money to develop definitions of autonomy”, yet “[t]he competing definitions for autonomy have led to confusion”. And that confusion “may be contributing to fears of unbounded autonomy”. The end result was “a waste of both time and money spent debating and reconciling different terms” which were “irrelevant to the real problems”.

While the desire to define autonomy is natural and seems reasonable, such efforts will inevitably be counterproductive. The U.S. militaries’ definitional efforts “unsatisfactorily (...) tried to express autonomy as a widget or discrete component (...).” Attempts to develop ‘autonomy roadmaps’ based on trying to correlate certain levels and types of computer functions needed for a certain level of autonomy were equal parts well-intended and ill-advised.

6 Ibid; the report acknowledged that “the word ‘autonomy’ often conjures images in the press and the minds of some military leaders of computers making independent decisions and taking uncontrolled action”.
8 Ibid.
9 Ibid.
10 U.S. Department of Defense (n 4).
The concept of autonomy is neither usefully nor helpfully thought of in levels. This is because of, among other reasons, the dynamic nature of functions within a system. Many functions can be executed “concurrently as well as sequentially (…) and can have a different allocation scheme to the human or computer at a given time”.

Autonomy is better thought of across not one but several spectrums. And within each spectrum, the amount or quality of machine autonomy varies and changes as the system operates. As a result, plotting autonomy as a linear and single axis progressively and discretely demarcated by whether humans are in, on, or out of a functional loop both oversimplifies and misrepresents. It amounts to a form of conceptualising autonomy in levels, which, as discussed, is neither useful nor helpful beyond serving as a descriptor. Nor is there consensus on where and how to delineate in, on, and out of the loop.

Similarly, attempting to broadly differentiate machine functions as being either automatic, automated, or autonomous lacks practical utility. These terms may be used to understand as one of the spectrums through which we conceptualise autonomy, the complexity of the machine.

11 U.S. Department of Defense (n 4).
12 Three such spectrums are the nature of the human-machine command and control relationships, the complexity of the machine and the type of decision being automated. See P. Scharre, Between a Roomba and a Terminator: What is Autonomy, War on the Rocks, 18 February 2015, http://warontherocks.com/2015/02/between-a-roomba-and-a-terminator-what-is-autonomy/.
This spectrum ranges from the lower end, automatic, to automated, to autonomous at the higher end.\textsuperscript{13} But again, this is only one spectrum. And the utility of that one spectrum is limited, as there are no clear boundaries between automatic, automated, or autonomous,\textsuperscript{14} a challenge which CCW States Parties have already encountered.\textsuperscript{15}

Consider the following assessment of a household cleaning device, the Roomba robotic vacuum cleaner:

“The Roomba must navigate a house full of obstacles while ensuring that the carpet is cleaned. (…) The Roomba user provides high-level goals (vacuum the floor, but don’t vacuum here, vacuum at this time of day, etc.). The Roomba must make some choices itself (how to identify the room geometry, avoid obstacles, when to recharge its battery, etc.). The Roomba also has some automated behaviour and encounters situations it cannot resolve on its own (e.g., it gets stuck, it can’t clean its own brushes, etc.). Overall, the Roomba has marginal autonomy, and there are numerous situations it cannot deal with by itself. It is certainly not intelligent. However, it does have basic on-board diagnostic capability (“clean my brushes!”) and a strategy for

\textsuperscript{13} Ibid.
\textsuperscript{14} Ibid.
\textsuperscript{15} See above; additionally, the focus on autonomously performed functions works toward the effect that such a system would be the subject of the CCW discussion, while one which automatically selected and engaged targets, or was automated, would not.
vacuuming a room about whose size and layout it was initially ignorant.”

Where should the Roomba be placed within the human in, on, or out of the loop? Is the Roomba automated? Autonomous? Similarly, where within the loop should a household thermostat or microwave oven be placed? The answer, which further illustrates that we cannot draw system-wide conclusions from this spectrum, is that we do not know without having more information about the system. Certain thermostats and microwave ovens would likely be considered automated, others capable of sensing and adjusting their operation may be autonomous. But

---


18 Asaro distinguishes automated from autonomous on the grounds that unsupervised automated systems “involve repetitive, structured, routine operations without much feedback information (such as a dishwasher)”, while autonomous systems operate in “dynamic, unstructured, open environments based on feedback information from a variety of sensors (such as a self-driving car)”; P. Asaro, On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal
again the lines are fuzzy, and machine complexity is but one spectrum.

Ultimately,

“autonomy isn’t a discrete property of a work system, nor is it a particular kind of technology; it’s an idealized characterization of observed or anticipated interactions between the machine, the work to be accomplished, and the situation. To the degree that autonomy is actually realized in practice, it’s through the combination of these interactions”.19

Autonomy is better thought of as “a capability of the larger system enabled by the integration of human and machine abilities”.20 This approach recognizes that the operation of all machines requires some degree of human involvement. That means that there are no fully autonomous systems.21 The point of autonomy being bounded extends to weapons systems as well – there is no such thing as a fully autonomous weapon.22

---

19 Bradshaw (n 6); another way of thinking about autonomous machines in is terms of the extent of self-directedness and self-sufficiency.

20 U.S. Department of Defense (n 4).

21 Ibid.

22 As the President of the ICRC observed in 2011 (referring to ‘truly’ as opposed to ‘fully’ autonomous weapons), “such systems have not yet been weaponised. Their development represents a monumental pro-
Indeed, as one commentator observed, “the question of when we will get to ‘full autonomy’ is meaningless. There is not a single spectrum along which autonomy moves. (...) A better framework would be to ask what tasks are done by a person and which by a machine”.23 Utilising that framework would in turn better focus the inquiry on the critical functions of selecting and engaging targets.

**Lethal Autonomous Weapons Systems**

The LAWS conversation must acknowledge that weapons systems capable of selecting and engaging targets without further human intervention have existed for decades. Framing the issue in terms of fully autonomous systems is not such an acknowledgement, it is rather an attempt to avoid that underlying reality.24 The inescapable problem is that the concern for LAWS is grounded in systems capable of determining what to shoot or fire at and


23 See Scharre (n 12).

24 Framing the LAWS discussion in terms of future, full autonomy allows both States Parties which currently field what would otherwise qualify as LAWS, and civil society groups advocating States Parties to support a ban, to not talk about current weapons systems. The problem is that if the LAWS discussion is to focus on weapons systems capable of selecting and engaging targets without further human intervention, then current systems will have to be included.
then shooting or firing. And such systems already exist.\textsuperscript{25} Unless and until the international community can identify what it is beyond autonomy in the critical function of selection and engagement of targets that is troubling, then there is little advantage in framing the discussion in terms of full autonomy.

\textit{Examples of current LAWS}

In a report following a 2014 experts meeting, the International Committee of the Red Cross (ICRC) provided a sampling of existing weapon systems with autonomy in the critical functions of acquiring, tracking, selecting, and attacking targets:\textsuperscript{26}

- Patriot surface-to-air missile system; a missile defence system that automatically detects and tracks targets before firing interceptor missiles;\textsuperscript{27}

\textsuperscript{25} See Scharre/Horowitz (n 17), at Appendix b, (detailing that over thirty countries have employed fifteen different autonomous weapon systems dating as far back as 1980).

\textsuperscript{26} International Committee of the Red Cross, Expert Meeting: Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects, March 2014, \url{https://www.icrc.org/en/document/report-icrc-meeting-autonomous-weapon-systems-26-28-march-2014}. The vast majority of these weapons systems are designed to target material, aircraft, vessels at sea, and inbound missiles. Other estimates are that “[a]s many as 40 nations are currently developing military robotics”, and “some weapons already in use may be considered autonomous”; S. Gross, The U.S. Should Oppose the U.N.’s Attempt to Ban Autonomous Weapons, The Heritage Foundation, 5 March 2015.

\textsuperscript{27} Ibid., p. 67; introduced in the late 1970s, the Patriot is employed by at least sixteen countries.
• Aegis Weapon System; a ship-based system combining radar to automatically detect and track targets with various missile and gun systems; 28
• Phalanx Close-in Weapon System; a ship-based 20 mm gun system that autonomously detects, tracks, and attacks targets; 29
• The Goalkeeper Close-in Weapon System; “an autonomous and completely automatic weapon system for short-range defence of ships against highly manoeuvrable missiles, aircraft and fast manoeuvring surface vessels”; 30
• Counter Rocket, Artillery, and Mortar System; a land-based fixed weapons system that employs the same technology as the Phalanx Close-in Weapon System to target and attack rockets, artillery, and mortars; 31
• Iron Dome; a ground-based air defence system which automatically selects targets and fires interceptor missiles; 32
• NBC MANTIS (Modular, Automatic and Network-capable Targeting and Interception System); an

28 Ibid; introduced in 1978, Aegis is employed by at least five countries.
31 Ibid.
32 Ibid.
automated ground-based air defence system using 35 mm guns to automatically target rocket, artillery, and mortars.\textsuperscript{33}

The majority of these weapons systems are not new, to varying degrees they have been in use for decades. As a result, it is challenging to assemble a coherent contemporary argument as to why 1980s weapons systems employed with minimal issues are now problematic. Perhaps focusing on the looming future and the prospect of fully autonomous weapons is, superficially anyway, easier than trying to articulate a retrospective argument. But framing the LAWS discussion in terms of full autonomy needlessly distracts and confuses.

\textit{Distraction of fully autonomous weapons}

What is it about autonomy, above and beyond machines selecting and engaging targets without human intervention, that is worrisome, and is CCW the appropriate venue for those concerns? Recognising that full autonomy, if it is possible, would be unbounded, and depicting that, in two dimensions, it is challenging, consider full autonomy as the oval below, comprised of any number of functions (presumably infinite), depicted by the other ovals.

\textsuperscript{33} Ibid.
There is space, an autonomy ‘delta’/\(\Delta\), between full autonomy and autonomy in the function of selecting and engaging targets. Among other functions, this delta could be manifested by, for example, logistics trucks\(^{34}\) or aircraft\(^{35}\) capable of self-driving or flying, of recharging or reloading themselves, or performing internal diagnostic assessments and repairs.

To ground the LAWS debate in full autonomy requires articulating what is concerning about systems *beyond* the capability to select and engage targets without human in-

\(^{34}\) See L.M. Bacon, Unmanned Vehicles Heading for Battlefield, Army Times, 2 September 2013 (describing military plans for driverless trucks).

\(^{35}\) See J. Gould, Army Seeks to Cut Casualties with Unmanned Delivery Craft, Army Times, 8 December 2014; A. Mehta, Sikorsky Plans First Flight of Autonomous Black Hawk, Army Times, 14 May 2014 (describing military plans for unmanned aerial delivery systems, which would not be remotely piloted but autonomously perform flight functions).
tervention. Otherwise, as this section advocates, the discussion should focus on the critical functions of selecting and engaging targets.

Consider a hypothetical weapons system “A”, which performs all functions autonomously except selection and engagement of targets.

Now consider the inverse of A, hypothetical system “B”, in which humans perform virtually all the functions, but B, not a human operator, selects and engages targets.

Which of the functions the different weapons systems perform are concerning, and by what criteria? None of the functions A performs without human intervention potentially causes unnecessary or unjustifiable suffering to combatants, or affect civilians indiscriminately. And those functions the system performs are not a ‘specific weapon’. These non-weapons functions are a reminder that autonomy is merely a technology descriptor.

The point of introducing systems A and B is to underscore that while there may well be concerns about autonomous non-weapons functions (the self-driving, the self-loading, etc.), they are not most appropriately addressed within the CCW, a weapons treaty regime. Fully autonomous systems would constitute systems A+B. But what is it beyond autonomously performed weapons functions that is potentially concerning? Stated otherwise, why not limit

36 Otherwise, the argument could – and this section contends that it should – focus on weapons systems capable of selecting and engaging targets without further human intervention.
the LAWS discussion to those critical weapons system functions? Doing so makes the conversation much more manageable and in line with the CCW. As has been demonstrated, a dialogue where LAWS are framed in full autonomy tends to drift between science, including artificial intelligence and the concept of singularity, and science fiction, including robots somehow countermanding human-imposed operating constraints.

37 This paper does not mean to suggest that refocusing the CCW LAWS discussion on critical functions would yield clarity or rapid resolution, far from it. Indeed, there are a number of very challenging questions to be sorted within the context of critical functions – what it means to select, what it means to engage, and what and how we think of human control – that are quite daunting. Yet, consider that three years into the CCW LAWS discussions there has been no mention of what is meant by select or engage. But framing in terms of critical functions is much closer to, if not a bounded inquiry, and States Parties are more likely to at least agree on what it is that is being discussed. That by itself would constitute a sizable step forward from the status quo.

38 McCarthy, the professor who developed the term artificial intelligence [AI] in 1956, defined it as “science and engineering of making intelligent machines, especially intelligent computer programs”; see N. Bostrom, Superintelligence Paths, Dangers Strategies, 2014. But AI means different things to different researchers..

39 While it is unclear whether full autonomy could ever exist, some claim that it could or would change once ‘technological singularity’ is reached. And of course, like every other term or concept in this discussion, technological singularity means very different things to different people, as illustrated by this article, which provides 17 different definitions: N. Danaylov, 17 Definitions of the Technological Singularity, Institute for Ethics & Emerging Technologies, 22 August 2012, http://ieet.org/index.php/IEET/more/danaylov20120822.

40 This of course includes Asimov’s classic “I, Robot” short story series, which developed the three ‘laws of robotics’ and a host of creative
We do not know what full autonomy means or if it is even achievable, and if so, when. This is a challenging foundation upon which to engage in constructive dialogue. Moreover, why does autonomy other than in selection and engagement of targets matter, at least within the CCW construct?

**CCW**

International consensus is of course challenging to obtain in any area, but by framing LAWS in terms of full autonomy, the discussion undermines itself. Because of the inability to reconcile potential, future, fully autonomous systems with long-standing weapons systems capable of selecting and engaging targets without further human intervention, the dialogue thus far has been limited to general terms. But this lack of specificity has either set the conditions for, or allowed the discussion to become circular, depending on one’s individual perspective.

This is reflected in what has transpired at the CCW since the treaty body took on LAWS in 2013. In 2014, the majority of the debate was spent discussing how to define autonomy.\(^4\) Not only was an agreement on the definition not

---

reached, a majority of the delegations thought it premature to even attempt to find one.\textsuperscript{42} As discussed above, trying to precisely define autonomy does not seem possible. And the conversation was only further complicated by the idea that one indefinable concept – autonomy – might be conceived in reference to an almost equally indefinable concept, ‘meaningful human control’.

\textsuperscript{43} This trend proceeded in a CCW LAWS meeting in November 2015, where delegates agreed to continue discussions without resolving any substantive issues, or even reaching consensus as what those issues were. And, as already discussed, in the meetings in April 2016, Human Rights Watch insisted that the conversation was about future systems, while the ICRC reiterated that current systems would have to be included.\textsuperscript{44}

There is no reason to believe that this will change at the upcoming LAWS GGE, unless and until the manner changes in which LAWS are considered. The CCW provides a forum for States Parties to discuss certain conventional weapons which are deemed to be excessively injurious or

\textsuperscript{1}AFA515C1257CF30047A8C7/$file/Report_AdvancedVersion_10June.pdf.

\textsuperscript{42} Ibid, at para. 17, reporting that “[t]he issue of a definition was raised by a number of delegations”. While some suggested that a clarification would be necessary at a certain stage if more substantial work were to be undertaken, most of the delegations indicated that it was too early to engage in such a negotiation.

\textsuperscript{43} Ibid, at para. 20, reporting that “[m]any interventions stressed that the notion of meaningful human control could be useful to address the question of autonomy”.

\textsuperscript{44} See ICRC (note 3).
indiscriminate. Framing the LAWS discussion in terms of full autonomy confuses an already complicated area, and introduces non-weapons functions that are outside CCW’s weapons treaty purview. Most importantly, full autonomy distracts from CCW’s humanitarian focus, banning or restricting the use of specific types of weapons that cause unnecessary or unjustifiable suffering to combatants, or indiscriminately affect civilians.

**Conclusion**

Advances in autonomy will, and indeed already do, herald a technological revolution.\(^{45}\) Driverless cars, the prospect of package delivery via largely autonomous aerial delivery systems, and infinitely more applications are now being used or developed.

The LAWS debate is proving to be a proxy for broader technology \(^{46}\) and morality questions.\(^{47}\) And while those questions are fairly asked and debated, in terms of LAWS, the CCW discussions will not benefit, or will never pro-

---


gress, from yet another circular lap mired in trying to define the indefinable, and framed in terms of full autonomy.

The international community needs to leave the cul-de-sac into which the LAWS discussion has driven. Focusing on the critical function of selection and engagement of targets may offer one such possibility. But the most important step is to recognise the circular nature of the discussions to date, which, absent a change in how LAWS are considered, will only continue.