Article36

DISCUSSION PAPER | MAR 2020

Autonomy in weapons systems – considering approaches to regulation

A short briefing note - which was originally a background paper to a small group meeting of states, international organisations and civil society actors.

Article 36 is a UK-based not-for-profit organisation working to promote public scrutiny over the development and use of weapons.*

www.article36.org
info@article36.org
@Article36

* This paper was written by Richard Moyes.

This short paper aims to assist thinking about the relationship between regulatory options and our conceptualisation of the subject matter in discussions on autonomy in weapons systems. The purpose is to facilitate open thinking about different possible ways of regulating, recognising that identifying options requires some shared understanding of the subject matter in question.

Towards a shared understanding of the subject matter

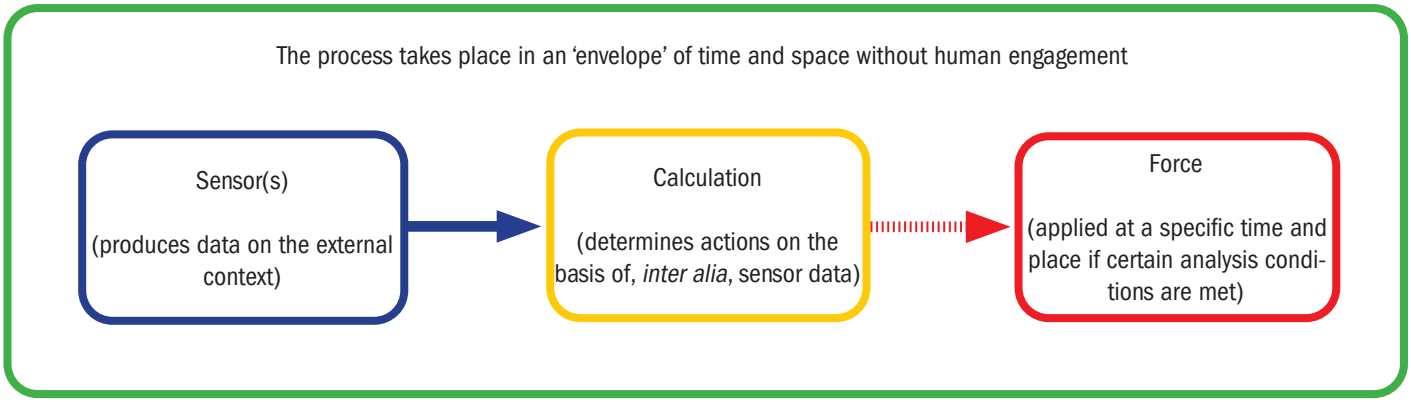
The international discussion on autonomy in weapons systems is rendered challenging by diverse starting assumptions about both the parameters of the subject matter and the types of regulation that are being sought. In particular, the adoption of the wording 'lethal autonomous weapons' (LAWS) as an organising term in the debate (although it remains undefined), and calls for a ban on 'fully autonomous weapons', both promote certain assumptions that the issue in question can be usefully approached with reference to a single distinct weapon category. The utility of such a 'single label' approach is however uncertain – and is certainly only one of a number of approaches that could be taken.

A process of functioning

As certain states within the CCW have noted, the categorisation of technologies must be seen in the relation to the purpose being served. 'Autonomy' in targeting is capable of presenting concerns in static land-based units, mobile sea-borne vehicles, swarms of aerial vehicles, or multi-element systems combining inter-linked sensors and weapon delivery capabilities on different platforms, etc. Such systems may be very different, in terms of physical construction, in their space of operation, and in their role from the strategic to the tactical – and thus they would not be labelled collectively for the purposes of planning operations, purchasing or deciding which warehouse to store them in. Such systems only share a common space in so far as they might utilise a certain *process of functioning*. It is this process of functioning, rather than the physical technology, that presents the central issues of concern in this discussion.

Sensor-calculation-force

All of the diverse systems that have been put on table as raising questions with respect to autonomy share a process of functioning whereby force is applied directly as a result of calculations based on certain sensor inputs, and without a human evaluation of those sensor inputs. More specifically it should be noted that the specific



time or location where those sensor inputs will be received is not set by a human or considered to be inevitable.[1] Different actors have asserted that this process may need to be more or less complex in order to qualify as 'the issue' with respect to autonomy in weapons systems. But if such arguments are put to one side, it is clear that this technical process is foundational in all cases.

The underpinning features for concerns about autonomy are a combination of both the sensor-calculation-force process and the human relationship to the 'envelope' of time and space within which that sensor-force process is functioning. Human control in this context is primarily dependent upon understanding the functioning of the sensor-calculation-force process and understanding the context within which that function will operate.

In so far as different actors may perceive the need for limits in relation to autonomy, or to assert requirements for human control, these need to find expression either as limits or requirements on *how the sensor-calculation-force process works*, or as limits or requirements regarding *the human relationship to the functioning of that process in a context*.

Target profiles

A particular component of the sensor-calculation-force process that merits consideration is the target profile, or target profiles, that a system uses. If force is to be applied on the basis of sensor inputs, there must be patterns of sensor data that will result, in some or all circumstances, in force being applied. Such patterns of sensor data are necessarily a 'proxy' or a 'representation' of an actual target, and they may be more or less precisely tailored to a specific target object (or phenomenon). For example, certain anti-vehicle warheads do not detect armoured vehicles, *per se*, but heat-shapes that may be emitted by such vehicles. Similarly, certain anti-ship torpedoes do not identify ships, but rather acoustic signatures that certain ships (or specific ships) will emit. Understanding the role of such target profiles, which may be more or less complex, is an important locus for thinking about the functioning of a sensor-calculation-force process, and for thinking about human operator understanding of such a process.

Key parameters

Building on the comments above, some key parameters that could be subject to management are as follows:

Sensor-calculation-force process

- What patterns of sensor data are considered a 'target' (target profile)?
- What objects or phenomena fall within the target profile(s)?
- How were the target profiles constructed?
- Can the target profile(s) change during an attack?
- What type of force does the system apply?
- How many applications of force can it undertake during an attack?

Human relationship to that process

- Can the target profile(s) be practically understood?
- When will the process start?
- How long will that process function for?
- Where will it function?
- What is known about the context of functioning – including objects or phenomena that will fall within the target profile, and any potentially adjacent collateral risks?

With the significance of autonomous functioning being highly dependent upon the parameters of time and space (context) within which that functioning is allowed to occur, it seems clear that effective management of the issue is not only a matter of technological features or capabilities, but also of human choices in the use of technologies.

Approaches to regulation

Whilst calls for a ban on "fully autonomous weapons" assert a clear line – suggesting that all of the unacceptable technologies can be grouped under a single label – the subject matter of autonomy in weapons systems implies that an effective response might need to encompass both technologies and human actions. Calls for ensuring "meaningful human control" (or its terminological variants) create space for such different approaches.

Existing structures for regulating weapons technologies provide sufficient examples of how such layered regulation can be articulated.

For example, the broad category of (land)'mines' is subject to regulation under a number of instruments. Taken together we can see that certain types are prohibited (based on their 'anti-personnel' target profile under the 1997 Mine Ban Treaty), whereas others ('mines other than anti-personnel mines' - MOTAPM) are subject to requirements relating to their technical functioning, and obligations regarding the location and duration that they remain operational (in CCW Amended Protocol II).[2]

Oftentimes the form of regulation is conditional upon the level of categorisation being taken as the starting point for analysis. Building on the same example, the use, production, stockpiling etc. of anti-personnel mines are prohibited (for parties to the MBT), but landmines as a whole (a higher level of categorisation) are subject to various regulations – including certain prohibitions, technical restrictions and positive obligations.

These examples are simply to highlight, in approaching the issue of autonomy in weapons systems, that different forms of regulation can coexist, applying different obligations in relation to different variations that share the same basic process of functioning.

Where are the regulatory proposals?

Turning to the politics of the current CCW discussion, there is a striking absence of detailed proposals regarding what might be done. Past CCW discussion on other themes have often seen individual states, or sometimes small groups of states, proposing specific responses related to the weapon issue in question. Such proposals do not have to be comprehensive – but can address one aspect of the issue that is of particular concern. For example, in discussion on antivehicle mines in the past, Germany highlighted concerns regarding 'sensitive fuzes'. They were not asserting that sensitive fuzes were the only matter that should be considered, but they contributed analy-

sis of the issues in that particular area and suggestions as to what could be done. With input from different delegations a typology of issues – sensitive fuzes, limitations on active life, etc. – came to form a recognised landscape that tied particular 'problems' to possible regulatory responses.

Recent work in the CCW has tended to prioritise collective work on the 'Guiding Principles'. This has the benefit of contributing to a sense of collective action and engagement – but it can also serve to perpetuate an avoidance of key issues, including the fundamental conceptual foundations, in favour of the consensus. It might also be considered that certain delegations would resist agreement on conceptual foundations out of recognition that such an agreement would have political significance.

In 2020 and 2021 the credibility of multilateral responses to the issues raised by autonomy in weapons systems will likely hinge on their ability to build a shared conceptual framework and agreement on the key problematic issues within that framework. To that end, focused proposals by individual states are likely needed in order to spur discussion.

END NOTES

1 After all, an artillery shell contains a sensor that determines when detonation should occur – but the conditions that will initiate that sensor input are rendered effectively inevitable by it being fired into the air and both the location and time at which that input is expected to occur is established by a human operator.

2 An 'anti-personnel mine' is defined as a type of mine and is distinguished by what is intended to trigger, via its sensors, an application of explosive force. In addition to the prohibitions on anti-personnel mines in the Mine Ban Treaty, the CCW APII applies general and specific obligations to other types of mine (MOTAPM), including inter alia: removing them from active functioning under certain conditions (3.2); a prohibition on target profiles based on the magnetic influence of mine detectors (3.5); loose obligations to limit the life of certain types of MOTAPM, and to document where they have been used (6.1,3).