

Defining the Scope of Autonomy

Issues for the Campaign to Stop Killer Robots

The Campaign to Stop Killer Robots was launched in April 2013 with the objective of achieving a ban on the development, production and deployment of lethal autonomous weapons. In May 2014, the issue will be discussed by a UN expert meeting under the auspices of the Convention on Conventional Weapons (CCW) in Geneva. At this stage, it is inevitable that there will be much debate and discussion over the scope and meaning of any future prohibition. The Campaign is still being shaped, and what will be necessary for its success is that over the next few years a group of states and governments coalesce around a shared understanding of the problem and its solutions.

This policy brief¹ provides a pragmatic theory of humanitarian disarmament and discusses a set of key challenges related to the question of *scope* – that is, what exactly would be covered by a possible international agreement on lethal autonomous robots. The Campaign must tackle these issues if it is to move from a startup phase and into activism with sustained political impact. Attention is given to the Brimstone, a UK anti-tank weapon and one of the most advanced ‘fire and forget’ missiles in use, whose degree of autonomy is also a matter of some contention.

Nicholas Marsh

Peace Research Institute Oslo (PRIO)

A Pragmatic Theory of Successful Humanitarian Disarmament

A humanitarian disarmament treaty is designed to reduce human suffering rather than to manage the affairs of great powers. In the experience of this author, two things are required for a successful campaign for a humanitarian disarmament treaty.

- First, the campaign needs to stigmatize the type of weapon concerned by showing that it causes an unacceptable level of harm. Doing so involves highlighting aspects such as the propensity of the weapon to cause death and injury among civilians. A key element of the campaigns to ban anti-personnel land mines or cluster munitions was the message that these weapons remain lethal long after the fighting has ended, and so pose a deadly risk to civilians.
- The second task is to persuade governments that they do not *need* the weapon in question. States are being asked to voluntarily give up the use of a weapon, and usually prohibitive treaties also cover development, production, stockpiling and trade. For a government to abjure a weapon it needs to first accept that the weapon is not militarily essential.

A key element of the campaigns to ban cluster munitions and landmines involved spreading the message that these weapons were not essential – that alternatives existed, the role envisioned for them no longer existed, or that their use is so abhorrent as to be counterproductive. A counter example is the case of nuclear weapons. No sensible person doubts their devastating effects upon civilians. Nonetheless, nuclear weapons remain in state arsenals because the nuclear states still view them as being necessary. Nuclear disarmament will therefore not occur until the governments of those states are persuaded that such weapons cannot be used under any circumstances; that they are considered militarily *useless*.²

The Campaign to Stop Killer Robots needs to strike a balance between a prohibition with a wide ranging scope and a pragmatic accommodation of the interests of likely State supporters of the Campaign (who need to be reassured that they won't have to give up something they perceive to be militarily essen-

tial).

If it is not possible to persuade states of the necessity of a prohibition, the Campaign will most likely not find support from the states that will be required to form a coalition and negotiate a successful treaty.

Defining the Threshold of Autonomy

One of the most contentious issues is likely to concern the threshold at which a weapon system is deemed to be 'fully autonomous'. Discussions on this issue will necessarily be both technical and intensely pragmatic. The threshold that is set would determine which systems are banned and which are allowed to continue in operation.

Setting the threshold of autonomy is going to involve significant debate because machine decisionmaking exists on a continuum. There is a grey area in which systems can be said to be *partly* autonomous, and there is no consensus as to exactly what counts as a *fully* autonomous machine. A key task for the Campaign will be to create consensus on this issue among both NGOs and the states that would have to negotiate and then implement a ban.

Scholar Noel Sharkey has defined a lethal fully autonomous robot as one that operates in an open and unstructured environment; receives information from sensors; and processes the information in order to move, select targets and fire – all without human supervision.³ The Human Rights Watch report *Losing Humanity: The Case Against Killer Robots* similarly defines a fully autonomous robot as being one that is 'capable of selecting targets and delivering force without any human input or interaction'.⁴ Later, the report refers to a fully autonomous weapon that would identify targets and trigger itself, and also to Sharkey's definition noted above.⁵

Likewise, Christof Heyns, the UN Special Rapporteur on extrajudicial, summary or arbitrary executions, considered the technology and defined *lethal autonomous robots* as being 'weapon systems that, once activated, can select and engage targets without further intervention by a human operator. The important element is that the robot has an autonomous "choice" regarding selection of a target and the use of lethal force'.⁶

The emphasis in all these definitions of a

lethal fully autonomous robot is the decisionmaking capability of the machine – that it is able to navigate around its environment and discriminate between targets.

The above definitions used by the Campaign, Noel Sharkey and Christof Heyns coalesce around a definition of machine which is capable of lethal force, and can autonomously move around the battlefield and select a target. However, this is not a universally shared understanding of the essential characteristics of a lethal autonomous weapon. The UK Ministry of Defence defines an 'autonomous system' in the following manner:

*An autonomous system is capable of understanding higher level intent and direction. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about a desired state.... Autonomous systems will, in effect, be self-aware and their response to inputs indistinguishable from, or even superior to, that of a manned aircraft. As such, they must be capable of achieving the same level of situational understanding as a human.*⁷

Sharkey criticizes the UK Ministry of Defence definition⁸ on the grounds that not only does the technology to enable human-like cognition not exist, but it is very unlikely to exist in the foreseeable future. The point for Sharkey is that full autonomy occurs when machines have sole responsibility for selecting targets and using weapons – which could happen in machines with decisionmaking capabilities far below levels of human cognition. Using the above definition, the UK Ministry of Defence could employ machines with the ability to move, locate and select targets, and then fire at them, without defining them as 'autonomous' if they lacked a particular aspect of human cognition. The next section examines how these definitional disputes apply to specific weapon systems.

The Case of 'Fire and Forget' Weapons: The Brimstone

This section highlights the ambiguous status of a particular weapon, the Brimstone anti-tank missile, one of a class of 'fire and forget' weapons that includes missiles with autonomous capabilities.

The Brimstone is a 'fire and forget' weapon system – one of a range of weapons that are designed to be used precisely as that term

suggests. Most are missiles that guide themselves to their target once launched, allowing the firer to concentrate on other tasks, such as escaping enemy retaliation. Some simply fly to preset coordinates, while others actively look for targets.

The Brimstone, a UK-built anti-tank weapon, is currently one of the most advanced ‘fire and forget’ missiles. Development started in the 1990s, and the system has been in service since 2005. A description of it provided by the UK’s Royal Air Force states that:

Brimstone is a fully autonomous, fire-and-forget, anti-armour weapon, effective against all known and projected armoured threats.... During the search phase of the engagement, Brimstone’s [radar] seeker searches for targets in its path, comparing them to a known target signature in its memory. The missile automatically rejects returns which do not match (such as cars, buses, buildings) and continues searching and comparing until it identifies a valid target. The missiles can be programmed not to search for targets until they reach a given point, allowing them to safely overfly friendly forces, or only to accept targets in a designated box area, thus avoiding collateral damage.⁹

These capabilities appear to place the Brimstone within the above-mentioned definition of an autonomous lethal robot as defined by the Campaign to Stop Killer Robots, as well as by Noel Sharkey and Christof Heyns. It would not fall squarely within the UK Ministry of Defence’s definition, however, as while the Brimstone can distinguish between potential targets, it is unable to make human-like assessments of intent. The UK Ministry of Defence position document on autonomous systems, however, does mention that the Brimstone is an example of a weapon that ‘probably’ achieves a degree of autonomous operation in certain environments.¹⁰ The specific question of whether the Brimstone is a fully autonomous weapon is also considered briefly in an article by the UK Royal Aeronautical Society’s Air Power Group. This reiterates that the Brimstone does not meet the UK armed forces definition of an autonomous system, but does not comment on the wider debate.¹¹

The operational use of the Brimstone actually highlights the very issues raised by the Campaign. Originally designed for the Cold War,

its ability to autonomously select targets was ill-suited to contemporary operations – especially Afghanistan. There, because of the conflict’s complex nature, rules of engagement required that a human monitor the engagement right up until impact of the missile. A ‘fire and forget’ missile was inappropriate in such an environment.¹² In response, the UK Ministry of Defence commissioned an urgent update so that the missile could also be guided with a laser seeker that would guide the missile to a target directly selected by a human operator. The ability of the Brimstone to autonomously select targets with its radar, however, has been used. For example, the first use of the autonomous radar seeker was reported on 16 September 2011: over Libya, a salvo of 24 missiles using radar guidance was used to destroy seven or eight vehicles travelling in a convoy.¹³

The Brimstone is not unique. There are tens of different types of ‘fire and forget’ missiles in existence. Many just have the capability to fly to preset coordinates and thus do not select targets without human intervention. Nevertheless, the Brimstone is not the only missile to have an autonomous target-selection capability. In particular, some missiles used to attack shipping actively search and identify targets. The missiles frequently have ranges of over 100 kilometres, and thus enemy ships will have moved during the period between launch and the missile reaching the target area.

There is not sufficient space in this brief to examine in detail the exact nature of the technology. However, I will highlight two claims made by the manufacturers of such weapons that indicate the development of autonomous target selection. The Norwegian Joint Strike Missile is advertised by its manufacturer the Kongsberg Group as having ‘Autonomous Target Recognition with identification of targets to ship class level’ and also employs a library of potential targets.¹⁴ Similarly, the company MBDA advertises its MM40 BLOCK 3 Exocet missile by stating ‘terminal guidance relies on a sophisticated J-band active seeker to discriminate and select targets at sea’.¹⁵

Issues for Campaigners and Policy-makers

The Brimstone and other ‘fire and forget’ weapons with similar capabilities are clearly

relevant to the current debate prompted by the Campaign to Stop Killer Robots. Five areas of particular concern are presented below:

- First, at the technical level, is the detailed consideration of exactly what counts as lethal and fully autonomous. The status of ‘fire and forget’ weapons has been included in [media articles](#) and commentary on the issue of autonomous lethal weapons.¹⁶ This policy brief argues that the Brimstone and other weapons cross the definitional threshold provided by the Campaign and other commentators such as Noel Sharkey and Christof Heyns.
 - Second, a narrow definition of autonomy – such as the one used by the UK Ministry of Defence, which specifies ‘human like capabilities’ – would exclude not only all current technology, but likely all machine decisionmaking in existence during the foreseeable future. One potential hurdle for the Campaign is likely to be states that support the idea of a ban, but set the threshold for autonomy so high that it will not affect any of the robot systems they wish to deploy.
 - Third, a governmental preference for an excessively high threshold of autonomy raises a paradox. A key motivation for the Campaign is that robots cannot make the ethical and contextual assessments that humans can. However, a robot with a decisionmaking capability as advanced as human cognition could presumably make such judgements.
 - Fourth, it is one thing for the Campaign to attempt to preemptively stop the development of lethal autonomous weapons that do not currently exist: it is hard to argue that a weapon is military essential when it has not yet even been developed. It will be much more difficult, though not impossible, to get states to give up arms that are already deployed. At the very least, if the Brimstone and other ‘fire and forget’ missiles are assumed to fall under the definition used by the Campaign, then considerable effort may need to be expended on explaining *why* they are not to be considered militarily essential. For pragmatic reasons, it might be possible to negotiate that the scope of an international ban should exclude certain systems even if they cross the threshold of full autonomy. But, even so, creating such a loophole would be a difficult decision.
- Finally, the focus of the Campaign to Stop Killer Robots has been very much on weapons

platforms – air- or ground-based robots such as unmanned aerial vehicles – that if fully autonomous could decide to fire weapons. The reality of current autonomous lethal systems is much more prosaic. The Brimstone and other ‘fire and forget’ missiles do not look like robots as popularly conceived, even if they meet the definition provided by the Campaign, Noel Sharkey and Christof Heyns. This mirrors the general development of automation. In decades gone by, science-fiction authors conceived of intelligent humanoid robots that would mimic the activities of humans. In fact, technology has developed very differently. For example, instead of having one general-purpose robot that would take over people’s everyday tasks, there are now numerous examples of machines that use autonomous decisionmaking to fulfil specific limited tasks: for example, a **smartphone** that accepts spoken instructions and recommends music; **robots** that will clean a room without direct human control; and **cars** equipped with sensors that detect an incoming cyclist or pedestrian and automatically brake to prevent a collision. The battlefield is following a similar trajectory. Instead of Terminator-like humanoid machines, the lethal autonomous robots of today, and most likely the future, are those in which sensors and artificial decisionmaking are applied to a specific and limited role. The challenge for the Campaign will be to capture the public imagination in order to build a sufficiently large coalition able to prohibit such a heterogeneous array of lethal autonomous weapons. ■

29 January 2014. The author thanks the other members of the panel – Christian Holmboe Ruge, Alexander Harang and Simon O’Connor – and members of the audience for their insightful comments. Kristin Bergtora Sandvik provided very astute feedback on a draft. The author also appreciates the assistance provided by Maral Mirshahi in editing the brief prior to publication.

² Lewis, Jeffrey (2013) *Poison Control – Why Can’t We Get Rid of Nukes the Same Way We Got Rid of Chemical Weapons?* *Foreign Policy*, 14 October. URL: http://www.foreignpolicy.com/articles/2013/10/14/poison_control_nuclear_chemical_weapons (accessed 27 March 2014).

³ Sharkey, Noel (2012) *Automating Warfare: Lessons Learned from the Drones*. *Journal of Law, Information & Science* 21(2): p. 2.

⁴ Human Rights Watch (2012) *Losing Humanity: The Case Against Killer Robots*, p.2.

⁵ *Ibid*, p.6.

⁶ Heyns, Christof (2013) *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions*. A/HRC/23/47. New York: United Nations, pp. 6–7.

⁷ UK Ministry of Defence (2011) *The UK Approach to Unmanned Aircraft Systems*, pp. 2-2 to 2-3.

⁸ Sharkey, Noel (2012) *Automating Warfare: Lessons Learned from the Drones*. *Journal of Law, Information & Science* 21(2): pp. 2-3.

⁹ Handy, Brian (ed.) (2007) *Royal Air Force Aircraft & Weapons*. DCC(RAF) Publications, p. 87. URL: http://www.raf.mod.uk/rafcms/mediafiles/0186cc2a_1143_ec82_2ef2bfff37857da.pdf (accessed 5 March 2014).

¹⁰ UK Ministry of Defence (2011) *Joint Doctrine 2/1/1: The UK Approach to Unmanned Aircraft Systems*, p. 5-4. URL: <http://dronewarsuk.files.wordpress.com/2011/04/uk-approach-to-uav.pdf> (accessed 27 March 2014)

¹¹ Aerospace (2013) ‘The ‘drone’ debate the ethics of armed unmanned aircraft’. November issue, pp. 14-19. URL: http://aerosociety.com/Assets/Docs/Publications/The_Ethics_of_Armed_UAS.pdf (accessed 4 April 2014).

¹² See Markowitz, Mike and John D. Gresham (2012) *Dual-Mode Brimstone Missile Proves Itself in Combat*. *Defense Media Network*, 26 April. URL: <http://www.defensemedianetwork.com/stories/dual-mode-brimstone-missile-proves-itself-in-combat/> (accessed 2 February 2014).

¹³ UK Ministry of Defence (2011) *RAF Conducts Precision Strikes over Libya*. Press Release, 16 September. URL: <https://www.gov.uk/government/news/raf-conducts-precision-strikes-over-libya> (accessed 2 February 2014).

¹⁴ Kongsberg Gruppen (2013) *Kongsberg and Boeing Complete Joint Strike Missile (JSM) Check on F/A-18 Super Hornet*. Press Release, 29 November. URL: <http://www.kongsberg.com/en/kog/news/2013/november/kongsberg-and-boeing-complete-joint-strike-missile/> (accessed 2 February 2014).

¹⁵ MBDA Missile Systems (2014) *EXOCET MM40 – New Generation Long Range Anti-Ship Missile System*. URL: http://www.mbda-systems.com/mediagallery/files/exocet_mm40_ds.pdf (accessed 5 March 2014).

¹⁶ *Wired* (2008) *Robot Wars Continues*. 31 March. URL: <http://www.wired.com/dangerroom/2008/03/the-robot-war-c/> (accessed 5 March 2014).

Notes

¹ This PRIO Policy Brief is based upon remarks made by Nicholas Marsh at a seminar on autonomous lethal robots organized by the Norwegian Centre for Humanitarian Studies, which took place at PRIO on

THE AUTHOR

Nicholas Marsh is a Researcher at PRIO and since 2001 has worked there on the trade, production, use and governance of weapons. His current research interests include the use of weapons by non-state parties involved in civil war, and the international trade in small arms and light weapons. Email: nic@prio.org

THE PROJECT

This policy brief is a product of the Critical Humanitarian Technology Project. The project examines the influence of new technologies upon humanitarian action, and aims to inform the development of a Norwegian policy on humanitarian technology. For more information about the project, please visit: www.humanitarianstudies.no.

PRIO

The Peace Research Institute Oslo (PRIO) is a non-profit peace research institute (established in 1959) whose overarching purpose is to conduct research on the conditions for peaceful relations between states, groups and people. The institute is independent, international and interdisciplinary, and explores issues related to all facets of peace and conflict.