



MEANINGFUL HUMAN CONTROL *in* WEAPON SYSTEMS: *A Primer*

By: Michael C. Horowitz and Paul Scharre

ABOUT CNAS WORKING PAPERS:

Working Papers are designed to enable CNAS analysts to either engage a broader community-of-interest by disseminating preliminary research findings and policy ideas in advance of a project's final report, or to highlight the work of an ongoing project that has the potential to make an immediate impact on a critical and time-sensitive issue.



PREFACE

Information technology is driving rapid increases in the autonomous capabilities of unmanned systems, from self-driving cars to factory robots, and increasingly autonomous unmanned systems will play a significant role in future conflicts as well. “Drones” have garnered headline attention because of the manner of their use, but drones are in fact remotely piloted by a human with relatively little automation and with a person in control of any weapons use at all times. As future military systems incorporate greater autonomy, however, the way in which that autonomy is incorporated into weapon systems raises challenging legal, moral, ethical, policy, and strategic stability issues.

Nation-states and non-governmental experts in the United States and abroad are already debating how advances in autonomy will influence warfare – and what, if anything, should be done. A group of over 50 non-governmental organizations (NGOs) have launched a “Campaign to Stop Killer Robots,” calling for a ban on autonomous weapons. In May of 2014, state parties to the United Nations Convention on Certain Conventional Weapons held initial discussions on autonomous weapons, and another round of discussions will be held in April 2015.¹

Governments and militaries are only beginning to grapple with how to address the challenges and opportunities associated with increased autonomy. Technology is moving fast in this area. Few states have guidelines on how autonomy should be included in future weapon systems, with the United States a notable exception.²

The Center for a New American Security (CNAS) has launched a project on Ethical Autonomy, which will examine the legal, moral, ethical, policy and strategic stability dimensions of increased autonomy in future weapon systems. The goal of the Ethical Autonomy project at CNAS is to help states, experts, advocates, academics, and militaries grapple with the challenging issues of autonomy in future weapons. This dialogue is necessary to ensure an appropriate balance between ethical and strategic stability considerations, technological opportunities and future warfighting needs.

The Ethical Autonomy project will result in a series of working papers examining various issues associated with autonomy in weapons, which will ultimately culminate in a final report in late 2015. This second working paper revolves around the issue of meaningful human control, a phrase coined by some proponents of a ban on autonomous weapon systems to describe the importance of having human control over the use of weapons. It will be followed by subsequent papers that will delve deeper into specific legal, moral, ethical, policy, or strategic stability issues raised by autonomous weapons.

About the Authors

Michael C. Horowitz is an Adjunct Senior Fellow at CNAS and an Associate Professor of Political Science at the University of Pennsylvania.

Paul Scharre is a Fellow and Director of the 20YY Warfare Initiative at CNAS.

The Ethical Autonomy project is a joint endeavor of CNAS’ Technology and National Security Program and the 20YY Warfare Initiative, and is made possible by the generous support of the John D. and Catherine T. MacArthur Foundation.

1. The United Nations Office at Geneva, “The Convention on Certain Conventional Weapons,” [www.unog.ch/80256EE600585943/\(httpPages\)/4F0DEF093B4860B4C1257180004B1B30?OpenDocument](http://www.unog.ch/80256EE600585943/(httpPages)/4F0DEF093B4860B4C1257180004B1B30?OpenDocument); The United Nations Office at Geneva, “The Convention on Certain Conventional Weapons: Latest Information,” <http://www.unog.ch/80256EE600585943/%28httpPages%29/3CFCEEEF52D553D5C1257B0300473B77?OpenDocument>.

2. United States Department of Defense, *Autonomy in Weapon Systems*, 3000.09 (November 21, 2012), www.dtic.mil/whs/directives/corres/pdf/300009p.pdf.

EXECUTIVE SUMMARY

In discussions on lethal autonomous weapon systems at the United Nations (UN) Convention on Certain Conventional Weapons in May 2014, “meaningful human control emerged” as a major theme. Many who support a ban on autonomous weapon systems have proposed the requirement of meaningful human control as one that ought to apply to all weapons, believing that this is a bar that autonomous weapons are unlikely to meet.

In particular, one issue motivating the focus on meaningful human control is concern about a potential “accountability gap.” The concern is that if a future weapon incorporating autonomy lacked meaningful human control, then if there was a failure that led to it striking the wrong targets, it is possible that no human would be held accountable for those engagements.

To investigate this topic, we assessed statements made by those advocating for meaningful human control and studied the use of weapons today, which have varying degrees of autonomy.³ The results suggest that meaningful human control has three essential components:

1. Human operators are making informed, conscious decisions about the use of weapons.
2. 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.
3. The weapon is designed and tested, and human operators are properly trained, to ensure effective control over the use of the weapon.

These standards of meaningful human control would help ensure that commanders were making conscious decisions, and had enough information when making those decisions to remain legally accountable for their actions. This also allows them to use weapons in a way that ensures moral responsibility for their actions. Furthermore, appropriate design and testing of the weapon, along with proper training for human operators, helps ensure that weapons are controllable and do not pose unacceptable risk.

3. We would like to thank the many individuals who contributing to our understanding of the issue of meaningful human control, including those from various governments, non-governmental organizations, and academics. We especially would like to thank those who participated in CNAS working groups on this and other topics relating to autonomous weapons. We have benefited greatly from these discussions with individuals representing a diverse array of viewpoints. While we have drawn upon their insights, this paper reflects our views and should not be construed in any way as characterizing the conclusions of those discussions or reflecting the opinions of any working group participants. Any errors in fact or judgment are ours alone.

INTRODUCTION

Ensuring appropriate human control over weapons, such that they are used effectively and as intended, has been a critical task for militaries around the world since the invention of the bow and arrow, or perhaps even the spear. Projectiles can miss their targets. Even hand-held edged weapons that are large and unwieldy run the risk of accidents in the hands of an untrained user. Many weapons, from slings to swords to bows and arrows, require skill and experience to be used safely and effectively. As humankind has developed increasingly powerful weapons over time, concern has grown as well about ensuring that weapons are used in an appropriate fashion. Nuclear weapons, for example, require stringent safeguards to prevent unauthorized use.

Future weapon systems that incorporate increased autonomy require careful consideration to ensure that humans remain in control of the use of those weapons. Autonomy, by its very nature, entails programming machines to perform some tasks or functions that would ordinarily be performed by humans. At the same time, humans must remain in control of deciding whether and how those weapons should be used. How should that control be exercised? What principles apply to thinking about autonomy and human control in the use of weapons?

The laws of armed conflict, or “laws of war,” lay out a number of principles and rules that apply to the use of weapons against enemies during armed conflict. These include the principle of military necessity; the principle of distinction, which requires a party conducting an attack to distinguish between civilians and combatants; the proportionality rule when conducting attacks; the rule against employing weapons calculated to cause unnecessary suffering to combatants; and the rule about taking precautions to minimize civilian casualties when conducting attacks, among others.⁴ These principles and rules apply to the use of all weapons against enemies during armed conflict, including the use of autonomous weapons.⁵

Many who support a ban on autonomous weapon systems are concerned that the existing laws of war are insufficient to account for the unique challenges posed by autonomous weapon systems, and have proposed an additional requirement, one of “meaningful human control,” that they argue ought to apply to weapons incorporating increased autonomy.⁶ Drawing on general principles of the law of armed conflict, they argue that there is already an implicit requirement for meaningful human judgment in every individual decision

4. Other rules and principles also govern the use of force against an enemy, but these are some of the most important legal principles. For more on the applicability of the law of armed conflict, or international humanitarian law, to autonomous weapons, see Kenneth Anderson, Daniel Reisner, and Matthew C. Waxman, “Adapting the Law of Armed Conflict to Autonomous Weapon Systems,” *International Law Studies*, 90 (September 2014), papers.ssrn.com/sol3/papers.cfm?abstract_id=2477095; and Kenneth Anderson and Matthew C. Waxman, “Law and Ethics for Autonomous Weapons: Why a Ban Won’t Work and How the Laws of War Can,” *Columbia University Public Law & Legal Theory Research Paper Series*, Research Paper No. 13-351, April 2013, papers.ssrn.com/sol3/papers.cfm?abstract_id=2250126.

5. For ease of reading, we will use the shorthand “autonomous weapons” throughout this paper instead of the more formal “autonomous weapon system.” It is more accurate to think of “weapon systems,” however, that comprise not only the munition itself, but also the launching platform, sensor, targeting process and any other elements such as communications links that are necessary for an engagement to occur, some of which may be distributed across multiple physical platforms. In this paper, “autonomous weapon” and “autonomous weapon system” should be treated as interchangeable, and it should be understood that we are referring to not only the munition but the entirety of the weapon system.

6. For example, see Bonnie Docherty, “Taking on Killer Robots,” *Justsecurity.org*, May 23, 2014, <http://justsecurity.org/10732/guest-post-killer-robots/>; Jonathan Fowler, “UN talks to tackle potential killer machines,” *Times of Israel*, May 13, 2014, <http://www.timesofisrael.com/un-talks-to-tackle-potential-killer-machines/>; Steve Goose, “Statement To the Convention on Conventional Weapons Informal Meeting of Experts on Lethal Autonomous Weapons Systems” (Human Rights Watch), <http://www.hrw.org/news/2014/05/13/statement-convention-conventional-weapons-informal-meeting-experts-lethal-autonomous>; Article 36, “Killer Robots: UK Government Policy on Fully Autonomous Weapons,” April 2013, http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf; Article 36, “Memorandum for delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS),” <http://www.article36.org/wp-content/uploads/2014/05/A36-CCW-May-2014.pdf>.

to use lethal force.⁷ They argue that the potential for increased autonomy to diminish human control over the use of force in the future necessitates making this requirement explicit. In particular, one of the motivations behind a principle of “meaningful human control” is the concern that greater autonomy in weapons could lead to an “accountability gap,” where so many key tasks have been programmed to be performed by a machine that, if there were an incident resulting in excessive civilian deaths, no human would have responsibility and be held accountable.

In discussions on lethal autonomous weapon systems at the United Nations (UN) Convention on Certain Conventional Weapons in May 2014, meaningful human control emerged as a major theme.⁸ This raises the question, however, of what should constitute meaningful human control. There is no clear definition or agreement at this point, although, as the UN Institute for Disarmament Research points out, “the idea of Meaningful Human Control is intuitively appealing even if the concept is not precisely defined.”⁹ Without a clear definition, however, “meaningful human control” risks being only a pleasant-sounding catchphrase. At best, it merely shifts the debate to, “what is meaningful?” It also risks appearing to resolve the issues raised by increased autonomy in weapons, when in reality it becomes an empty platitude, and one that would be devoid of a common meaning. At worst, a failure to define the term clearly could, if embedded in international discussions, lead to flawed policy choices.

Establishing a clear understanding of what is meant by meaningful human control is therefore an important task. A discussion about the phrase could serve as a starting point for understanding the difference between systems that could be considered under an appropriate level of human control and those where humans lack appropriate control. It is also necessary for a more effective dialogue between proponents of a ban, governments, and industry as the development and use of more advanced weapon systems, whether autonomous weapons or not, moves forward in parallel to these discussions.

Any usable concept of meaningful human control must account for the varied ways in which weapons have been used to date, however, and must address what is potentially new about greater autonomy in weapon systems. For example, at present, a fighter pilot in an engagement might have only seconds to decide whether to fire a missile at an enemy fighter. When that engagement occurs at beyond-visual-range, that pilot has meaningful human control even though the pilot makes the decision to fire entirely based on information received from sensors and computer processors – machines – and computers then guide the missile onto the target.

In what follows, we attempt to contribute to the discussion about meaningful human control by laying out the potential goals of defining meaningful human control, the context in which meaningful human control occurs in the use of weapons systems today, and then the key components of a successful definition of meaningful human control.

This paper is part of a continuing dialogue and is not intended as a final answer to the question of

7. Peter Asaro, “On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making,” *International Review of the Red Cross*, 94 no. 886 (Summer 2012), 687-709, www.icrc.org/eng/assets/files/review/2012/irrc-886-asaro.pdf.

8. Sarah Knuckey, “Governments Conclude First (Ever) Debate on Autonomous Weapons: What Happened and What’s Next,” *Justsecurity.org*, May 16, 2014, <http://justsecurity.org/10518/autonomous-weapons-intergovernmental-meeting/>.

9. United Nations Institute for Disarmament Research, “Considering how Meaningful Human Control might move the discussion forward,” *The Weaponization of Increasingly Autonomous Technologies*, No. 2, 2014, <http://www.unidir.org/files/publications/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf>, 2.

meaningful human control, but rather yet as another step in understanding the concept. Fundamentally, meaningful human control over the use of weapon systems should focus on helping ensure that weapon systems are used in ways that comply with international law. Keeping this goal in mind is critical for productive discussions moving forward.

WHY REQUIRE MEANINGFUL HUMAN CONTROL?

What do we want to get out of the concept of meaningful human control? Or, to put it another way, what is gained by adding the notion of meaningful human control to existing concepts under the laws of armed conflict for ensuring that weapons are used appropriately?

While many discussions of autonomous weapons include references to human control, including the 2012 U.S. Department of Defense Directive on autonomy in weapons, the phrase “meaningful human control” was first used by the non-governmental organization (NGO) Article 36 in a 2013 report on how the United Kingdom is thinking about autonomous weapon systems.¹⁰ Since then, however, it has been used by a wide range of commentators. While many of those writing on the concept of meaningful human control emphasize different ideas, there are two general schools of thought about how to answer the question of why meaningful human control is important.

The first is that meaningful human control is not, in fact, an additional requirement but rather a principle for the design and use of weapon systems in order to ensure that their use can comply with the laws of war. This way of thinking about meaningful human control starts from the assumption that the rules that determine whether the use of a weapon is legal are the same whether a human delivers a lethal blow directly, a human launches a weapon from an unmanned system, or a human deploys an autonomous weapon system that selects and engages targets on its own.¹¹

The second possible answer is a more maximalist position, which is that the existing principles under the laws of war are necessary but not sufficient for addressing issues raised by increased autonomy, and that meaningful human control is a separate and additional concept. From the maximalist perspective, even if an autonomous weapon could be used in a way that would comply with existing laws of war, it should be illegal if it could not meet the additional standard of meaningful human control. This view envisions meaningful human control as a new principle of the law of war to be on par with proportionality, distinction, military necessity, and others. Or rather, some have argued meaningful human control is already an implicit principle that has never needed explication until now, because there was never a way to exercise force without direct human control.¹²

This more maximalist position begs the question of what problem, or problems, autonomous weapons raise that existing principles under the laws of war cannot address. If a hypothetical future autonomous weapon system could be used in a manner that meets all of the requirements under the laws of armed conflict (e.g.

10. United States Department of Defense, *Autonomy in Weapon Systems*, 3000.09. Article 36, “Killer Robots: UK Government Policy on Fully Autonomous Weapons.”

11. Alex Leveringhaus and Gilles Giacca, “Robo-Wars: The Regulation of Robotic Weapons,” *University of Oxford Martin Policy Paper*, 2014, 15, <http://www.oxfordmartin.ox.ac.uk/downloads/briefings/Robo-Wars.pdf>. Also see Anderson et al, “Adapting the Law of Armed Conflict to Autonomous Weapon Systems.”

12. For example, Peter Asaro argues that this standard is derived from rights to human dignity embodied in the Geneva Convention, among other documents. Peter Asaro, “On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making,” in *Ethics of 21st Century Military Conflict*, Erica L. Gaston & Patti Tamara Lenard eds., (New York: International Debate Education Association, 2012), pp. 257-269 and Peter Asaro, Presentation in Geneva on May 14, 2014, http://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2014/statements/14May_Asaro.pdf.

distinction, proportionality), then what problem exists that requires an additional requirement for meaningful human control?

There are at least three concerns raised by those who have written on this issue, that they argue are not explicitly addressed in existing law of war principles and that meaningful human control might address:

1. **Accountability** – Chief among the motivations for a principle of meaningful human control is addressing the potential for an “accountability gap.” It is conceivable to imagine a future autonomous weapon that, when functioning properly, meets all of the requirements under the laws of war, but in the event of a failure that leads to it striking the wrong target, no human is deemed responsible for the targets that were erroneously engaged and thus no one is held accountable. In this scenario, even though humans developed, tested, manufactured, and deployed an autonomous weapon, the human controller did not choose the specific targets engaged – the malfunctioning weapon did. Thus, it is possible that no human would be responsible for the error.¹³
2. **Moral responsibility** – Machines are not moral agents. It is humans who are bound by the laws of war – and humans who adhere to it or break it. Even if a person committed a war crime with an autonomous weapon, it would be the human who committed the crime, using the autonomous weapon as the tool for committing the crime. For this to remain true, however, then humans must remain not only legally accountable but also morally responsible for the actions of autonomous weapons. Furthermore, some decisions pertaining to the use of weapons require legal and moral judgments, such as weighing likely expected civilian casualties against military advantages from conducting attacks. Some have argued that regardless of whether machines could perform these functions in a legally compliant manner, humans ought to perform them since they are also moral judgments.¹⁴ In this respect, “meaningful” refers to humans retaining moral responsibility for the use of weapons, even weapons that might incorporate high degrees of autonomy.
3. **Controllability** – Militaries generally have no interest in developing weapon systems that they cannot control. However, a military’s tolerance for risk could vary considerably across cultures and strategic positions. The desire for a battlefield advantage could push militaries to build weapons with high degrees of autonomy that diminish human control, particularly if they see such weapons as necessary to confront emerging threats or to keep pace with other militaries. While any weapon has the potential for failure and accidents, autonomous weapons arguably add a new dimension, since a failure could, in theory, lead to the weapon system selecting and engaging a large number of targets inappropriately. Thus, one potential concern is the development of weapons that are legal when functioning properly, but that are unsafe and have the potential to cause a great deal of harm if they malfunction or face unanticipated situations on the battlefield.

This means there are two interrelated issues that have to be resolved when thinking about meaningful human control. The first is the definition – figuring out a clear and useful definition of the concept. The second is what “meaningful human control” means in relation to international law. Without staking a

13. For a real-world example of this scenario, see the 2003 U.S. Patriot fratricides. During the early phases of the Iraq War, the U.S. Patriot air defense system erroneously engaged two friendly coalition aircraft, downing the aircraft and killing the pilots. See John Hawley et al., *The Human Side of Automation: Lessons for Air Defense Command and Control*, Army Research Laboratory, March 2005, <http://www.arl.army.mil/arlreports/2005/ARL-TR-3468.pdf>.

14. For example, see Christof Heyns, “Robot wars: after drones, a line we must not cross,” *The Guardian*, June 19, 2013, <http://www.theguardian.com/commentisfree/2013/jun/19/robot-wars-drones-life-death-decisions>; and Article 36, “Killer Robots: UK Government Policy on Fully Autonomous Weapons,” 2.

position as to which view of the notion of meaningful human control is preferred – whether it is a principle that generally helps ensure compliance with discrimination, proportionality, and other laws of war, or a new principle in addition to them, we propose that any useful understanding of what is meant by “meaningful human control” must address these concerns: accountability, moral responsibility, and controllability.

PRACTICAL CONSIDERATIONS

Several key factors may influence whether the concept of meaningful human control is successful in garnering support from states and other stakeholders interested in the topic of autonomous weapons and the use of force.

Any definition of meaningful human control that rules out the use of weapon systems that today are used without controversy, including systems that make civilian casualties less likely, is not practical or likely to be adopted.¹⁵ Moreover, since autonomous weapons, from the perspective of the Campaign to Stop Killer Robots, are generally an issue of concern for future systems, then a definition of meaningful human control that rules out large sets of current weapons systems likely misses what is potentially new about autonomy that warrants thinking about meaningful human control.¹⁶

For example, the 2014 definition of meaningful human control proposed by the International Committee on Robotic Arms Control (ICRAC) includes a provision that, for meaningful human control to be exercised, a commander must have “full contextual and situational awareness of the target area and be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack.” Yet, humans have been employing weapons where they lack perfect, real-time situational awareness of the target area since at least the invention of the catapult.

The ICRAC statement also argues, “[T]here must be a means for the rapid suspension or abortion of the attack.”¹⁷ However, the essence of a projectile weapon, since the first time a human hurled a rock in anger, is the inability to suspend and abort the attack after launch. It is only with modern, advanced weapons that commanders have had the ability to retarget or abort a projectile in flight.

These proposed requirements articulate an idealized version of human control divorced from the reality of warfare and the weapons that have long been considered acceptable in conducting it. Not only would they arguably outlaw whole classes of weapons dating back to the sling and stone, but they also fail to capture what is new about autonomy. If meaningful human control is defined in such a way that it has never existed in war, or only very rarely, then such a definition sheds little light on the new challenge posed by increased autonomy.¹⁸

If the intent of the notion of meaningful human control is to make a distinction between how force has been historically used and some potential point in the future where humans have less meaningful control,

15. Paul Scharre and Michael C. Horowitz, “An Introduction to Autonomy in Weapon Systems,” *Center For A New American Security Working Paper*, February 2015, http://www.cnas.org/sites/default/files/publications-pdf/Ethical%20Autonomy%20Working%20Paper_021015_v02.pdf.

16. Paul Scharre, “Autonomy, “Killer Robots,” and Human Control in the Use of Force, Part II,” *Just Security*, July 9, 2014, <http://justsecurity.org/12712/autonomy-killer-robots-human-control-force-part-ii/>.

17. Frank Sauer, “ICRAC statement on technical issues to the 2014 UN CCW Expert Meeting,” May 14, 2014, <http://icrac.net/2014/05/icrac-statement-on-technical-issues-to-the-un-ccw-expert-meeting/>.

18. Scharre makes a similar point, Scharre, “Autonomy, “Killer Robots,” and Human Control in the Use of Force, Part II.”

then a definition of meaningful human control must, at a minimum, account for the varied contexts in which force has been used to date.

Additionally, a plausible definition of meaningful human control should cover the use of weapons across all of the various ways they may be used. A definition that works very well for thinking about the use of force by soldiers storming a building in a city but which does not reflect the realities of the use of force from the air against enemy formations is unlikely to be useful in a wide range of scenarios pertaining to autonomy in weapons and unlikely to succeed in garnering widespread adoption.

Finally, an effective definition of meaningful human control must reflect a realistic vision for how weapons are used by soldiers in all domains, including on the ground, in the air, and on and under the sea. A definition derived from idealized visions of weapons' use that do not reflect the realities of the battlefield and how decisions are made will necessarily run into opposition from key stakeholders and be less useful in helping countries around the world make sense of autonomous weapon systems.

EXISTING DEFINITIONS OF MEANINGFUL HUMAN CONTROL

While some degree of human control over weapons seems not only appropriate but also essential, what is intended by adding the modifier “meaningful” to characterize the type of human control that is desired? For advocates of the phrase, the specific term “meaningful” is less important than the concept behind it. As an Article 36 briefing paper notes:

*It should be noted that whilst this paper uses the term ‘meaningful human control’ there are other terms that refer to the same or similar concepts. These include ‘significant’, ‘appropriate’, ‘proper’, or ‘necessary’ ‘human judgement’ or ‘human involvement’.*¹⁹

One way to think about what is captured by the term “meaningful” is by thinking about what it excludes. Consider a person who sits in a room and is supposed to press a button every time a light bulb in the room goes on. If the person does this as instructed, and a weapon fires each time the person presses the button, a human has fired the weapon, but human control over the weapon is far from meaningful.²⁰ Alternatively, the Platonic form of meaningful human control is when a person swings a sword, axe, or knife – or uses their bare hands – to directly end the life of an enemy combatant.

The reality is that for many engagements today, as we describe below, humans launch weapons based solely on information received from computer sensors. Yet in those cases, humans are clearly in control of the weapons and responsible for its effects. This is because the human making that decision is not merely pushing a button reflexively when a light goes on, but rather using information gained from computer sensors, along with other information he or she has about the weapon, potential targets, and the overall situation, to make a judgment about whether a given engagement is lawful and appropriate in that instance. A useful definition of meaningful human control, therefore, must be able to distinguish between this scenario and one where a human lacked meaningful control.

19. Article 36, “Memorandum for delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS).”

20. Asaro, “On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making.” Article 36 uses a similar example in “Memorandum for delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS).”

In their original 2013 policy paper introducing the concept of meaningful human control, Article 36 argues that meaningful human control has three necessary requirements:

- **Information** – a human operator, and others responsible for attack planning, need to have adequate contextual information on the target area of an attack, information on why any specific object has been suggested as a target for attack, information on mission objectives, and information on the immediate and longer-term weapon effects that will be created from an attack in that context.
- **Action** – initiating the attack should require a positive action by a human operator.
- **Accountability** – those responsible for assessing the information and executing the attack need to be accountable for the outcomes of the attack.²¹

The International Committee for Robot Arms Control, in their May 2014 statement on meaningful human control, listed these “minimum necessary standards” for meaningful human control:

- *First, a human commander (or operator) must have full contextual and situational awareness of the target area and be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack.*
- *Second, there must be active cognitive participation in the attack and sufficient time for deliberation on the nature of the target, its significance in terms of the necessity and appropriateness of attack, and likely incidental and possible accidental effects of the attack.*
- *Third, there must be a means for the rapid suspension or abortion of the attack.*

While, as stated previously, the degree of information that the ICRC statement requires is not practicable in warfare, both the Article 36 and ICRC statements emphasize the general theme of informed action by a human. While the standard for information required may be unrealistic in these proposals, informed action is central to the concept of meaningful human control. The lack of informed decision-making is, after all, the problem with the person-pushing-a-button-when-a-light-comes-on scenario. In such a scenario, the person is taking action, but without any information about the context for the use of force or the consequences of their action. This raises the question of how much information is required for a human operator to make a meaningful decision about the use of force. Before arbitrarily proposing a standard for future military operations, however, it is worth first considering how meaningful human control is exercised in weapons today.

HOW IS MEANINGFUL HUMAN CONTROL EXERCISED OVER THE USE OF WEAPONS TODAY?

When thinking about meaningful human control, it is important to recognize that human decision-making about the use of weapons, at present, already varies significantly in different situations. The context in which a weapon is used alters how humans exercise control over the use of that weapon. Consider several possibilities:

Ground engagement: Consider a ground engagement where a commander is deciding whether to send a

21. Article 36, “Killer Robots: UK Government Policy on Fully Autonomous Weapons.”

squad into a building to kill enemy combatants. This is a common scenario discussed in dialogue concerning meaningful human control. Is the point of meaningful human control when a senior commander decides on the rules of engagement, when a local commander authorizes the mission, or when each individual that enters the building fires a weapon? Each decision is significant and meaningful. Furthermore, while higher-level commanders may have time to deliberate the potential consequences of their orders, troops entering a room may have only fractions of a second to determine whether a person is a civilian or a combatant.

Air-to-air engagement: The concept of meaningful human control is made even more complicated when considering the use of weapons in the air. A fighter pilot deciding whether to fire at an enemy fighter might have mere seconds to make a decision about whether to release a missile or fire guns. In a beyond-visual-range engagement, machines – sensors and computer processing – provide all of the information the pilot is using to make the decision. Moreover, if it is an air-to-air missile, after launch, a computer on the weapon then guides the weapon into the target. Essentially, the degree of human control a pilot has when firing a missile in a beyond-visual-range engagement is radically different from a soldier, in a building, firing at an enemy combatant.

Air-to-ground engagement: There are a number of different ways human control might be executed in air-to-ground engagements. In an engagement against a pre-planned fixed target, a pilot might carry out a strike authorized through a deliberate targeting process by a number of individuals – lawyers, staff personnel, and commanders. The pilot might be responsible for carrying out the mission, but not for identifying and selecting the target. Moreover, when it comes to finding the specific target and ensuring the munition hits the target once the mission is underway, the work is done by sensor pods and computers – machines – that provide the totality of the information the pilot has before launching the weapon.

In response to emergent threats during an ongoing mission, on the other hand, a pilot might have to make the decision to use weapons personally, and in a very compressed amount of time. When supporting ground forces, if a qualified ground controller is on the ground directing close air support, primary responsibility and decision-making for the use of weapons might rest with the controller on the ground. In all three scenarios for air-to-ground engagement, there is human control over those weapons and accountability, but that responsibility might fall with different individuals. In all three scenarios, as with an air-to-air engagement, computer systems play a significant role both in helping the pilot find a target and in guiding weapons to their targets.

Naval engagement: As in air engagements, commanders launching missiles or torpedoes in surface or undersea warfare are usually relying heavily on computers to aid in finding targets, identifying them, and maneuvering munitions toward the targets once the person selects the target and the missile is launched. This is especially true for engagements beyond visual range, whether ship-to-ship or launching cruise missiles against targets on land.

Human-supervised autonomous defensive systems: At least 30 countries possess human-supervised autonomous defensive systems to defend ships, vehicles, or ground bases against short-warning attacks

from incoming missiles, rockets, or planes.²² These are designed for situations where the time of engagements may be too short for humans to adequately respond, necessitating automation. These systems have been in existence for decades without raising significant questions about their legality or appropriateness.²³ Human control is exercised by determining the system's rules of engagement the decision to activate the system, and real-time human supervision of its operation to turn it off if the context is no longer appropriate for use. But despite this control, humans do not need to push a button to fire at each target.

Thus, it seems necessary for the definition of meaningful human control to be flexible enough to account for significant variations in the context in which force is already used. In all of these cases, human control is exercised differently, but that control can be said to be “meaningful” in one sense or another. In all of these examples:

1. Humans are making informed, conscious decisions about the use of the weapon (no one is merely pushing a button when they see a light blink on).
2. The information they have to make that decision is sufficient for them to ensure the lawfulness of the action they are taking, given what they know about the target, the weapon, and the context for action. When key related actions are performed by others (such as one person choosing the target and another person releasing the weapon), commensurate responsibility for those actions falls with the people taking those actions.

Because the above two conditions are met, the humans are responsible for the actions they are taking, and therefore accountable for those actions. To the extent that they are relying on others to make important decisions, responsibility for those other decisions lies with other humans.

Furthermore, in all these examples, human operators have effective control over the use of weapons. This is the case even if some of them are “fire and forget” weapons that cannot be recalled after launch. This is because trained human operators have a clear understanding of how the weapon will function in certain environments as well as its limitations, so they can use it appropriately.

The varied ways in which weapons are used today suggests that setting an arbitrary standard for how much information a person must have before using a weapon, such as ICRC's standard of “full contextual and situational awareness” is not consistent with how weapons are actually used in warfare. Rather, Article 36's approach of “adequate” information is more appropriate.

How much information is adequate? It should be enough information about the target, the weapon, and the context for engagement for the person to make an informed decision about the lawfulness of their action. This does not mean that each human operator involved in the chain of decision-making need have the complete picture of information about the target, the weapon, and the context for engagement. As happens today for soldiers entering a building or a pilot dropping a bomb on a pre-planned target, human operators may rely on decisions that have been made by other humans. However, relying on others does not mean blind trust or abrogating moral judgment. A single individual may not be responsible for all aspects of

22. These include, but are not limited to, Australia, Bahrain, Belgium, Canada, Chile, China, Egypt, France, Germany, Greece, India, Israel, Japan, Kuwait, the Netherlands, New Zealand, Norway, Pakistan, Poland, Portugal, Qatar, Russia, Saudi Arabia, South Africa, South Korea, Spain, Taiwan, the United Arab Emirates, the United Kingdom, and the United States. Paul Scharre and Michael C. Horowitz, “An Introduction to Autonomy in Weapon Systems.”

23. United Nations Institute for Disarmament Research, “Considering how Meaningful Human Control might move the discussion forward,” 5.

decision-making relating to attacking a target, but any given person can be held accountable for his or her own actions related to that attack.

CONTROLLABILITY

One additional aspect of meaningful human control that some commentators have raised pertains to controllability. In their report, UNIDIR raised a number of important aspects related to weapons design for controllability, including system predictability, operator understanding of the weapon system, geographic and time limitations on the weapon, and mobility of the weapon system. In addition, the U.S. Department of Defense Directive on Autonomy in Weapon Systems includes a long list of requirements designed to ensure “appropriate levels of human judgment over the use of force.” These include:

- Hardware and software-level safeties;
- Test and evaluation in realistic operating environments against adaptive adversaries;
- Human-machine interface to ensure transparent operation to trained human operators;
- Clear procedures for trained operators to activate and deactivate system functions;
- Time constraints on system operations;
- Anti-tamper and cyber-security measures; and
- Appropriate training for human operators and doctrine for employment.

All of these are important for ensuring that human operators are able to employ complex systems with high degrees of autonomy in conflict environments, and that these systems operate in a manner consistent with operator intentions.

ESSENTIAL COMPONENTS OF MEANINGFUL HUMAN CONTROL

The concept of meaningful human control is intended to address three core problems: accountability, moral responsibility, and controllability. If meaningful human control is to adequately address the challenge of increased autonomy in future weapons, it must take into account the varied ways in which weapons are used today. Informed decision-making is a critical component of meaningful human control, yet the degree of information that humans have today about the weapon, the target, and the context may vary considerably across different scenarios. The key common feature is that humans have sufficient information to assess the lawfulness of the action they are taking. This ensures both accountability and moral responsibility for the actions human operators take. In addition, controllability must be an important consideration in weapon design, as it always is. Although requirements to abort or retarget a weapon once launched lack feasibility and precedent in warfare, there are a host of key design considerations that must be considered to ensure effective and adequate control over a weapon. A weapon system that is uncontrollable by its very nature is already likely to be difficult to use in ways that comply with the laws of war and in ways that would meet military objectives.

Thus, meaningful human control has three essential components:

1. Human operators are making informed, conscious decisions about the use of weapons.
2. 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.

3. The weapon is designed and tested, and human operators are properly trained, to ensure effective control over the use of the weapon.

From a legal and practical perspective, the accountability issue is critical. From a practical perspective, meaningful human control would mean a clearly identifiable person who makes the decision to deploy lethal force. The above standard of meaningful human control would help ensure that commanders were making conscious decisions, with sufficient information to remain legally accountable for their actions. Even if the commander is not pulling the trigger, the commander can still be accountable for the decision, as commanders are in a variety of military contexts. However, further research and thinking will be necessary to try and determine exactly how to ensure accountability across different scenarios.

MEANINGFUL HUMAN CONTROL AT WHAT LEVEL?

One important outstanding issue in how to define and apply the concept of meaningful human control is the level of human decision-making at which that control should be exercised. There are multiple levels where control could be applied in a given engagement. At the highest level of abstraction, a commander deciding on the rules of engagement for a given use of force is exercising meaningful human control over the use of force. Below that, there is an individual commander ordering a particular attack against a particular target deemed legitimate under international law. Along a different axis, meaningful human control might refer to the way a weapon system is designed in the first place – ensuring it is programmed, as it moves from concept to reality, to ensure it is possible to use the system in a way that complies with international law.

Determining the appropriate level at which meaningful human control is required is important. Article 36 has asserted that meaningful human control should govern individual attacks, although they argue that one attack could encompass multiple targets.²⁴ Is meaningful human control required over individual attacks? Why or why not? And, most importantly, how does the level of meaningful human control affect concerns about accountability, moral responsibility, and controllability?

CONCLUSION AND ADDITIONAL ISSUES TO CONSIDER

The issue of meaningful human control over the use of weapon systems, including both how to define the concept and what it means in relation to international law, is an important element in the overall debate over autonomous weapons. Some fear that autonomous weapon systems will inherently lack the human control necessary to ensure that they are used in ways that comply with the laws of war. In addition, some have raised additional considerations such as accountability, moral responsibility, and controllability that they argue are not explicitly captured in current law of war principles. Moreover, some argue that meaningful human control over the use of force is already an implied part of international law, meaning weapons that lack meaningful human control ought to be illegal, but that requirement should be made explicit.

To helpfully inform debates about autonomy in weapon systems, the phrase meaningful human control requires greater context. This paper is intended to define the essential elements of meaningful human control.

²⁴ “In addition, it is important that this control relates to individual attacks. Whilst it is recognized that an individual attack may include a number of specific target objects, human control will cease to be meaningful if an autonomous weapon system is undertaking multiple attacks that require specific timely consideration of the target, context and anticipated effects.” Article 36, “Killer Robots: UK Government Policy on Fully Autonomous Weapons,” 4.

They include human operators making informed, conscious decisions; human operators having sufficient information about the target, the weapon, and the context to ensure the lawfulness of their actions; and design and testing of the weapon and training for human operators to ensure effective control over the use of the weapon.

There are many outstanding issues, however, including the level at which meaningful human control should be required, whether meaningful human control is an overarching concept to ensure compliance with the laws of war or a new principle in its own right, and whether meaningful human control is even the right concept for thinking about human control over autonomous weapons. Further discussion and dialogue is needed on autonomy and human control in weapon systems to better understand these issues and what principles should guide the development of future weapon systems that might incorporate increased autonomy.