Stay In Command : ICRC Perspective

Matilda Byrne: [00:00:00] Welcome to SafeGround, a small organisation with big ideas working in disarmament, human security, climate change and refugees. I'm Matilda Byrne.

Welcome to Stay in Command a series which discusses fully autonomous weapons or lethal autonomous weapons systems and different dimensions and concerns.

Today, we'll be getting insight from the international committee of the red cross or ICRC who are active in research and dialogue on this important emerging weapons issue. The ICRC is an impartial, neutral, and independent organisation whose exclusively humanitarian mission is to protect the lives and dignity of victims of armed conflict.

I'm joined now by Neil Davison from the ICRC headquarters in Geneva, where he is a senior advisor in the department of international law and policy. He has been working on weapons and disarmament issues for almost 20 years, the last 9 at the ICRC. Thank you for being here.

[00:00:59] **Neil Davison:** Pleasure. Good to be here.

[00:01:00] **Matilda Byrne:** So before we dive into talking about the issue specifically of lethal autonomous weapons, I was wondering if you could speak more to the ICRC's general mandate and mission and its work, and in particular, how it approaches a weapons issues?

[00:01:14] **Neil Davison:** Well, I mean, our mandate is to assist and protect victims of armed conflict and other situations of violence and our work on weapons really focuses on two factors, looking at the potential risks for civilians and those fighters no longer taking part in the conflict, and interconnectedly the compatibility of weapons or their use with international humanitarian law, law of war, including the principles of humanity, which is sort of where ethics meets the law.

[00:01:45] So, when we're looking at new weapons technologies, we tend to have obviously less information from the field on the actual consequences. So we try to assess the foreseeable impact and it can be quite difficult. There's often quite a lot of claims made about how new weapons may or may not be used, the capabilities that they may or may not have. And some of these claims are often not actually borne out in practice and don't match reality and driven by quite unrealistic scenarios. So, we really emphasize having a realistic assessment of the weapon, the technology, and its likely use. This is the approach we've applied to looking at autonomous weapon systems, over the past 10 years or so.

[00:02:27] **Matilda Byrne:** Great. And so with that work on autonomous weapons systems and seeing this advancement of autonomy in weapons towards even lethal autonomous weapon systems or fully autonomous weapons, what does the ICRC see as the main concerns around and these emerging weapons?

[00:02:45]**Neil Davison:** Mmm, I should say actually just at the outset that for about the last five years we've been calling for internationally agreed limits on autonomous weapons systems and in some ways, the concerns about these types of weapons are quite simple. We understand these weapons - We don't use the terminology lethal autonomous weapons,

just autonomous weapon systems - but, these are systems that select and apply force to targets without human intervention. So they fire themselves essentially based on the interaction of their sensors and the environment. And it's different from a lot of other weapons issues, because it's something that could be applied to any weapon really, in theory.

And for us, from a humanitarian perspective, the crucial thing to understand is that the user of an autonomous weapon system of any type, does not actually choose or know specifically, the target they will hit, nor even exactly when or where it would be hit and that's really the root of the problem because the consequences, therefore, are always unpredictable to a degree. And that's even leaving aside the issue of malfunctions, which we all know happen with any complex systems, especially software based systems. So, it's that unpredictability, which is found we would say in all autonomous weapons systems, which can even be compounded at design level where you start to use, let's say AI and machine learning software that is not properly understood, or that even changes its functioning over time, that can even compound the unpredictability at a design level. So basically, you know, this problem of unpredictable consequences effectively means potential risks for civilians and civilian objects; homes, schools, hospitals, as well as, you know, combatants who are no longer fighting, injured, surrendering. And the root of it is this: if someone is in a conflict using such a weapon system, without knowing exactly what it's going to hit and where, and when, how do they assess the risks and how do they manage those risks?

One way to do this is actually to add extreme constraints on the types of situations and tasks they're used for. So today you see some autonomous weapons used already. For example, defense systems have autonomous modes for shooting down incoming missiles, but they're very narrow tasks, only against objects, measures are taken to ensure there are no civilians or civilian objects, civilian aircraft in that area while they're activated in autonomous mode and all sorts of other constraints. Now, the danger really is that looking at an expanding array of systems in the air, on the ground, at sea; there'll be mobile, they'll be AI enabled potentially, they could be used to target people rather than military objects and used predominantly where wars are taking place today - in cities and towns. And so here, this unpredictability that I mentioned presents in our view, a serious danger for civilians. And that is our sort of overarching concern. But, they do also raise serious legal questions and fundamental ethical concerns that I'm happy to go into in a bit more detail.

[00:05:40] **Matilda Byrne:** Yeah, definitely. I think, you know, the humanitarian imperative is quite clear in terms of the risks for civilians. But in addition, if you could break down maybe some of the key legal issues that are posed by autonomous weapons.

[00:05:53] **Neil Davison:** Well, the legal issues, also in a way, are quite straightforward. So essentially humanitarian law rules on the conduct of hostilities. They require those people carrying out attacks in armed conflict to make specific judgements, to ensure their attacks are lawful and generally to protect civilians and they must ensure that they only attack legitimate military targets - that's the rule of distinction, and they must ensure that any dangers for civilians that may arise from their attack are proportionate to the military advantage - that's the rule of proportionality, and they must also be able to cancel or

suspend an attack, should the situation change, so should the risk for civilians change, that might affect their assessment of proportionality or their ability to distinguish - they need to be able to take precautions and that's the rule of precautions in attack.

[00:06:42] So, I mean, the key thing to understand here is that these types of judgements are not only required of human combatants, they're also highly context dependent. So they require an assessment in the circumstances of a specific attack. And this is where autonomous weapons raise a major challenge for that process because with autonomous weapons, you're moving from a very specific type of decision-making with normal use of weapons, where you choose a specific target and you choose to attack it at a specific time and place, to a sort of generalized decision-making where you activate a weapon and it triggers itself. So you have less knowledge about what's going to happen. So the question is how can you effectively make these assessments and apply the rules? How can you judge the proportionality? How can you take precautions? It's very difficult. I mean, I come back to what I said before: one way of doing this in a way, and it's what's done today with the existing autonomous weapons, is to ensure there are no civilians or civilian objects there. That's one way of doing it in a very highly constrained way. The system is still unpredictable in a sense, you don't know exactly when it's going to fire or against what, but you've taken measures to sort of ensure that unpredictably doesn't matter, it doesn't present risks to civilians, but you know, again, looking to the future, looking at the range of armed, unmanned systems where there's interest to, to make them autonomous and looking at most conflict scenarios today, there are civilians present. And so this is going to be a major problem in terms of upholding the law.

[00:08:14] **Matilda Byrne:** Definitely. And I think you were talking about in terms of upholding the law, that it is sort of carried out by humans that are making these contextual judgements in terms of all of the different circumstances and doing these things like evaluating the proportionality of attack and taking precaution and things like this, which leads me to this notion of human control, which lots of people are talking about in terms of the discussion of autonomous weapons. And so I was hoping you could explain the notion of human control from the ICRC's perspective and why it is important.

[00:08:46] **Neil Davison:** Sure. Yeah, I mean, human control and judgment is fundamental to the discussion because, because like I say, adding autonomy to a weapon system is a feature, it's not a specific category - it could be applied to any weapon. So it's really a method of using force in that sense. But human control really underpins the legal obligations that I mentioned, and human judgment. It also underpins ethical responsibilities. And I think the important thing to understand here, there's a misconception or there are often misleading comparisons made, but machine calculations are not equivalent to human judgment, and they never will be because humans are persons with legal obligations and moral responsibilities. Machines, weapons, software will always be inanimate objects- they do not have these, these obligations or responsibilities. So, you know, the issue of human control is that in order to uphold these legal obligations and ethical responsibilities, you have to have some control over the weapons, the machines you're using and the consequences that results. And, you know, that's a critical issue.

The more difficult question is exactly what is the type and extent of human control needed, legally and also ethically. And in a way, in some of our work recently, there's also a parallel or a reinforcing requirement for human control from a military operational perspective, because militaries, uh, want to have control over the weapons they use and the effects they cause. So there should be a collective interest in that. And so the question is determining what that is in practice. Where do we draw the line and what is acceptable or not? The ICRC has made a few, a few suggestions on that, which I can expand on a bit.

[00:10:36] Um, but perhaps just returning to the ethical aspects a bit like human judgment for applying the law, the issue from an ethical perspective, I mean, having talked to many ethicists about this over the years in our work, is that, you know, it's really concerned about loss of human agency in life and death decisions. So this is, it's really most acute with autonomous weapon that presents risk to human life and especially those that were designed or used to target people. I mean, I think the way to capture it is the sense that widely speaking, you see this also in public opinion surveys, that you know, an algorithm, a machine should not decide who lives or dies, an algorithm should not be able to kill it. So, what does this mean? What does preserving human agency in those life and death decisions mean? Um, you know, one way to look at it; it means there has to be some effective human deliberation about that decision. And if there hasn't been that deliberation, you can say that there hasn't been morally responsible decision-making, nor the recognition of the human dignity, the dignity of those who may be killed or injured. Another way to put it is that if that deliberation hasn't taken place, it's a kind of dehumanizing process that sort of undermines our shared humanity. And I think there are obviously parallels with here and other parts of society, where there are current ethical debates about the degree to which algorithms and machines inform our decision-making or take over certain tasks, that may have serious consequences for life, of course, decisions to kill and use weapons being the most serious: you've got to think twice about that.

[00:12:18] **Matilda Byrne** Definitely. I think with the progression of AI and society in particular, it really is important to take pause and reflect on where do we need to draw a lines? Where should there be limits, what decisions should never be given to a machine such as those over life and death. And so, I guess it kind of then leads back to the question you mentioned earlier is then what is the extent of control required within autonomous weapons to safeguard these important principles and also kind of address the legal concerns. So, the ICRC has suggested limits on autonomy in a few different ways. And so I was hoping you could elaborate what this means, and I guess some of the practical suggestions, so things like operational limits, so we're think about temporal or spatial contexts and things like that, what we're really talking about when we talk about human control, that's required over autonomous weapons.

[00:13:11] **Neil Davison:** Sure, yeah. We've been looking at this for a number of years, trying to find a practical answer to this difficult question. Last summer we put out a report, jointly actually with the Stockholm International Peace Research Institute, where we, we looked at the demand for human control, from a legal perspective, from an ethical perspective and from a military operational perspective. And we made an assessment of, and like I said before, I think there was a demand from all those perspectives. And we made some

recommendations about what that might look like in practice. And essentially we proposed three types of limits or control measures that are overlapping. And the first is control on the weapon parameters. So these types of controls can inform limits on the types of autonomous weapons, the types of tasks they use, particularly the types of targets they're used against. So, one way would be to delineate between weapons used to target people and those to target objects, particularly, you know, military objects.

[00:14:12] There are constraints also there in terms of how long a system operates in autonomous mode and the geographical scope, the area of its operation - and those are things that it's perhaps more difficult to be definitive on in all circumstances, they may be quite context dependent. Certainly, the more complex the environment, the shorter and the smaller area you need in order to have a certain type of control. If you've got complex urban area and you don't know where your weapon is going to fire, then you've got, you know, you've got problems.

[00:14:49] The other issue is still talking about control of weapon parameters, is requirements for deactivation measures, and these can be both, kind of, remote intervention by someone who's supervising the system and, or including, you know, self deactivation mechanism, but you know, somewhere to turn it off, essentially.

[00:15:09] So that's the first area, the second area, and like I say, these are overlapping, the second area is controls on the environment. So these types of controls can inform limits on the situations and locations in which the autonomous weapons might be kind of lawfully acceptably used and I think the major consideration here, like I say, is the presence and density of civilians and civilian objects. And this overlaps, for example, with the issue about the duration and time and space that I mentioned, but also with, you know, the types of constraints on targets that are, that are set.

[00:15:43] The third area are controls through human machine interaction. So these types of controls can inform requirements for human supervision of such systems, ability to intervene, deactivate it, should the situation change. I mean, that's very much linked to the obligation to take precautions, in international humanitarian law. In addition, an important factor here is predictable and transparent functioning. So like I say, we always have some unpredictability in the consequences of using an autonomous weapon, but where you might have even more problem is, is unpredictability at the design level. So if you don't know how systems function, if you don't effectively know how the software that controls your weapon works, then it's going to be majorly problematic.

[00:16:29] So, so we think these three types of control measures, like I say, can inform limits, constraints agreed at the international level on the, the types of autonomous weapons used and the types of targets they are used against, types of situations in which they may or may not be used and requirements for how humans supervise, intervene, deactivate and design such systems in a way that they function predictably.

[00:16:59] **Matilda Byrne:** Great. Yeah, and I guess, this notion of human control, is very much this idea that's developing and exactly what it means to be applied to autonomous weapons or weapons systems broadly, and it's something that is also being discussed by

different countries. So the governments of the world, when they convene in the diplomatic processes, human control is something that increasingly is being talked about in different ways by different countries. And I was just wondering if you could speak to what you think is encouraging about the ongoing diplomatic talks in this area and on the issue of autonomous weapons more broadly.

[00:17:32] **Neil Davison:** Mmm. Yeah, well I think the discussions certainly that took place in September last year have taken a turn towards the more encouraging. You know over many years there was a lot of quite unfocused discussion and there is now a recognition among most States that this issue of human control, involvement, judgment - some governments prefer different phrases - that's, that's the central issue. I'd say it's fair to say there's a recognition for requirement for, for the human control, involvement or judgment. There is also recognition now among many governments about the types of measures that will be needed to ensure that, and those are some of the ones that I mentioned before. And I think thirdly, there's a recognition that these types of measures really are at the roots of the work they need to do. In the terminology of the UN Convention on Certain Conventional Weapons discussions in Geneva, they would say verification, consideration and development of the normative and operational framework. I think, you know, in a more simple terminology that would be essentially internationally agreed limits.

[00:18:42] So in that sense, it is encouraging. And you do have against that background, a sort of enduring disagreement about what you do with those limits. So, the majority of States want to see a new legally binding rules, whether a new protocol to the CCW or otherwise. But other States perhaps have not called for new rules, they want to see perhaps more policy standards or best practices that are sort of non-binding, but somehow agreed politically. So you have kind of enduring disagreement about the process, but you do, I would say have some increasing focusing of an agreement and convergence of views on the substance which is encouraging.

[00:19:28] You know, on the other hand, I think it's becoming a bit of a crunch time now for the CCW and its work on autonomous weapons - seven years of discussions, in different informal and more formal settings, a lot of work done, a lot of issues explored in a lot of detail. And it's now a time to take action to build on that and to crystallize what has been learned into some practically applicable policy solution.

[00:20:00] The ICRC, we think it's really a fundamental issue for the future of warfare. We believe that international agreement is really needed quite urgently. Each week, there are new reports of weapons developed, deployed, transferred with increasingly autonomous functions. It's not often clear exactly how they function whether they're yet autonomous. Essentially we're on a line we're on a fence that we're about to cross potentially with potentially quite serious consequences for civilians, for the law, and for humanity. So, if we want to prevent those risks, then governments really need to take action soon.

[00:20:45] **Matilda Byrne:** Yeah, definitely. And I guess, I think it would be fair to say that, to address that kind of ethical imperative that exists in terms of dehumanization and the risk to civilians, as well as upholding the law, these sort of internationally agreed limits that you're speaking to is really what's required and action needs to be taken in order to really

crystallize that for the international community and set these new standards. Is there anything else that you wanted to add?

[00:21:15] **Neil Davison:** Um, I don't think so. I think that, well maybe I would just add that, like I said, it's a crunch time for these discussions. You know, there are difficulties at the moment with even holding the meetings in Geneva because of the current situation with the pandemics and meetings have been postponed.

[00:21:33] But at the end of this year, it's still scheduled the five yearly review conference of the convention on certain conventional weapons, the CCW, so that we see really as a critical juncture in this debate and in the political response. So we'll be looking to promote the practical limits we've identified and build support for that action towards the end of this year.

[00:22:01] **Matilda Byrne:** Absolutely. I think a lot of people, their eyes are kind of looking forward to that review conference and really hoping that States can band together and get some really decisive action happening at such a critical time on this crucial issue. So, um, thank you so much, Neil, for your insights today and bringing us the ICRC perspective.

[00:22:20] **Neil Davison:** Pleasure. Thank you for the invitation.

[00:22:22]**Matilda Byrne:** If you want to know more, look for us on Facebook, Twitter, and Instagram Australia campaign to stop killer robots all use the hashtag AusBanKillerRobots - become part of the movement. So we stay in command.

[00:22:37] Thank you for listening and please share with your friends. For access to this and other episodes along with full transcription and relevant links and information, head to safeground.org.au/podcasts. Our are podcasts come to you from all around Australia and we would like to acknowledge the traditional owners throughout and their continuing connection to country land, waters and culture. Stoke audio provided by videvo downloaded from www.videvo.net. Thank you for listening to Safe Ground.