Presentation at the United Nations Convention on Certain Conventional Weapons

Paul Scharre

April 13, 2015

Autonomy is playing a bigger role in many industries and aspects of our lives, from self-driving cars to autonomous Twitter bots. Similarly, autonomy will be used increasingly in military operations. A central question, then, is how much autonomy?

When we think about autonomy, we should think about three dimensions of autonomy.

The first is the relationship between the human and the machine.

- In semi-autonomous systems, a human is "in the loop." The machine will take some action and then stop and wait for a human to take a positive action before it continues.
- In supervised autonomous systems, a human is "on the loop." The machine will take action and will not wait for the human, but the human can intervene to stop the machine's operation.
- In fully autonomous systems, a human is "out of the loop." The machine will take some action and the human cannot intervene.

These are important distinctions when thinking about the risk of the machine getting something wrong.

The second dimension of autonomy, which is separate from the first one, is the degree of intelligence of the machine. People use words like automatic, automated, autonomous, and intelligent to refer to a spectrum of complexity of machines. But there are no clear distinctions between these categories, and different people can disagree on how much complexity is required for a machine to move from automated to autonomous.

The third dimension, and undoubtedly the most important one, is which tasks the autonomous system is performing. Both a toaster and a mine are very simple automatic systems, but the tasks they are performing are very different. Rather than talk about systems being "autonomous" or not, we should be careful to specify which tasks are autonomous. Consider a self-driving car. A self-driving car would stop, go, and change lanes on its own. It might even pick the route on its own. But presumably a human is choosing the destination.

This dimension – which task is being performed – is the most important dimension of autonomy. For each task, we can ask whether for that task the system is semi-autonomous, supervised autonomous, or fully autonomous. That is: Is a person doing the task? Is the machine doing it under human supervision? Or is the machine doing it and the person cannot intervene? And this is different from the intelligence

of the machine. We can have very intelligent systems that still have a human "in the loop" or "on the loop." All of these three dimensions of autonomy are separate aspects of a system.

Our discussion this week focuses on one particular decision or task in war, the decision that specific targets are to be engaged by the use of force. Now, autonomy is already used for many tasks in warfare. These include identifying potential targets, tracking them, the timing of when to fire, and maneuvering or homing onto targets. In some cases, military weapons like missiles and torpedoes are fired and, once released, humans have no more control over the weapons – they are "fire and forget" weapons. This is not a new development. These weapons have been in use by militaries for over 70 years. But, with a few rare exceptions, humans still make the decisions about which specific targets are to be engaged. Homing missiles have onboard seekers, but are generally narrow in their field of view and their ability to loiter to search for targets. The missile is maneuvering toward the target, but humans are still in control of which targets are destroyed because they are launching the missile at a specific target.

Autonomous weapons would be different. They would be different not necessarily in their degree of intelligence, but rather in the task they would be performing. Autonomous weapon systems would be ones where a human launched the weapon or otherwise put it into operation, but once activated, the weapon would be selecting and engaging targets on its own. Thus, the task of deciding "this particular target will be engaged with force" which was once done by a human is now done by a machine. The autonomous weapon would still be put into operation by a human, however, and would still be operating under programming instructions and rules of engagement written by humans.

Now, for some, this brings to mind visions from science fiction movies of beady-eyed humanoid robots stalking through villages. But Hollywood movies are not necessarily helpful in envisioning the future of autonomous weapons. Autonomous weapons, if developed, are more likely to first come in the form of loitering missiles, adaptive cyber weapons, or undersea sub-hunting autonomous vehicles. Precisely because it is so difficult to build a robot that could accurately discriminate among combatants in ground conflicts where civilians are present, autonomous weapons are most likely to be introduced, if they are built, in areas where there are few if any civilians, such as undersea, in the air, or in space or cyberspace.

Nor, I want to point out, are we necessarily talking about systems that are intelligent, learning, self-aware, or have free will. The chess-playing computer Deep Blue is not self-aware, but it plays chess on its own. An autonomous weapon need not be very intelligent. It is simply a weapon that selects and engages targets on its own.

Why would militaries want autonomous weapons? Autonomy can be useful for a wide range of functions, but why would people want to take a human "out of the

loop" for the decision about which specific target to engage? I can see three possible reasons:

The first is speed. Today, at least 30 nations already have autonomous weapons of a limited type: defensive systems with human-supervised autonomous modes to defend against short-warning attacks. These include air and missile defense systems as well as active protection systems for ground vehicles. These systems are used to defend military vehicles, bases, and civilian populations from short-warning attacks from rockets and missiles where the volume and speed of engagements would overwhelm human operators. At least thirty countries have systems like this today, but they are used relatively narrowly. They are used for fixed defense of installations or onboard defense of human-occupied vehicles. They have a person "on the loop" who can supervise their operation and intervene in the case the system malfunctions. And the human operators have physical access to a system. If it begins to malfunction and it does not respond to their software commands to cease operation, they can physically disable the system. Weapons of this type, which are a very limited kind of autonomous weapon, have been in use for decades. Moreover, they are likely to be increasingly important as precision-guided weapons proliferate, necessitating autonomous defenses to protect military assets and civilian populations. While these systems have been used narrowly, it is possible to envision autonomous weapons being used in other contexts because of their advantages in speed, particularly in cyberspace.

The second reason militaries might to pursue autonomous weapons is to continue operations with unmanned vehicles in communications-degraded or –denied environments. In war, militaries will seek to jam or disrupt one another's communications, and some environments, such as undersea, are intrinsically difficult for communications. This is problematic for unmanned vehicles where there is no person onboard. Now, not all autonomous systems are autonomous weapons. Even without communications, militaries could operate unmanned vehicles that conduct surveillance or even strike pre-approved fixed targets, much like cruise missiles today, all still while keeping a person "in the loop" for which specific targets to engage. Militaries could also keep a human "in the loop" to authorize emergent targets of opportunity with relatively low bandwidth, on the order of a 56K dial-up modem from the 1990s. But if there was no communications at all and militaries wanted to hunt and strike mobile targets with unmanned vehicles, or give an unmanned vehicle the ability to defend itself even if it didn't have any communications with human controllers, then autonomous weapons might be desirable.

And finally, countries may want autonomous weapons simply due to the fear that others might develop them. While no nation has said that they are building autonomous weapons, the perception that they might be valuable could spur militaries to race ahead to build them, simply out of the fear of being left behind.

Now, autonomous weapons raise many concerns. These include legal, moral, ethical, and strategic stability concerns. As we begin to mature our understanding of the issues raised by LAWS, I would suggest that we need to begin refining these issues and parsing them more clearly.

Specifically, some issues surrounding LAWS are about whether or not they could be used in ways that comply with international humanitarian law principles, such as proportionality, distinction, and others. Other issues regarding LAWS would be issues that would arise if LAWS **could** be used in compliance with IHL. That is, some issues are not legal issues at all, but are rather moral, ethical, or strategic stability concerns. Just because these are not legal issues does not mean they are unimportant, however. At the same time, just because an issue is important or just because LAWS raise a concern does not mean that it is a legal issue. Some things are legal, but immoral. Some things are legal and moral, but unwise.

There are also important distinctions between different types of LAWS. Some issues apply primarily to anti-personnel LAWS, weapons that would target people. Others would apply to any kind of LAWS, including anti-vehicle or anti-materiel LAWS.

Similarly, some issues depend on where LAWS are used. If one is principally concerned about civilian casualties, then the use of LAWS in areas where civilians are not present, such as undersea or in space, is perhaps less of a concern. If one is, however, concerned also about the strategic stability implications of LAWS, then weapons in space, cyberspace, or undersea could be quite troubling.

Some issues are unique to LAWS, others are not unique to LAWS but are exacerbated by autonomy, and still others apply to LAWS as well as any other military weapon. We should be clear as we discuss these topics to distinguish among them.

Finally, many of the discussions on LAWS to-date look at LAWS principally from a humanitarian dimension. That is, they examine how LAWS would change the conduct of war. These are important concerns, but LAWS also raise potentially serious questions about crisis stability and the conditions that could start a war. This is especially the case in situations where multiple LAWS might be interacting in uncertain environments and at high speeds.

On May 6, 2010, the U.S. stock market experienced a "flash crash" where it lost nearly 10% of its value in just a few minutes. This was caused by an automated stock trade interacting with high-frequency trading algorithms at speeds too fast for humans to intervene. Similarly, autonomous weapon systems interacting with an uncontrolled environment and with each other could lead to unanticipated outcomes. This could be exacerbated by cyber hacking, tricking an autonomous weapon with false data, human error, or malfunctions. Because countries are not likely to share their algorithms with one another, we need to discuss how countries

might develop “rules of the road” for how we might incorporate autonomy in weapons to avoid such an outcome. A “flash war” would benefit no one.

I hope that delegates will consider these issues during the course of this week as the discussion develops. And I thank delegates for meeting for a second round of discussions this year on this important topic. This is a challenging issue, one that we are still grappling to understand, and we have much to discuss.

Thank you.