Why You Should Fear “Slaughterbots”—A Response

Lethal autonomous weapons are not science fiction; they are a real threat to human security that we must stop now

By Stuart Russell, Anthony Aguirre, Ariel Conn and Max Tegmark

This is a guest post. The views expressed here are solely those of the authors and do not represent positions of IEEE Spectrum or the IEEE.

Paul Scharre’s recent article “Why You Shouldn’t Fear ‘Slaughterbots’” dismisses a video produced by the Future of Life Institute, with which we are affiliated, as a “piece of propaganda.” Scharre is an expert in military affairs and an important contributor to discussions on autonomous weapons. In this case, however, we respectfully disagree with his opinions.

Why we made the video

We have been working on the autonomous weapons issue for several years. We have presented at the United Nations in Geneva and at the World Economic Forum; we have written an open letter signed by over 3,700 AI and robotics researchers and over 20,000 others and covered in over 2,000 media articles; one of us (Russell) drafted a letter from 40 of the world’s leading AI researchers to President Obama and led a delegation to the White House in 2016 to discuss the issue with officials from the Departments of State and Defense and members of the National Security Council; we have presented to multiple branches of the armed forces in the United States and to the intelligence community; and we have debated the issue in numerous panels and academic fora all over the world.

Because autonomous weapons do
Our primary message has been consistent: Because they do not require individual human supervision, autonomous weapons are potentially scalable weapons of mass destruction (WMDs); essentially unlimited numbers can be launched by a small number of people. This is an inescapable logical consequence of autonomy. As a result, we expect that autonomous weapons will reduce human security at the individual, local, national, and international levels.

Despite this, we have witnessed high-level defense officials dismissing the risk on the grounds that their “experts” do not believe that the “Skynet thing” is likely to happen. Skynet, of course, is the fictional command and control system in the Terminator movies that turns against humanity. The risk of the “Skynet thing” occurring is completely unconnected to the risk of humans using autonomous weapons as WMDs or to any of the other risks cited by us and by Scharre. This has, unfortunately, demonstrated that serious discourse and academic argument are not enough to get the message through. If even senior defense officials with responsibility for autonomous weapons programs fail to understand the core issues, then we cannot expect the general public and their elected representatives to make appropriate decisions.

The main reason we made the video, then, was to provide a clear and easily understandable illustration of what we mean. A secondary reason was to give people a clear sense of the kinds of technologies and the notion of autonomy involved: This is not “science fiction”; autonomous weapons don’t have to be humanoid, conscious, and evil; and the capabilities are not “decades away” as claimed by some countries at the U.N. talks in Geneva. Finally, we are mindful of the precedent set by the ABC movie The Day After in 1983, which, by showing the effects of nuclear war on individuals and families, had a direct effect on national and international policy.

Where we agree

Scharre agrees with us on the incipient reality of the technology; he writes, “So while no one has yet cobbled the technology together in the way the video depicts, all of the components are real.” He concludes that terrorist groups will be able to cobble together autonomous weapons, whether or not such weapons are subject to an international arms control treaty. This is probably true at a small scale; but at a small scale, there is no great advantage to terrorists in using autonomy. It is almost certainly false at a large scale. It is extremely unlikely that terrorists would be able to design and manufacture thousands of effective autonomous weapons without detection—especially if the treaty verification regime, like the Chemical Weapons Convention, mandates the cooperation of manufacturers that produce drones and other precursor components.
We concur with Scharre on the importance of countermeasures, while noting that a ban on lethal autonomous weapons would certainly not preclude the development of antidrone weapons.

Finally, we agree with Scharre that the stakes are high. He writes, “Autonomous weapons raise important questions about compliance with the laws of war, risk and controllability, and role of humans as moral agents in warfare. These are important issues that merit serious discussion.” It is puzzling, however, that he does not consider the issue of WMDs to merit serious discussion.

Where we disagree

Scharre attributes four claims to us and then attempts to refute them. To make things less confusing, we will negate those four claims to produce four statements that Scharre is effectively asserting in his article (the exact wording of these assertions has been confirmed in subsequent correspondence with Scharre):

1. Scharre: Governments are unlikely to mass-produce lethal micro-drones to use as weapons of mass destruction.

One might ask, “In that case, why not ban them?” Prior to the entry into force of the Chemical Weapons Convention in 1997, the major powers did mass-produce lethal chemical weapons including various kinds of nerve gas, for use as weapons of mass destruction. After they were banned, stockpiles were destroyed and mass production stopped. Banning lethal autonomous micro-drones would criminalize their production as well as their use as WMDs, making it much less likely that terrorists and others would be able to access large quantities of effective weapons.

There is some reason to believe, however, that the claim is simply not true. For example, lethal micro-drones such as the Switchblade are already in mass production. Switchblade, a fixed-wing drone with a 0.6-meter wingspan, is designed as an anti-personnel weapon. Contrary to Scharre’s claim, they can easily be repurposed to kill civilians rather than soldiers. Moreover, Switchblade now comes with a “Multi-Pack Launcher.” Orbital ATK, which makes the warhead, describes the Switchblade as “fully scalable.”
Switchblade is not fully autonomous and requires a functioning radio link; the DoD’s CODE (Collaborative Operations in Denied Environments) program aims to move towards autonomy by enabling drones to function with at best intermittent radio contact; they will “hunt in packs, like wolves” according to the program manager. Moreover, in 2016, the Air Force successfully demonstrated the in-flight deployment of 103 Perdix micro-drones from three F/A-18 fighters. According to the announcement, “Perdix are not pre-programmed synchronized individuals, they are a collective organism, sharing one distributed brain for decision-making and adapting to each other like swarms in nature.” While the Perdix drones themselves are not armed, it is hard to see the need for 103 drones operating in close formation if the purpose for such swarms were merely reconnaissance.

Under pressure of an arms race, one can expect such weapons to be further miniaturized and to be produced in larger numbers at much lower cost. Once autonomy is introduced, a single operator can deploy thousands of Switchblades or other lethal micro-drones, rather than piloting a single drone to its target. At that point, production numbers will ramp up dramatically.

In the major wars of the 20th century, over 50 million civilians were killed. This horrific record suggests that, in an armed conflict, nations will not refrain from large-scale attacks. And, as WMDs, scalable autonomous weapons have advantages for the victor compared to nuclear weapons and carpet bombing: They leave property intact and can be applied selectively to eliminate only those who might threaten an occupying force.

“As WMDs, autonomous weapons have advantages for the victor compared to nuclear weapons and carpet bombing: They leave property intact and can be applied selectively to eliminate only those who might threaten an occupying force”

2. Scharre: Nations are likely to develop effective countermeasures to micro-drones, especially if they become a major threat.

While Scharre’s article attributes to us the claim “There are no effective defenses against lethal micro-drones,” he effectively concedes that the claim is true as things stand today. His own position is that the situation depicted in the video, where mass-produced antipersonnel weapons are available but no effective defenses have been developed, could not occur, or could occur only as a temporary imbalance.
Scharre cites as evidence for this claim a New York Times article. The article does not exactly inspire confidence: It describes the problem of lethal micro-drones as “one of the Pentagon’s most vexing counterterrorism conundrums.” It describes as “decidedly mixed” the results from DoD’s Hard Kill Challenge, which aims to see “which new classified technologies and tactics proved most promising.” The DoD’s own conclusion? “Bottom line: Most technologies still immature.” The Hard Kill Challenge is the successor to the Black Dart program, which ran annual challenges beginning in 2002. After more than 15 years, then, we still have no effective countermeasures.

Scharre states that lethal autonomous micro-drones “could be defeated by something as simple as chicken wire,” perhaps imagining entire cities festooned with it. If this were a workable form of defense, of course, then there would be no Hard Kill Challenge; Switchblades would be useless; and Iraqi soldiers wouldn’t be dying from attacks by lethal micro-drones.

Scharre notes correctly that the video shows larger drones blasting through walls, but he obviously failed to notice that the family home in the video is encased in a steel grille—as are parts of the university dorm, which is plastered with “safe zone” signs directing students in case of drone attack. Scharre claims that the attack/defense cost ratio favors the defender, but this seems unlikely if one needs to be 100 percent protected, 100 percent of the time, against an attack that can arrive anywhere. When the weapons are intelligent, one hole in the defensive shell is enough. Adding more defensive shells makes little difference.

Moreover, the weapons are cheap and expendable, as Scharre correctly points out in a recent interview: “The key is not just finding a way to target these drones. It’s finding a way to do it in a cost-effective way. If you shoot down a $1,000 drone with a $1 million missile, you’re losing every time you’re doing it.” We agree. This doesn’t sound like a ratio that favors the defender.

As to whether we should have complete confidence in the ability of governments or defense corporations to develop, within a short time frame, cheap, effective, wide-area defenses against lethal micro-drones: We are reminded of the situation of the British population in the early days of World War II. One would think that if anyone had a motive to develop effective countermeasures, it would be the British during the Blitz. But, by the end of the Blitz, after 40,000 bomber sorties against major cities, countermeasures were no more than 1.5 percent effective—even lower than at the beginning, 9 months earlier.
3. Scharre: Governments are capable of keeping large numbers of military-grade weapons out of the hands of terrorists.

According to Scharre, the video shows “killer drones in the hands of terrorists massacring innocents.” In fact, as the movie goes to great lengths to explain, the perpetrators could be “anyone,” not necessarily terrorists. Attacks by autonomous weapons will often be unattributable and can therefore be carried out with impunity. (For this reason, governments around the world are extremely concerned about assassination by autonomous weapon.) In the movie, the most likely suspects are those involved in “corruption at the highest level,” i.e., persons with significant economic and political power.

Scharre writes, “We don’t give terrorists hand grenades, rocket launchers, or machine guns today.” Perhaps not, except when those terrorists were previously designated as freedom fighters—but there is no shortage of effective lethal weaponry on the market. For example, there are between 75 and 100 million AK-47s in circulation, the great majority outside the hands of governments. Roughly 110,000 military AK-47s went missing in a two-year period in Iraq alone.

“Attacks by autonomous weapons will often be unattributable and can therefore be carried out with impunity, and governments around the world are indeed concerned about assassination by autonomous weapon”

Produced in large quantities by commercial manufacturers, lethal autonomous micro-drones would probably be cheaper to buy than AK-47s. And much cheaper to use: They don’t require a human to be trained, housed, fed, equipped, and transported in order to wield lethal force. The ISIS budget for 2015 was estimated to be US $2 billion, probably enough to buy millions of weapons if they were available on the black market.

Image: Slaughterbots/YouTube

The creators of the “Slaughterbots” video argue that autonomous weapons like lethal micro-drones could be used not only by terrorists but anyone with significant economic and political power.

4. Scharre: Terrorists are incapable of launching simultaneous coordinated attacks on the scale shown in the video.
As noted above, the attack in the video against several universities was carried out not by terrorists but by unnamed persons in high-level positions of power. (We considered showing a mass attack against a city occurring as part of a military campaign by a nation-state, but we decided that appearing to accuse any particular nation of future war crimes would not be conducive to diplomacy.) No matter who might wish to perpetrate mass attacks using autonomous weapons, their job will be made far more difficult if arms manufacturers are legally banned from making them.

It’s also important to understand the difference that autonomy makes for the ability of non-state actors to carry out large-scale attacks. While coordination across multiple geographical locations is unaffected, the scale of each attack can be immeasurably greater. (We note that Scharre misconstrues the video on this point. He sees only 50 drones emerge from the van, whereas in fact most of the larger drones are carriers for multiple, shorter-range lethal micro-drones that are deployed automatically in the final moments of the attack. One van per university suffices, so the movie implies coordination across 12 locations—not so different from the 10 locations described as feasible in the article cited by Scharre.)

Whereas a nation-state can, in principle, launch attacks with thousands of tanks or aircraft (remotely piloted or otherwise) or tens of thousands of soldiers, such scale is possible for non-state actors only if they use autonomous weapons. Thus, an arms race in autonomous weapons shifts power from nation-states—which are largely constrained by the international system of treaties, trade dependencies, etc.—to non-state actors, who are not.

Perhaps this is what Scharre is referring to in the interview cited above, when he says, “We’re likely to see more attacks of larger scale going forward, potentially even larger than this and in a variety of things—air, land, and sea.”
In summary, we, and many other experts, continue to find plausible the view that autonomous weapons can become scalable weapons of mass destruction. Scharre’s claim that a ban will be ineffective or counterproductive is inconsistent with the historical record. Finally, the idea that human security will be enhanced by an unregulated arms race in autonomous weapons is, at best, wishful thinking.

Stuart Russell is a professor of computer science at the University of California, Berkeley, and coauthor of the standard textbook “Artificial Intelligence: A Modern Approach.”

Anthony Aguirre is a professor of physics at the University of California, Santa Cruz, and cofounder of the Future of Life Institute.

Ariel Conn oversees media and outreach for the Future of Life Institute.

Max Tegmark is a professor of physics at MIT, cofounder of the Future of Life Institute, and author of “Life 3.0: Being Human in the Age of Artificial Intelligence.”

### Robotics News

Biweekly newsletter on advances and news in robotics, automation, control systems, interviews with leading roboticists, and more.

### About the Automaton blog

IEEE Spectrum's award-winning robotics blog, featuring news, articles, and videos on robots, humanoids, automation, artificial intelligence, and more.

Erico Guizzo, Editor

Evan Ackerman, Senior Writer

Fan Shi, Contributor

Subscribe to RSS Feed