Routledge
Taylor & Francis Group

Survival
Global Politics and Strategy

ISSN: 0039-6338 (Print) 1468-2699 (Online) Journal homepage: http://www.tandfonline.com/loi/tsur20

Autonomous Weapon Systems and Strategic Stability

Jürgen Altmann & Frank Sauer

To cite this article: Jürgen Altmann & Frank Sauer (2017) Autonomous Weapon Systems and Strategic Stability, Survival, 59:5, 117-142, DOI: 10.1080/00396338.2017.1375263

To link to this article: http://dx.doi.org/10.1080/00396338.2017.1375263

Autonomous Weapon Systems and Strategic Stability

Jürgen Altmann and Frank Sauer

In July 2015, an open letter from artificial-intelligence experts and roboticists called for a ban on autonomous weapon systems (AWS), comparing their revolutionary potential to that of gun powder and nuclear weapons.[1] According to a 2012 Pentagon directive, AWS are weapon systems which, 'once activated … can select and engage targets without further intervention by a human operator'.[2] Proponents of AWS have suggested that they could offer various benefits, from reducing military expenditure to ringing in a new era of more humane and less atrocious warfare. By contrast, critics – some characterising AWS as 'killer robots'[3] – expect the accompanying political, legal and ethical risks to outweigh these benefits, and thus argue for a preventive prohibition.

AWS are not yet operational, but decades of military research and development, as well as the growing technological overlap between the rapidly expanding commercial use of artificial intelligence (AI) and robotics, and the accelerating 'spin-in' of these technologies into the military realm, make autonomy in weapon systems a possibility for the very near future. Military programmes adapting key technologies and components for achieving autonomy in weapon systems, as well as the development of prototypes and doctrine, are well under way in a number of states.

Jürgen Altmann is a lecturer in experimental physics at Technical University of Dortmund, working on the prospective assessment of new military technologies and the analysis of preventive arms-control measures. **Frank Sauer** is a senior research fellow and lecturer in international relations at Bundeswehr University in Munich, working on international security and arms control. He is the author of *Atomic Anxiety: Deterrence, Taboo, and the Non-Use of U.S. Nuclear Weapons* (Palgrave Macmillan, 2015). Both authors are members of the International Committee for Robot Arms Control (ICRAC).

Survival | vol. 59 no. 5 | October–November 2017 | pp. 117–142 DOI 10.1080/00396338.2017.1375263

Accompanying this work is a rapidly expanding body of literature on the various technical, legal and ethical implications of AWS. However, one particularly crucial aspect has – with exceptions confirming the rule[4] – received comparably little systematic attention: the potential impact of autonomous weapon systems on global peace and strategic stability.

By drawing on Cold War lessons and extrapolating insights from the current military use of remotely controlled unmanned systems, we argue that AWS are prone to proliferation and bound to foment an arms race resulting in increased crisis instability and escalation risks. We conclude that these strategic risks justify a critical stance towards AWS.

Defining the debate

It is worth noting that some weapon systems, so far used only for defensive purposes, have long been able to identify, track and engage incoming targets on their own. These systems can already be set up so that humans are cut out of decision-making, a capability deemed necessary because there can be instances in which there is not enough time for humans to react, as during attacks with missiles or mortar shells.

These defensive weapons are stationary or fixed on ships or trailers, and are designed to fire at inanimate targets. They repeatedly perform pre-programmed actions within tightly set parameters and time frames in comparably structured and controlled environments. Consequently, they are commonly thought to be only the precursors to AWS, and might be described as *automatic*, as distinct from the *autonomous* systems currently being developed. The latter will be able to operate without human control or supervision in dynamic, unstructured, open environments, attacking various sets of targets, including inhabited vehicles, structures or even individuals. They will operate over an extended period of time after activation – and will potentially be able to learn and adapt their behaviour.

It can be difficult, however, to differentiate between automatic and autonomous systems in practice, with many systems falling into a considerable grey area. Autonomous functionality in weapon systems develops over a continuum. Some advanced 'automatic' systems are already behaving in ways that might be considered autonomous – for instance, when

automatically (autonomously?) targeting the source of incoming fire. Such systems also blur the line between 'defensive' and 'offensive'. Nevertheless, juxtaposing automatic and autonomous systems is a helpful mental exercise to grasp what AWS are going to be like, and what benefits they, according to their proponents, will provide.

Such benefits include the possibility that new systems will combine superior performance with lower costs due to a reduced need for personnel. Moreover, AWS are said to render constant control and communication links obsolete. Daisy-chained, line-of-sight connections can already allow for control and communication without necessarily revealing a system's location. But dispensing with a communications link altogether could offer even stronger insurance against communications disruption or hijacking. Much more importantly, being able to do without an up- and downlink removes the inevitable delay between the human operator's command and the system's response, thus generating a clear tactical advantage over a remotely controlled, 'slower' adversarial system. Finally, some proponents hope that, since AWS experience neither fear nor stress, and do not overreact, they might render warfare more humane and prevent some of the atrocities of war. Not only are machines devoid of negative human emotions, they also lack a self-preservation instinct, so they could well delay returning fire, it is argued. They are supposed to allow not only for greater restraint but better discrimination between civilians and combatants, resulting in an application of force that accords with international humanitarian law.[5]

Critics counter that militarised AI systems are – and for the foreseeable future will be – incapable of distinguishing between combatants and civilians, as well as being unable to assure a proportionate application of military force, which renders the battlefield use of AWS illegal.[6] Also, should an autonomous weapon system nevertheless be fielded and end up causing disproportionate loss of life among (or injury to) civilians, or damage to civilian objects, it is unclear who might be held legally responsible, since machines can obviously not be court-martialled.[7]

Critics concerned with the ethical, rather than legal, implications of AWS argue that such systems are intrinsically amoral because delegating kill decisions to an algorithm in a machine – which is not accountable for its actions

in any meaningful ethical sense – infringes on fundamental human values including dignity and the right to life.[8] Such humanitarian concerns are also reflected in public opinion. Representative polling data suggests that a majority of US citizens oppose the use of AWS, with 40% 'strongly opposing' them.[9] An online poll conducted by the Open Roboethics Initiative in 14 different languages supports these findings at the global level.[10]

Finally, operational risks are cause for concern. For instance, the potential of AWS for high-tempo fratricide, way beyond the speed of human intervention, incentivises militaries to avoid full autonomy in weapon systems, and instead to retain humans in the chain of decision-making as a fail-safe mechanism.[11] We argue that concerns of this nature are relevant not just at the operational level, but point to the potentially detrimental impact of AWS on overall strategic stability.

Two dimensions of instability

The goal of upholding stability to prevent a catastrophic nuclear war was a central feature of the Cold War. Destabilisation loomed with the arms build-up, in particular with the development of ballistic missiles carrying multiple independently targetable re-entry vehicles (MIRVs), and of missile defence. The former dramatically increased fears of a first strike and thus the pressure to launch on warning, that is, before the arrival of enemy warheads 10–30 minutes later. 'Accidental nuclear war' scares, fuelled by human and technical errors in early-warning systems, informed the decisions to limit anti-ballistic-missile systems and to preferentially reduce MIRVed missiles and warhead counts.[12] The goal of stability was also taken up in the realm of conventional military armaments, mainly in the Treaty on Conventional Armed Forces in Europe (CFE Treaty).[13]

The lessons of the Cold War are worth remembering. They suggest that instability has two dimensions. The first encompasses military instability with regard to the proliferation of arms and the emergence of arms races. During the Cold War, the perceived risk of 'horizontal proliferation' – that is, the spread of nuclear weapons beyond the existing nuclear-weapons states – gave rise to the Non-Proliferation Treaty and various export-control regimes. The risk of vertical proliferation – that is, an uncontrolled build-up

of arms that drives up military expenditure and exacerbates the security dilemma, thus increasing the likelihood of crises – was reflected in the various strategic arms-limitation and -reduction agreements between the US and the Soviet Union. As the US Office of Technology Assessment (OTA) put it in 1985,

> Arms race stability involves the effect of planned deployments on the scope and pace of the arms race … If a deployment on one side is likely to lead to a responding deployment on the other side which is in turn likely to induce a still higher level of deployment on the first side, the first side's deployment might be seen as 'destabilizing' the arms competition.[14]

Generally speaking, any quantitative or qualitative arms race between – in this example – two potential adversaries involves an element of instability. But a race's pace can vary widely. Destabilisation becomes a particular concern when qualitatively new technologies promising clear military advantages seem close at hand. When potential adversaries make special efforts to get ahead themselves, or at least to avoid falling behind, this can trigger a dynamic intensified by mutual observation of – as well as speculation in light of uncertainty about – the other side's advances. If the situation is perceived as urgent, and precedents have been or are about to be set, there are compelling incentives for accelerating the development of technology and incorporating it into militaries, a process that is then more likely to outpace and render moot any attempt at agreement on mutual, preventive prohibitions.

The second dimension of strategic instability is crisis instability and escalation, either across the threshold from peace to war, or, when war has already broken out, to a higher level of violence – in particular from conventional to nuclear weapons. With respect to nuclear weapons, crisis stability during the Cold War was seen, according to the OTA, as

> the degree to which strategic force characteristics might, in a crisis situation, reduce incentives to initiate the use of nuclear weapons … Weapon systems are considered destabilizing if in a crisis they would

> add significant incentives to initiate a nuclear attack, and particularly to attack quickly before there is much time to collect reliable information and carefully weigh all available options and their consequences.[15]

In terms of conventional forces, the preamble of the CFE Treaty encompasses crisis stability in its commitment to 'establishing a secure and stable balance of conventional forces at lower levels … eliminating disparities detrimental to stability and security [and] eliminating … the capability for launching surprise attack and for initiating large-scale offensive action in Europe'.[16]

Both dimensions are closely connected. New kinds of weapons, developed as an outcome of an arms race, can increase crisis instability, with MIRVed missiles being a prominent Cold War example. And (perceived) crisis instability can create motives for diversifying weapon carriers and fuel the arms race in turn, as the development of nuclear submarines demonstrates.

Proliferation and arms-race instability

As early as 2007, the US Department of Defense wrote in its Unmanned Systems Roadmap that for processor technology 'the ultimate goal is to replace the operators with a mechanical facsimile [of] equal or superior thinking speed, memory capacity, and responses gained from training and experience'. The document also stated that the 'primary technical challenges for weapon release from unmanned systems include the ability to reliably target the right objective'.[17] The goal of weapon autonomy pervades all subsequent road maps.[18] Autonomous weapon-system functions have since been tested on land, under water, on the sea and, most notably, in the air. In fact, current trends with respect to unmanned combat aerial vehicles (UCAVs or 'combat drones') provide indicators for what to expect with regard to AWS. Unlike today's high-profile UCAVs, such as the *Reaper*, which are propeller driven, slow, carry comparably small payloads and have few to no capabilities for operating in contested airspace, future systems will be less dependent on human control, faster, stealthy and capable of delivering bigger payloads.

The X-47B, for instance, has demonstrated autonomous take-off from and landing on an aircraft-carrier deck, as well as autonomous aerial refuelling.

This technology demonstrator was developed by the US Navy's Unmanned Carrier-Launched Airborne Surveillance and Strike programme (UCLASS). Similarly, the British *Taranis* UCAV was described by the UK Ministry of Defence as 'fully autonomous' and able to 'defend itself against manned and other unmanned enemy aircraft' with 'almost no need for operator input'.[19] However, the ministry also stated that 'the operation of weapons systems will always be under human control'.[20]

While AWS test beds such as *Taranis* and the X-47B rely on familiar designs, in this case the airframes of a fast, stealthy, next-generation drone with substantial payload capabilities, future systems will display an autonomous swarming capability, and thus AWS will also come in much smaller sizes. In October 2016, for instance, the US Department of Defense demonstrated a swarm of 103 *Perdix* micro drones capable of 'advanced swarm behaviors such as collective decision-making, adaptive formation flying, and self-healing'.[21] In the future, such micro drones are to be 3D printed in large batches and deployed from (manned) flying systems. This dispensing method has already been successfully tested at Mach 0.6 speed by two F/A-18 *Super Hornets* releasing a *Perdix* drone swarm. The US Navy's LOCUST programme is also seeking to develop swarming, disposable unmanned aerial vehicles (UAVs).[22]

The overall goal for this new ecosystem of flying assets is to replace not just the old generation of drones but also manned aircraft, thus continuing the trend towards keeping human pilots out of harm's way and providing superior unmanned air-to-ground *and* air-to-air capabilities across the board.[23] In air-to-air combat, the big, fast autonomous drones currently envisioned will be able to fly high-g manoeuvres no human pilot would be able to endure. More importantly, they would ensure much shorter reaction times. On-board sensors combined with artificial 'intelligence' – either located onboard or distributed in the swarm and based on decision-making algorithms endowed with the authority to initiate an attack without awaiting human input – are to make these weapons autonomous and hence provide a decisive edge over remotely controlled and human-piloted adversary systems alike.

While the development of AWS is currently most advanced in the air and under water – that is, in less cluttered environments – the example of

autonomous (swarms of) UCAVs demonstrates the generally valid proposition that for future unmanned systems, operational speed will reign supreme, regardless of the domain. In that sense, technological developments in AI and robotics, as well as current expectations regarding future armed conflict (and the need for speed), jointly point towards AWS. In fact, US deputy secretary of defense Bob Work stated in March 2016 that even the final delegation of lethal authority to autonomous systems will inexorably happen as a result of this race for speed.[24] According to Work, the United States 'will not delegate lethal authority for a machine to make a decision … The only time we'll delegate authority is in things that go faster than human reaction time, like cyber or electronic warfare.' Yet, he conceded that such self-restraint may be unsustainable if an authoritarian rival acts differently. 'We might be going up against a competitor who is more willing to delegate authority to machines than we are and, as that competition unfolds, we'll have to make decisions on how we can best compete', Work said. 'It's not something that we have fully figured out, but we spend a lot of time thinking about it.'[25]

Operational speed will reign supreme

To further deepen our understanding of AWS, it is useful to take a step back and underline that they need not necessarily take the shape of a specific weapon system akin to, for instance, a drone or a missile. AWS also do not require a specific military-technology development path, the way nuclear weapons do, for example. As AI, autonomous systems and robot technologies mature and begin to pervade the civilian sphere, militaries will increasingly be able to make use of them for their own purposes, as the development of information and communication technology suggests. Naturally, any military adaptation of a dual-use technology will need to fulfil specific military requirements that do not exist in a civilian environment, or are less relevant for mass markets. Nevertheless, AWS development will profit from the implementation or mirroring of a variety of civilian technologies (or derivatives thereof) and their adoption for military purposes, technologies which are currently either already available or on the cusp of becoming ready for series production in the private sector. This trend is already observable in the case of armed drones. Light detection and ranging (LIDAR) systems

are another example. These are the optical sensors used by the automotive industry to give self-driving cars a 360-degree picture of their surroundings. LIDAR prices have recently dropped from five figures to a few hundred dollars. The units have also become more rugged and much smaller.[26] Given that these components, which are necessary for endowing mobile systems with autonomy, are now cheaply and readily available off the shelf, there is every reason to expect the military to adapt, and, if required, adjust and refine, them for their own purposes.[27]

It is clear that the research and development for AWS-relevant technology is well under way and distributed across countless university laboratories and, especially, commercial enterprises that are making use of economies of scale and the forces of the free market to spur competition, lower prices and shorten innovation cycles. This renders the military research and development effort in the case of AWS different from those of past high-tech conventional weapon systems (the F-35 comes to mind), let alone nuclear weapons. So while the impact of AWS might be revolutionary in terms of their implications for warfare, their development within the context of the military is best described as evolutionary: the military is merely continuing and, with outside help and technology lifted from the private sector, accelerating an already existing trend to replace labour with capital and automate dull, dirty and dangerous military tasks.[28] For example, former secretary of defense Ashton Carter sought closer ties with Silicon Valley to hasten the incorporation of technological innovations into the US military after the US officially declared AI and robotics cornerstones of its new 'third offset' strategy to counter rising powers.[29]

Thus, AWS are easy to obtain compared with other paradigm-shifting weapons, such as nuclear weapons, which even now require the Herculean effort of a state-run, focused politico-military effort to produce. AWS do not require ores, centrifuges, high-speed fuses or other comparably 'exotic' components to be assembled and tested in a clandestine manner. Consequently, while nuclear technologies can be – and are – proliferation controlled, AWS are much harder to regulate. With comparatively fewer choke points that might be targeted by non-proliferation policies, AWS are potentially available to a wide range of state and non-state actors, not just those nation-states

that are willing and able to muster the considerable resources needed for the robotic equivalent of the *Manhattan Project*.[30] This carries significant implications for arms control.

There will of course be differences in quality. Sophisticated AWS will have to meet the same or similar military standards that current weapon systems, such as main battle tanks or combat aircraft, do. Moreover, technologically leading nations such as the US and Israel are carrying out research to produce autonomous systems that comply with international humanitarian law. Less scrupulous actors, however, will find AWS development much easier. Comparably crude AWS which do not live up to the standards of a professional military in terms of reliability, compliance with international humanitarian law or the ability to go head-to-head with systems of a near-peer competitor could, in fact, be put together with technology available today by second- or third-tier state actors, and perhaps even non-state actors. Converting a remotely controlled combat drone to autonomously fire a weapon in response to a simple pattern-recognising algorithm is already doable. Even the technological edge displayed by sophisticated AWS is unlikely to be maintained over the longer term. While sensor and weapon packages to a large degree determine the overall capabilities of a system, implementing autonomy ultimately comes down to software, which is effortlessly copied and uniquely vulnerable to being stolen via computer-network operations. Thus, while the development of AWS clearly presents a challenge to less technologically advanced actors, obtaining AWS with some degree of military capability is a feasible goal for any country already developing, for example, remotely controlled armed UAVs – the number of which rose from two to ten between 2001 and 2016.[31] Admittedly, the US and Israel are still in the lead with regard to developing unmanned systems and implementing autonomous-weapon functionality – China only recently test-fired a guided missile from a drone via satellite link for the first time.[32] But considering that drone programmes can draw from the vibrant global market for unmanned aerial vehicles of all shapes and sizes,

Implementing autonomy comes down to software

the hurdles regarding AWS are much lower than those of other potentially game-changing weapons of the past.

Proliferation of AWS could of course also occur via exports, including to the grey and black markets. In this way, autonomous systems could fall not only into the hands of technologically inferior state actors, but also those of non-state actors, including extremist groups. Hamas, Hizbullah and the Islamic State have already deployed and used armed drones. As sensors and electronics are increasingly miniaturised, small and easily transportable systems could be made autonomous with respect to navigation, target recognition, precision and unusual modes of attack.[33] Terrorist groups could also gain access to comparably sophisticated systems that they could never develop on their own. Again, autonomy in this context does not necessarily require military-grade precision – a quick and dirty approach would suffice for these actors. In fact, it stands to reason that terrorist groups would use autonomous killing capabilities indiscriminately in addition to using them, if possible, in a precise fashion for targeted assassinations.

It is still unclear how the development of unmanned systems on the one hand and specific countermeasures on the other will play out. Traditional aircraft-sized drones such as the X-47B or *Taranis*, to stick with these examples, are obviously susceptible to existing anti-aircraft systems. As for smaller-sized systems, various tools, from microwaves to lasers to rifle-sized radio jammers for disrupting the control link, are currently being developed as countermeasures. Simpler, less exotic methods such as nets, fences or even trained hunting birds might also prove effective for remotely controlled and autonomous systems alike. It is clear, however, that saturation attacks have been identified as a key future capability for defeating a wide range of existing and upcoming defensive systems – both human-operated and automatic.[34] The latter are a particular focus of research into swarming as a potential solution.[35] And military systems operating at very high speeds and in great numbers or swarms are bound to generate new instabilities, to which we will turn in our next section.

To first sum up our argument so far, there are obvious dual-use problems and an unusually high risk of proliferation when it comes to AWS. Should one of the technologically leading nation-states go forward with the

deployment of AWS, it would be comparably easy – and thus very likely – that others would follow suit.[36] In that sense, the development of AWS could well trigger a destabilising arms race.

Crisis instability and escalation

Increasing operational speeds mean that human involvement in AWS would be limited to, at best, general oversight and decision-making in instances where communication delays of up to a few seconds – and thinking and deliberation times of a few minutes – could be deemed acceptable, meaning they would not result in defeat or the loss of systems. Many situations would not allow for the luxury of human pondering, however. In such cases, the actions and reactions of individual AWS, as well as AWS swarms, would have to be controlled autonomously by algorithms – in other words determined only by programming software in advance and possibly through the adaptation and learning of the systems themselves. After all, as Paul Scharre put it, 'winning in swarm combat may depend upon having the best algorithms to enable better coordination and faster reaction times, rather than simply the best platforms'.[37]

One such swarm-combat situation could be a severe political crisis in which adversaries believe that war could break out. With swarms deployed in close proximity to each other, control software would have to react to signs of an attack within a split-second time frame – by evading or, possibly, counter-attacking in a use-them-or-lose-them situation. Even false indications of an attack – sun glint interpreted as a rocket flame, sudden and unexpected moves of the adversary, or a simple malfunction – could trigger escalation.

The nature of military conflict is such that these kinds of interactions could not be tested or trained for beforehand. In addition, it is, technically speaking, impossible to fathom all possible outcomes in advance. Clearly, the interaction of swarms, if fully autonomous, would be unpredictable, and could potentially result in an escalation from crisis to war, or, within armed conflict, to higher levels of violence. This is not a theoretical proposition deduced solely from systems theory and the argument of unavoidable 'normal accidents'.[38] On the contrary, comparable runaway interactions

between algorithms are already happening in the civilian sphere on a regular basis. In April 2011, the price of an out-of-print biology textbook rose within weeks to $23.7 million on the Amazon marketplace due to the price-setting algorithms of two vendors interacting with each other.[39] Eventually one of the vendors intervened; no damage was done because nobody purchased the book at this absurd price. Greater havoc was caused in the New York Stock Exchange 'flash crash' of 6 May 2010 in which computerised high-frequency trading played an essential role, and during which stock indices and important industry stocks collapsed.[40] In this case 'circuit breakers' established by monitoring authorities set in, suspending high-speed trading and preventing further avalanche effects. These oversight and intervention mechanisms have been improved since then, but debate continues as to whether they are sufficient to prevent another significant flash crash; mini-crashes and interventions occur daily.[41]

During the Cold War, and even afterwards, both the US and the Soviet Union received erroneous indications of nuclear attack on multiple occasions.[42] These varied from sunlight reflected off clouds to magnetic training tapes fed into the early-warning system by accident. In all these cases, human reasoning led to restraint instead of escalation; double checks revealed that the alarm had been false. At the time, double checking and reconsideration were possible due to flight times of between several hours (in the case of bombers and cruise missiles) and 10–30 minutes (for ballistic missiles launched from submarines or those covering intercontinental ranges), as well as systems for preventing unwanted crisis escalation, such as the 'hotline' for communication between Moscow and Washington established after the Cuban Missile Crisis. Humans, or rapid-reaction mechanisms pre-programmed by humans, can also act as a fail-safe in instances where an overarching authority exists to enforce a shared set of rules, as in the stock-exchange example – unlike in international politics.

With the goal of improved military effectiveness providing a strong incentive to increase operational speeds, and thus to allow AWS to operate without further human intervention, tried and tested mechanisms for double-checking and reconsideration that allow humans to function as fail-safes or circuit-breakers are discarded. This, in combination with

unforeseeable algorithm interactions producing unforeseeable military outcomes, increases crisis instability and is unpleasantly reminiscent of Cold War scenarios of accidental war.

Setting aside the increasing risk of unwanted escalation, AWS are also bound to introduce stronger incentives for premeditated (including surprise) attacks. This is because of a combination of three factors: casualty avoidance, cost reduction and, once again, swarming.

Firstly, unmanned systems, generally speaking, keep soldiers out of harm's way – which is positive, but which also reduces the political risk of military endeavours, especially in democracies.[43] Referring to the current generation of combat drones, Christof Heyns, the United Nations Special Rapporteur on extrajudicial, summary or arbitrary executions, put it this way: '[Drones] make it easier for States to deploy deadly and targeted force on the territories of other States.'[44] As unmanned systems become faster and smaller, as well as, eventually, autonomous – which will also make them stealthier due to radio silence, and allow them to become 'swarmier' – the resulting room for manoeuvre in political and military terms increases.

Unmanned systems keep soldiers out of harm's way

Secondly, the example of *Perdix* demonstrates that AWS need not be big, costly or high-tech. Instead, such systems can be cheap and disposable, produced using 3D printers and gaining strength from numbers, their 'intelligence' residing in a distributed fashion in the swarm or, if external communication is an option, at some higher level within the military 'system of systems' at large.

A closely related third consideration is that swarms would make mounting a successful defence especially difficult due to their resilience and their ability to attack from many directions, simultaneously, in an overwhelming fashion. Small and very small AWS (those measuring tens of centimetres at most) would suffer from limited power supply on board, but could be brought closer to the target by riding along on 'motherships', as has been demonstrated with *Perdix*. With payloads weighing a few hundred grams at most, the amount of destructive power of small drones would be limited

too. But if directed at political or military leaders or sensitive military infrastructure, they would produce relevant damage and provide entirely new means for carrying out assassinations and decapitation strikes.[45]

None of these points in isolation would introduce a radically novel element to military decision-making. After all, the fact that a weapon is cheap does not necessarily render it more likely to be used.[46] However, the combination of these three factors – brought about mainly by the development of hard-to-defend-against autonomous swarms – presents a strong incentive to seize the advantage of being the first on the offensive.

Considering the current climate between Russia and NATO, it stands to reason that old mechanisms of threat perception and worst-case thinking might see a comeback in the wake of AWS deployment.[47] Russia was reportedly alarmed when the idea of using stealthy drones for missile defence was floated in the US.[48] Swarms of AWS could be used to attack nuclear-weapon delivery systems, command and control systems, and sensitive infrastructure components such as antennas, sensors or air intakes. Even though an attacker might have little interest or confidence in the success of a disarming first strike of this type, the fact that such strikes were now possible would in itself increase nervousness and distrust between nuclear-armed adversaries.

This overlap between the conventional and the nuclear realm is not new, of course. It emerged with precision munitions and bunker-busting (or possibly electromagnetic-pulse) warheads during the 1990s and 2000s,[49] and is also documented in the New START treaty, the preamble of which states that the US and Russia are 'mindful of the impact of conventionally armed ICBMs and SLBMs on strategic stability'.[50] But AWS will likely perpetuate and intensify this trend, not least by opening up new possibilities for holding nuclear submarines carrying ballistic missiles at risk.[51] Thus, when nuclear weapons or strategic command and control systems are, or are perceived to be, at greater risk, conventional capabilities end up increasing instability at the strategic level.

Today's unmanned systems have already increased the risk that military force will be used in scenarios where manned systems would previously have presented decision-makers with bigger, caution-inducing hurdles – a connection recently confirmed in war-gaming exercises.[52] Of course,

swarming AWS need not necessarily lead to escalation under all conditions. In asymmetric scenarios involving adversaries who lack AWS capabilities, the escalatory mechanisms developed above would not take effect. In symmetric settings, by contrast, they would certainly exacerbate the overall development toward an increased risk of crisis instability and escalation.

Preventive arms control for AWS?

Due to their detrimental impact on global peace and strategic stability, AWS have sparked a lively international debate among arms-control experts. Since 2013 this debate has included the United Nations. The main venue for UN deliberations on AWS is the Convention on Certain Conventional Weapons (CCW) in Geneva (where AWS used to be referred to as 'lethal autonomous robots' and are now called 'fully' or 'lethal' autonomous weapon systems – the latter designation spawning the clumsy yet widely used acronym LAWS).[53]

Unease with AWS is growing in Geneva. After three 'informal meetings of experts' (leading to continued and intensified deliberations on the issue through a 'Group of Governmental Experts' in 2017), 19 governments have called for a preventive prohibition – commonly referred to as a 'ban' – on AWS, which could be concluded via a sixth CCW protocol.[54] Interestingly, while the notion that autonomous weapon systems may offer certain military benefits is being upheld mainly by Israel and the US, neither they nor any other state party to the CCW has so far argued unambiguously in favour of the development and deployment of AWS.

It is worth clarifying that arms control in the form of a preventive AWS prohibition or ban would not mean prohibiting or controlling specific technologies as such. The wide dissemination and dual-use potential of AI and robotics, the two prime technologies driving AWS, suggest that this would not only be a futile endeavour, it would also be severely misguided in light of the various benefits that could potentially flow from the maturation of these technologies with regard to civilian applications, the self-driving car being just one prominent example.[55] In close connection to this, a number of examples suggest that the private sector would welcome a ban on AWS since companies do not want their products to be associated with 'killer

robots'. In August 2017, 116 AI and robotics company leaders – including Tesla's Elon Musk and DeepMind's Demis Hassabis and Mustafa Suleyman – published an open letter, urging the United Nations to increase its arms-control efforts.[56] Google had stated that it was not interested in military robotics years before, eventually selling the robot-maker Boston Dynamics, which it had owned for a brief period and which was well known at the time for its close ties to the US military.[57] The Canadian robot-maker Clearpath Robotics even officially joined forces with the international Campaign to Stop Killer Robots, asking 'everyone to consider the many ways in which this technology would change the face of war for the worse' and calling for robotic products to be created solely 'for the betterment of humankind'.[58]

In short, preventive arms control for AWS would not mean the regulation or prohibition of specific technologies, nor even countable (stockpiles of) individual weapon systems.[59] Instead, it would mean regulating or prohibiting a defined military practice, particularly certain applications of specific technologies for military purposes. An example of such an approach can be found in the preventive prohibition on blinding laser weapons added to the CCW in 1995.[60] This prohibition protects soldiers' eyes on the battlefield without banning laser technology in all its other military and civilian uses.

When it comes to achieving a preventive prohibition on AWS, all governments are still – more or less – in the same boat. Some states may be in the lead in technological terms, but there is not yet a clear division between haves and have-nots. And while some may hope for AWS to yield specific benefits, no one is oblivious to the risks of developing such systems, at least judging from the current deliberations in Geneva.

Three alternatives to a ban, possibly in some combination, are conceivable. One is an internationally agreed moratorium on the development and deployment of AWS. This might buy time for further research and development, and additional consideration of the risks and benefits of AWS. This seems an unlikely and unwise next step, however, due to the dual-use nature of AWS technology and thus the immense ease with which such an agreement could be spoiled. A second alternative is a non-binding agreement on best practices or a 'code of conduct' with an emphasis on compliance with existing international humanitarian law and more rigorous unilateral

weapons reviews in accordance with Article 36 of Additional Protocol I to the Geneva Conventions.[61] This approach attempts to address legal concerns but does nothing to counter the detrimental effects of AWS on strategic stability. Moreover, it would be incomplete, because only a few states conduct Article 36 weapons reviews – without any obligation to share the results, one might add. The third possibility is an agreement between the major powers and states leading in AWS technology that curbs strategic-escalation risks. This could entail limiting or even precluding AWS interactions, especially swarms, as well as excluding specific targets, such as nuclear weapons, from AWS attacks.[62] The resulting loss of military capabilities would be so great, however, that it seems unlikely that any AWS-capable state would agree to it. Nor does this solution address concerns related to international humanitarian law. It is also worth mentioning that none of these options would address the fundamental ethical problem posed by AWS – that killing people with an anonymous, unaccountable algorithm arguably amounts to a violation of human dignity. Verifying compliance would also be a problem. Non-binding agreements would by definition not be subject to verification, but even an international, legally binding moratorium, which *would* be subject to verification, would present difficulties – just as in the 'zero solution' case of a prohibition. However, a ban might promise enough advantages to prompt the international community to actually muster the resources required for addressing the verification problem.

Verification would be a problem

In light of these considerations, the best solution would be to use the current window of opportunity to firmly establish and codify a bright red line against autonomous weapon systems that take life-and-death decisions out of human hands. This would not be easy, but the long list of concerns raised by AWS from the perspective of international law, ethics and, as argued here, global peace and strategic stability suggests that the international community would be well advised to collectively stop the race toward full autonomy in weapon systems. A legally binding, preventive, multilateral arms-control agreement comprehensively prohibiting the deployment and use of AWS would not only be the most logically consis-

tent, morally desirable and politically prudent solution, it would also deal with all current concerns in one fell swoop.[63]

Could a ban on autonomy in weapon systems ever be verified? After all, autonomy is, as noted, essentially a question of software, potentially consisting of little more than a checkbox to be clicked on a graphical user interface. In other words, AWS and remotely controlled weapon systems may appear identical from the outside. Any existing, hardware-based verification method, let alone a quantitative approach based on counting systems or measuring specific system features, is useless in this case. Yet no state would be willing to allow the software that runs its weapon systems to be inspected. Even if it did, cheating would be all too easy since the software could be changed back within minutes after inspection.

An alternative would be to approach the problem from an *ex post* point of view. A treaty requiring meaningful human control over weapon systems, 'specifically in the "critical functions" of selecting and attacking targets', as the International Committee of the Red Cross suggests,[64] would oblige states parties to install digital 'glass boxes' in all relevant weapon systems to keep secure and reliable records of all sensor and control data exchanged between the system and human operators or supervisors.[65] In cases of suspected illegal use of autonomous systems, any state party to the treaty could be asked to produce (in encrypted form, and handed to an international monitoring organisation) the relevant records to ensure an orderly forensic inquiry into questions of compliance.

Arms control for autonomy in weapon systems, especially regarding verification and compliance, would not be easy to accomplish. Arms control almost never is. However, as weapons technology advances, so too do technologies that can be leveraged for the purposes of arms control. It could be, for example, that novel, inherently manipulation-resistant database solutions such as blockchain could play a role in improving the glass-box approach. Any number of working, creative solutions as yet unknown to the world could potentially be achieved by mustering political will and dedicating more resources to an inquiry into these issues.

Maintaining meaningful human control over the use of weapon systems and life-and-death decision-making in warfare is a worthy and sensible

goal for legal, ethical and strategic reasons. A preventive prohibition of AWS would go a long way toward achieving it, and would come with the additional benefit of curbing the upward spiral towards operational speeds beyond human fail-safe capabilities. Speed is undoubtedly a tactical advantage on the battlefield, and humans are slower than machines. But strategic stability is essential for survival. When it comes under threat, some remainder of human slowness is a good thing.

Notes

1 Future of Life Institute, 'Autonomous Weapons: An Open Letter from AI & Robotics Researchers', http://futureoflife.org/open-letter-autonomous-weapons/.

2 US Department of Defense, 'Autonomy in Weapon Systems', Directive no. 3000.09, 21 November 2012, http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf.

3 See, for example, the Campaign to Stop Killer Robots website, http://stopkillerrobots.org/.

4 See Jürgen Altmann and Mark Gubrud, 'Anticipating Military Nanotechnology', *IEEE Technology and Society Magazine*, vol. 23, no. 4, 2004, pp. 33–40; Jürgen Altmann, *Military Nanotechnology: Potential Applications and Preventive Arms Control* (Abingdon: Routledge, 2006), Section 6.1.2; Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Farnham and Burlington, VT: Ashgate, 2009), Chapter 6; Jürgen Altmann, 'Preventive Arms Control for Uninhabited Military Vehicles', in Rafael Capurro and Michael Nagenborg (eds), *Ethics and Robotics* (Heidelberg: Akademische Verlagsgesellschaft, 2009); Jean-Marc Rickli, 'Strategic Stability, Non-State Actors and Future Prospects', Presentation at CCW Meeting of Experts on LAWS, Geneva, 16 April 2015, http://www.unog.ch/80256EDD006B8954/(httpAssets)/B6E6B974512402BEC1257E2E0036AAF1/$file/2015_LAWS_MX_Rickli_Corr.pdf; I. Lachow, 'The Upside and Downside of Swarming Drones', *Bulletin of the Atomic Scientists*, vol. 73, no. 2, 2017, pp. 96–101; and Wendell Wallach, 'Toward a Ban on Lethal Autonomous Weapons: Surmounting the Obstacles', *Communications of the ACM*, vol. 60, no. 5, 2017, p. 31.

5 Ronald C. Arkin, 'The Case for Ethical Autonomy in Unmanned Systems', *Journal of Military Ethics*, vol. 9, no. 4, 2010, pp. 332–41.

6 Noel Sharkey, 'Grounds for Discrimination: Autonomous Robot Weapons', *RUSI Defence Systems*, vol. 11, no. 2, 2008, pp. 86–9.

7 See Human Rights Watch, 'Mind the Gap: The Lack of Accountability for Killer Robots', 9 April 2015, https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots; and Robert Sparrow, 'Killer

Robots', *Journal of Applied Philosophy*, vol. 24, no. 1, 2007, pp. 62–77.

8 See Peter Asaro, 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making', *International Review of the Red Cross*, vol. 94, no. 886, 2012, pp. 687–709; Christof Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions', A/HRC/23/47, 9 April 2013, http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf; and Robert Hart, 'The Case for Banning Autonomous Weapons Rests on Morality, Not Practicality', *Bulletin of the Atomic Scientists*, 24 April 2017, http://thebulletin.org/case-banning-autonomous-weapons-rests-morality-not-practicality10707.

9 Charli Carpenter, 'Beware the Killer Robots: Inside the Debate over Autonomous Weapons', *Foreign Affairs*, 3 July 2013, http://www.foreignaffairs.com/articles/139554/charli-carpenter/beware-the-killer-robots#cid=soc-twitter-at-snapshot-beware_the_killer_robots.

10 Open RoboEthics Initiative, 'The Ethics and Governance of Lethal Autonomous Weapons Systems: An International Public Opinion Poll', 9 November 2015, http://www.openroboethics.org/wp-content/uploads/2015/11/ORi_LAWS2015.pdf.

11 Paul Scharre, 'Autonomous Weapons and Operational Risk', Center for New American Security, February 2016, available at https://www.cnas.org/publications/reports/autonomous-weapons-and-operational-risk.

12 See Scott D. Sagan, *The Limits of Safety: Organizations, Accidents and Nuclear Weapons* (Princeton, NJ: Princeton University Press, 1993).

13 Russia suspended its CFE participation in 2007 and halted it completely in 2015.

14 US Congress, Office of Technology Assessment, *Ballistic Missile Defense Technologies*, OTA-ISC-254 (Washington DC: US Government Printing Office, September 1985), pp. 119, 128. See also the literature in Appendix L of this report.

15 *Ibid*., pp. 119, 120.

16 'Treaty on Conventional Armed Forces in Europe', 19 November 1990, available at http://www.osce.org/library/14087.

17 US Department of Defense, 'Unmanned Systems Roadmap 2007–2032', 2007, pp. 53, 54, http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA475002.

18 The most recent example is US Department of Defense, 'Unmanned Systems Integrated Roadmap: FY2013–2038', 2013, http:/www.defense.gov/Portals/1/Documents/pubs/DOD-USRM-2013.pdf.

19 Jon Cartwright, 'Rise of the Robots and the Future of War', *Guardian*, 21 November 2010, https://www.theguardian.com/technology/2010/nov/21/military-robots-autonomous-machines.

20 Article 36, 'Killer Robots: UK Government Policy on Fully Autonomous Weapons', 19 April 2013, http://www.article36.org/weapons-review/killer-robots-uk-government-policy-

on-fully-autonomous-weapons-2/.

21 US Department of Defense, 'Department of Defense Announces Successful Micro-Drone Demonstration', *CHIPS: The Department of the Navy's Information Technology Magazine*, 9 January 2017, http://www.doncio.navy.mil/%285udzc155ibdgke454epoce55%29/CHIPS/ArticleDetails.aspx?ID=8575.

22 David Smalley, 'LOCUST: Autonomous, Swarming UAVs Fly into the Future', Office of Naval Research, 14 April 2015, https://www.onr.navy.mil/Media-Center/Press-Releases/2015/LOCUST-low-cost-UAV-swarm-ONR.aspx.

23 US Department of Defense, 'Unmanned Systems Roadmap 2007–2032', p. 49.

24 Aaron Mehta, 'Work: Autonomy in Flight Likely Before Ground', DefenseNews, 30 March 2016, http://www.defensenews.com/story/defense/land/army/2016/03/30/bob-work-autonomy-flight-ground-systems-robot-ai/82427024/.

25 'David Ignatius and Pentagon's Robert Work Talk About New Technologies to Deter War', *Washington Post*, 30 March 2016, https://www.washingtonpost.com/video/postlive/david-ignatius-and-pentagons-robert-work-on-efforts-to-defeat-isis-latest-tools-in-defense/2016/03/30/0fd7679e-f68f-11e5-958d-d038dac6e718_video.html.

26 Philip E. Ross, 'Velodyne Announces a Solid-State Lidar', IEEE Spectrum, 19 April 2017, http://spectrum.ieee.org/cars-that-think/transportation/sensors/velodyne-announces-a-solidstate-lidar.

27 In the case of LIDAR this adaptation might, for example, include ensuring that it can operate under highly adverse conditions, or making efforts to achieve signature reduction.

28 See this data set on the developing role of artificial intelligence in weapon systems: Heather M. Roff and Richard Moyes, 'Project: Artificial Intelligence, Autonomous Weapons, and Meaningful Human Control', Arizona State University, Global Security Initiative, Autonomy, Robotics & Collective Systems, 2016, https://globalsecurity.asu.edu/robotics-autonomy. See also Heather M. Roff, 'Weapons Autonomy is Rocketing', *Foreign Policy*, 28 September 2016, http://foreignpolicy.com/2016/09/28/weapons-autonomy-is-rocketing/.

29 See Chuck Hagel, 'Reagan National Defense Forum Keynote', Simi Valley, CA, 15 November 2014, http://www.defense.gov/News/Speeches/Speech-View/Article/606635; and Dan Lamothe, 'Pentagon Chief Overhauls Silicon Valley Office, Will Open Similar Unit in Boston', *Washington Post*, 11 May 2016, https://www.washingtonpost.com/news/checkpoint/wp/2016/05/11/pentagon-chief-overhauls-silicon-valley-office-will-open-similar-unit-in-boston/.

30 We are thankful to an anonymous reviewer for pointing this out to us.

31 The ten countries possessing these weapons by 2016 were China, Iran, Iraq, Israel, Nigeria, Pakistan, Somalia, South Africa, the UK and the US. New America's 'World of Drones', the source of this list, today names some 28 countries as having armed drones. However, this updated number includes states that only have develop-

ment programmes and demonstrators. See New America, 'World of Drones', http://securitydata.newamerica.net/world-drones.html.

32 Tair Eshel, 'China Tested an Upgraded CH-4 "Rainbow" Weaponized Drone', *Defense Update*, 5 June 2016, http://defense-update.com/20160605_improved_ch-4_rainbow.html.

33 Modes of attack could, for instance, include crashing the system through a specific window, landing on a desk and exploding or stinging like an insect.

34 See Scharre, 'Autonomous Weapons and Operational Risk'.

35 See David Hambling, *Swarm Troopers: How Small Drones Will Conquer the World* (Venice, FL: Archangel Ink, 2015).

36 Jeremy Hsu, 'Any Ban on Killer Robots Faces a Tough Sell', *Discover Magazine*, 29 April 2017, http://blogs.discovermagazine.com/lovesick-cyborg/2017/04/29/any-ban-on-killer-robots-faces-a-tough-sell/#.

37 Paul Scharre, 'Counter-Swarm: A Guide to Defeating Robotic Swarms', *War on the Rocks*, 31 March 2015, http://warontherocks.com/2015/03/counter-swarm-a-guide-to-defeating-robotic-swarms/.

38 For more on the concept of 'normal accidents', see Charles Perrow, *Normal Accidents: Living with High-Risk Technologies* (New York: Basic Books, 1984).

39 Michael Eisen, 'Amazon's $23,698,655.93 Book About Flies', Michaeleisen.org, 22 April 2011, http://www.michaeleisen.org/blog/?p=358.

40 US Commodity Futures Trading Commission and US Securities & Exchange Commission, 'Findings Regarding the Market Events of May 6, 2010', 30 September 2010, https://www.sec.gov/news/studies/2010/marketevents-report.pdf.

41 Gary Shorter and Rena S. Miller, 'High-Frequency Trading: Background, Concerns, and Regulatory Developments', Congressional Research Service, 19 June 2014, http://fas.org/sgp/crs/misc/R43608.pdf.

42 See Sagan, *The Limits of Safety*; Bruce G. Blair, *The Logic of Accidental Nuclear War* (Washington DC: Brookings Institution Press, 1993); Ron Rosenbaum, *How the End Begins: The Road to Nuclear World War III* (London: Simon & Schuster, 2011); and Eric Schlosser, *Command and Control: Nuclear Weapons, the Damascus Accident, and the Illusion of Safety* (London: Penguin, 2013).

43 Frank Sauer and Niklas Schörnig, 'Killer Drones – The Silver Bullet of Democratic Warfare?', *Security Dialogue*, vol. 43, no. 4, 2012, pp. 363–80.

44 Christof Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions', A/68/382, 13 September 2013, http://justsecurity.org/wp-content/uploads/2013/10/UN-Special-Rapporteur-Extrajudicial-Christof-Heyns-Report-Drones.pdf.

45 Two examples of such miniature aircraft, which have already been deployed, are TiGER from MBDA and *Switchblade* from AV Inc., which have wingspans of only tens of centimetres, and can carry a small explosive charge within a radius of three and ten kilometres, respectively. They operate via remote control and a video feed. See David Crane, 'MBDA TiGER (Tactical Grenade Extended Range) Small UAS/UAV (SUAS/

SUAV)/Mini-Flying Bomb/"Kamikaze Drone"', *Defense Review*, 31 October 2011, http://www.defensereview.com/mbda-tiger-tactical-grenade-extended-range-mini-flying-bombkamikaze-drone-for-tactical-reconnaissance-and-precision-kill-missions-is-low-observable-seriously-lethal-kittys-got-a-temper/; and AV Inc., 'Switchblade', 2017, http://www.avinc.com/uas/view/switchblade. Future concepts foresee micro weapons that fly or crawl, and biohybrids using implants in small mammals or insects. Research on biohybrids has been funded for some time by the US Defense Advanced Research Projects Agency. See US Air Force Research Laboratory, 'US Air Force Flapping Wing Micro Air Vehicle', 16 July 2009, https://www.youtube.com/watch?v=_5YkQ9w3PJ4; Sanjiv K.Talwar et al., 'Behavioural Neuroscience: Rat Navigation Guided by Remote Control', *Nature*, vol. 417, no. 6884, 2002, pp. 37–8; and Alper Bozkurt, 'Research Paves Way for Cyborg Moth "Biobots"', *NC State News*, 20 August 2014, http://news.ncsu.edu/2014/08/bozkurt-moth-jove-2014/.

46 We are thankful to an anonymous reviewer for prompting us to further clarify this important point.

47 See Patrick Tucker, 'The Pentagon Is Nervous About Russian and Chinese Killer Robots', Defense One, 14 December 2015, http://www.defenseone.com/threats/2015/12/pentagon-nervous-about-russian-and-chinese-killer-robots/124465/; and Heather Roff, 'The New Mineshaft Gap: Killer Robots and the UN', Duck of Minerva, 20 April 2015, http://duckofminerva.com/2015/04/the-new-mineshaft-gap-killer-robots-and-the-un.html.

48 George N. Lewis and Theodore A. Postol, 'How US Strategic Antimissile Defense Could Be Made to Work', *Bulletin of the Atomic Scientists*, vol. 66, no. 6, 2010, pp. 8–24.

49 Eugene Miasnikov, 'Long-Range Precision-Guided Conventional Weapons: Implications for Strategic Balance, Arms Control and Non-Proliferation', paper commissioned by the International Commission on Nuclear Non-proliferation and Disarmament, June 2009, http://www.armscontrol.ru/pubs/en/emo90918.pdf.

50 'Treaty Between the United States of America and the Russian Federation on Measures for the Further Reduction and Limitation of Strategic Offensive Arms', 8 April 2010, http://www.state.gov/documents/organization/140035.pdf.

51 Sebastian Brixey-Williams, 'Will the Atlantic Become Transparent?', Second Edition, British Pugwash, November 2016, http://www.basicint.org/sites/default/files/Pugwash_TransparentOceans_update_nov2016_v1%281%29.pdf. Regardless of whether the autonomous underwater vehicles are armed themselves, or serve simply to localise and track enemy submarines so that they can be attacked by other weapons, the destabilising effects on deterrence are the same.

52 Center for a New American Security, 'Game of Drones: Wargame Report', 29 June 2016, http://drones.cnas.org/reports/game-of-drones/.

53 See Frank Sauer, 'Stopping "Killer

Robots": Why Now Is the Time to Ban Autonomous Weapons Systems', *Arms Control Today*, vol. 46, no. 8, 2016, pp. 8–13; and Frank Sauer, 'Autonomous Weapons Systems. Humanising or Dehumanising Warfare?', *Global Governance Spotlight*, no. 4, 2014, http://www.sef-bonn.org/fileadmin/Die_SEF/Publikationen/GG-Spotlight/GGS_2014-04_en.pdf.

54 Details of the meetings can be found at United Nations Office at Geneva, '2014 Meeting of Experts on LAWS', http://www.unog.ch/80256EE600585943/%28httpPages%29/A038DEA1DA906F9DC1257DD90042E261?OpenDocument; United Nations Office at Geneva, '2015 Meeting of Experts on LAWS', http://www.unog.ch/80256EE600585943/%28httpPages%29/6CE049BE22EC75A2C1257C8D00513E26?OpenDocument; and United Nations Office at Geneva, '2016 Meeting of Experts on LAWS', http://www.unog.ch/80256EE600585943/%28httpPages%29/37D51189AC4FB6E1C1257F4D004CAFB2?OpenDocument. The 19 governments to have called for a ban include (as of April 2017) Algeria, Argentina, Bolivia, Chile, Costa Rica, Cuba, Ecuador, Egypt, Ghana, Guatemala, Holy See, Mexico, Nicaragua, Pakistan, Panama, Peru, State of Palestine, Venezuela and Zimbabwe. See Campaign to Stop Killer Robots, 'Report on Activities', Convention on Conventional Weapons, Fifth Review Conference, United Nations Geneva, 12–16 December 2016, http://www.stopkillerrobots.org/wp-content/uploads/2013/03/CCW_ReportRC_Feb2017.pdf.

55 Civilian applications of AI/robotics pose their own sets of problems, but these are beyond the scope of this article.

56 Future of Life Institute, 'An Open Letter to the United Nations Convention on Certain Conventional Weapons', https://futureoflife.org/autonomous-weapons-open-letter-2017.

57 Peter Kafka, 'Google Wants Out of the Creepy Military Robot Business', *Recode*, 17 March 2016, https://www.recode.net/2016/3/17/11587060/google-wants-out-of-the-creepy-military-robot-business.

58 Meghan Hennessey, 'Clearpath Robotics Takes Stance Against "Killer Robots"', Clearpath Robotics, 13 August 2014, https://www.clearpathrobotics.com/2014/08/clearpath-takes-stance-against-killer-robots/.

59 See Altmann, *Nanotechnology*, Chapter 5; and Jürgen Altmann, 'Nanotechnology and Preventive Arms Control', Forschung DSF, no. 3, 2005, sections 4.1, 4.2, http://www.bundesstiftung-friedensforschung.de/images/pdf/forschung/berichtaltmann.pdf.

60 United Nations Office at Geneva, 'The Convention on Certain Conventional Weapons', http://www.unog.ch/80256EE600585943/(httpPages)/4F0DEF093B4860B4C1257180004B1B30?OpenDocument.

61 This article states that in 'the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party'. See

'Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflict (Protocol I)', 8 June 1977, https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Article.xsp?action=openDocument&documentId=FEB84E9C01DDC926C12563CD0051DAF7.

62 We owe this suggestion to an anonymous reviewer.

63 Space constraints preclude us from discussing the scope of this prohibition in more depth; suffice it to say that there are valid arguments for extending it to development and testing, as well as to policing and other circumstances as well.

64 International Committee of the Red Cross, 'Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons', 1 September 2016, https://www.icrc.org/en/publication/4283-autonomous-weapons-systems.

65 More precisely, 'A time slice of the data stream immediately prior to and including the selection and engagement commands could be designated as the primary record of [every] engagement. This record would be held by the state party, but a cryptographic code called a "hash" of the record would be recorded by a "glass box" (not "black [box]" because its hardware and software would be known and open) together with a time stamp of the moment the engagement command was issued.' See Mark Gubrud and Jürgen Altmann, 'Compliance Measures for an Autonomous Weapons Convention', International Committee for Robot Arms Control (ICRAC), Working Paper Series #2, 2013, https://icrac.net/wp-content/uploads/2016/03/Gubrud-Altmann_Compliance-Measures-AWC_ICRAC-WP2-2.pdf. This proposal also contains details about what exactly might be prohibited and which exceptions with regard to systems in use for defence against incoming missiles might be accepted. It outlines a wide framework for compliance that begins with the philosophical and legal principle that violent force must always be under human control, and comprises transparency and confidence-building measures, inspections, technical safeguards and forensic investigation of suspicious incidents.