CENTAUR WARFIGHTING:

THE FALSE CHOICE OF HUMANS VS. AUTOMATION

*Paul Scharre**

Much of the debate on autonomous weapons presumes a choice between human versus autonomous decision-making over targeting and engagement decisions. In fact, in many situations, human-machine teaming in engagement decisions will not only be possible but preferable. Hybrid human-machine cognitive architectures will be able to leverage the precision and reliability of automation without sacrificing the robustness and flexibility of human intelligence. While human-machine teaming will not be possible in all circumstances, it does suggest the need to recalibrate some of the debate on autonomous weapons to more accurately distinguish between increased autonomy *in* weapons and "autonomous weapons."

Increasing automation in military systems poses both benefits and risks. Automation is good at many things—precision, reliability, and speed, among them. But autonomous systems are brittle. They lack the flexibility humans have to step outside their instructions and apply "common sense" to adapt to novel situations.

Much of the literature on the legal and ethical implications of autonomous weapons systems (AWS) hinges on these two conflicting characteristics of autonomous systems.[1] Some argue AWS could make warfare more lawful and humane because of their precision and reliability, much as precision-guided weapons have done.[2] Others cite situations where autonomous systems would be

* Paul Scharre is a Senior Fellow and Director of the 20YY Future of Warfare Initiative at the Center for a New American Security. From 2008-2013 he worked in the Office of the Secretary of Defense, where he played a leading role in establishing policies on autonomy in weapons. He is a former infantryman in the 75th Ranger Regiment with multiple tours to Iraq and Afghanistan.

[1] *See The Ethical Autonomy Project: Bibliography*, CTR. FOR A NEW AM. SEC., http://www.cnas.org/research/defense-strategies-and-assessments/20YY-Warfare-Initiative/Ethical-Autonomy/bibliography (last visited Feb. 22, 2016) (providing a comprehensive bibliography on autonomous weapons).

[2] *See, e.g.*, Kenneth Anderson & Matthew Waxman, *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, HOOVER INST. (Apr. 9, 2013), http://www.hoover.org/research/law-and-ethics-autonomous-weapon-systems-why-ban-wont-work-and-how-laws-war-can; Matthew Waxman & Kenneth Anderson, *Don't Ban Armed Robots in the U.S.*, NEW REPUBLIC (Oct. 17, 2013), http://www.newrepublic.com/article/115229/armed-robots-banning-autonomous-weapon-systems-isnt-answer; Kenneth Anderson, Daniel Reisner & Matthew Waxman, *Adapting the Law of Armed Conflict to Autonomous Weapon Systems*, INT'L L. STUD. (2014), https://www.usnwc.edu/getattachment/a2ce46e7-1c81-4956-a2f3-c8190837afa4/dapting-the-Law-of-Armed-Conflict-to-Autonomous-We.aspx; Michael Horowitz & Paul Scharre, *Do Killer Robots Save Lives?*, POLITICO (Nov. 19, 2014), http://www.politico.com/magazine/story/2014/11/killer-robots-save-lives-113010; Ronald C. Arkin, *Warfighting Robots Could Reduce Civilian Casualties, So Calling for a Ban Now Is Premature*,

likely to fail because they could not take into account important contextual factors that may lie outside their programming.[3] Both views have merit. In fact, it is possible to conceive of weapons that exhibit both characteristics that are more precise and discriminating most of the time but still sometimes fail, and potentially fail badly. Are we doomed, then, to choose between the brittleness of automation and human cognitive weaknesses? Artificial intelligences (AI) already perform equally well or better than humans at visual object recognition most of the time, and they are improving.[4] AI is matching or exceeding human performance in a wide range of tasks, from chess to driving to diagnosing medical conditions.[5] Is the price to gaining these advantages accepting the instances when autonomous systems perform in unsatisfactory ways, potentially even catastrophically?

No—humans versus machines is a false choice. The best systems will combine human and machine intelligence to create hybrid cognitive architectures that leverage the advantages of each. Hybrid human-machine cognition can leverage the precision and reliability of automation, without sacrificing the robustness and flexibility of human intelligence.[6]

IEEE SPECTRUM (Aug. 5, 2015, 1:00 GMT), http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/autonomous-robotic-weapons-could-reduce-civilian-casualties.

[3] *See Losing Humanity: The Case Against Killer Robots*, HUM. RTS. WATCH (Nov. 19, 2012), https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots (citing automation bias, a person's decision to trust an automated system over their judgment, as a reason why automated weapons could lead to more civilian casualties); Peter Asaro, *On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision making*, 94 INT'L REV. OF THE RED CROSS 687, 706 (2012) (stating that the particularities of the war environment, including enemy adaptation, degraded communications, environmental hazards, etc., create the possibility that automated systems will act in an unintended fashion); *The Problem*, CAMPAIGN TO STOP KILLER ROBOTS, http://www.stopkillerrobots.org/the-problem/ (last visited Feb. 17, 2016) (arguing that AWS would not be able to understand proportion or the difference between civilians and soldiers and therefore would be in violation of the laws of war); *Autonomous Weapons: An Open Letter from AI & Robotics Researchers*, FUTURE OF LIFE INST. (Jul. 28, 2015), http://futureoflife.org/AI/open_letter_autonomous_weapons (stating that automated weapons, in the wrong hands, could be used for assassinations, destabilizing nations, or selectively killing particular ethnic groups).

[4] *See* Tom Simonite, *A Google Glass App Knows What You're Looking At*, MIT TECH. REV. (Sept. 30, 2013), https://www.technologyreview.com/s/519726/a-google-glass-app-knows-what-youre-looking-at/ (discussing developing computer programing in the field of visual object recognition). The percentage of error by the best AI is limited to 15–17%; therefore, AI perform equally well or better than humans at visual object recognition 83–85% of the time. *Id*.

[5] *See What is Watson?*, IBM, http://www.ibm.com/smarterplanet/us/en/ibmwatson/what-is-watson.html (last visited Feb. 18, 2016) (providing an example of a technology platform that can analyze unstructured data, understand complex questions, and provide answers and solutions).

[6] *See, e.g.*, Bob Work, Deputy Sec'y of Def., U.S. Dep't of Def., Address at the CNAS Defense Forum (Dec. 14, 2015).

A. *"The best chess players in the world are human-machine teams."*[7]

As an example of the future of cognition, look no further than one of the most high-profile areas in which AI has bested humans: chess. In 1997, world chess champion Gary Kasparov lost to International Business Machine's Deep Blue,[8] cementing the reality that humans are no longer the best chess players in the world.[9] But neither, as it turns out, are machines.[10] A year later, Kasparov founded the field of "advanced chess," or centaur chess, in which humans and AI cooperate on the same team.[11] AI can analyze possible moves and identify vulnerabilities or opportunities the human player might have missed, resulting in blunder-free games.[12] The human player can manage strategy, prune AI searches to focus on the most promising areas, and manage differences between multiple AI.[13] The chess AI, or multiple AI, give feedback to the human player, who then decides what move to make.[14] By leveraging the advantages of human *and* machine, centaur chess[15] results in a more perfect game, better than humans or AI alone. Similarly, combining human and machine cognition for targeting and engagement decisions could yield the precision and reliability of automation, without sacrificing the robustness and flexibility that humans bring.

B. *Centaur warfighting: human-machine teaming in engagement decisions*

Human-machine teaming is a better approach than using humans or autonomous systems alone, bringing to bear the unique advantages of each.

[7] Paul Scharre, *Autonomous Weapons and Operational Risk*, CTR. FOR A NEW AM. SEC. 39 (Feb. 2016), http://www.cnas.org/sites/default/files/publications-pdf/CNAS_Autonomous-weapons-operational-risk.pdf; s*ee also* Tyler Cowen, *What Are Humans Still Good For? The Turning Point in Freestyle Chess May Be Approaching*, MARGINALREVOLUTION, http://marginal revolution .com/marginalrevolution/2013/11/what-are-humans-still-good-for-the-turning-point-in-freestyle-chess-may-be-approaching.html (showing that the best chess games of all time have been played by man-machine pairs).

[8] DAVID MOURSUND, BRIEF INTRODUCTION TO EDUCATIONAL IMPLICATIONS OF ARTIFICIAL INTELLIGENCE 22 (2006).

[9] *See id*. ("Although Kasparov was considered to be one of the strongest chess players of all time and the match was close, the computer won . . . .").

[10] *See* Mike Cassidy, *Centaur Chess Brings out the Best in Humans and Machines*, THE BLOOMREACH BLOG (Dec. 14, 2014), http://bloomreach.com/2014/12/centaur-chess-brings-best-humans-machines/ ("Teaming the two in chess, experts say, produces a force that plays better chess than either humans or computers can manage on their own.").

[11] *Id*.

[12] *See id*. ("What [computers] can do is push players towards perfection by charting out a series of flawless opening moves and then providing recommendations the rest of the way.").

[13] *See* Cowen, *supra* note 7 (listing four ways in which a human-computer chess player can add value to a sole computer chess player).

[14] *Id*. ("The human can see where *different* chess-playing programs *disagree*, and then ask the programs to look more closely at those variations, to get a leg up against the computer playing alone.").

[15] Cassidy, *supra* note 10.

Understanding how human-machine teaming might work for weapons engagements requires first disaggregating the different roles a human operator performs today with respect to selecting and engaging enemy targets.

With today's semi-autonomous weapon systems, humans currently perform three kinds of roles with respect to target selection and engagement.[16] In some cases, human operators perform multiple roles simultaneously:[17]

1. **The human as essential operator**: The weapon system cannot accurately and effectively complete engagements without the human operator.[18]
2. **The human as moral agent**: The human operator makes value-based judgments about whether the use of force is appropriate. For example, the human operator decides whether the military necessity of destroying a particular target in a particular situation outweighs the potential collateral damage.[19]
3. **The human as fail-safe**: The human operator has the ability to intervene and alter or halt the weapon system's operation if the weapon begins to fail or if circumstances change such that the engagement is no longer appropriate.[20]

An anecdote from the U.S. air campaign over Kosovo in 1999 includes an instructive example of all three roles in action simultaneously:

> On 17 April 1999, two F-15E Strike Eagles, Callsign CUDA 91 and 92, were tasked to attack an AN/TPS-63 mobile early warning radar located in Serbia. The aircraft carried AGM-130, a standoff weapon that is actually remotely flown by the weapons system officer (WSO) in the F-15E, who uses the infra-red sensor in the nose of the weapon to detect the target. CUDA 91, flown by two captains (Phoenix and Spidey) from the 494th Fighter Squadron, launched on coordinates provided by the Air Operations Center. As the weapon approached the suspected target location, the crew had not yet acquired the [enemy radar]. At 12 seconds from impact, the picture became clearer. . . . [The pilots saw the profile outline of what appeared to be a church steeple.] Three seconds [from impact], the WSO makes the call: "I'm ditching in this field" and steers the weapon into an empty field several hundred meters away. . . . Postflight review of the tape revealed no object that could be positively identified as a radar, but the profile of a Serbian Orthodox church was unmistakable.[21]

[16] *See* Paul Scharre & Michael C. Horowitz, *An Introduction to Autonomy in Weapon Systems*, 8 (CTR. FOR A NEW AM. SEC., Working Paper No. 021015, 2015), http://www.cnas.org/sites/default/files/publications-pdf/Ethical%20Autonomy%20Working%20Paper_021015_v02.pdf (explaining the types of AWS and humans' role in operating them).

[17] *Id*. at 16.

[18] *See id*. (distinguishing between human-supervised and semi-autonomous weapon systems).

[19] *Id*.

[20] *Id*.

[21] Mike Pietrucha, *Why the Next Fighter will be Manned, and the One After That*, WAR ON

In this example, the pilots were performing all three roles simultaneously.[22] In overriding the automated system and manually guiding the air-to-ground weapon they were acting as essential operators.[23] Therefore, without the guidance of the pilots, the weapon would not have been accurate or effective. They were also acting as moral agents.[24] They assessed the military necessity of the target as not worth the potential collateral damage to what appeared to be a church.[25] And they were acting as fail-safes, observing the weapon while it was in flight and making an on the spot decision to abort once they realized the circumstances were different from what they had anticipated.[26]

In a different scenario, however, human operators might only perform some of these roles. A GPS-guided bomb, for example, would not need manual guidance while in flight.[27] If such a bomb was network-enabled, giving operators the ability to abort in-flight, and the pilots had the ability to observe the target area immediately prior to impact, they still could perform the roles as moral agents and fail-safes, even if they were no longer essential operators once they launched the weapon.[28]

Other types of automation in non-military settings disaggregate these functions in various ways. A person kept on medical life support has machines performing the essential task of keeping him or her alive, but it is humans making the moral judgment whether to continue life support.[29] Commercial airliners today

THE ROCKS (Aug. 5, 2015), http://warontherocks.com/2015/08/why-the-next-fighter-will-be-manned-and-the-one-after-that/.

[22] *See id*. (demonstrating that during an incident where the desired target was not present, pilots played the role of essential operator, moral agent, and fail-safe to make a decision to avoid hitting a building unassociated with the assigned target).

[23] *See id*. ("The aircrew not only determined that they could not find the assigned target, but identified what they could find and assessed the consequences.").

[24] *Id*.

[25] *See id*. ("The implications of hitting a Serbian church with a 2000-lb. general-purpose warhead were worked through in real time.").

[26] *See* Pietrucha, *supra* note 21 ("This example illustrates another reality of weapons employment — sometimes the planned target isn't there. Hitting the wrong target can have significant effects on the conduct and outcome of a conflict . . . [and must be] worked through in real time.").

[27] *See* Carlo Kopp, *GPS Aided Guided Munitions*, AIR POWER AUSTL. (Jan. 27, 2014), http://www.ausairpower.net/TE-GPS-Guided-Weps.html (discussing the mechanics of a GPS guided missile).

[28] This assumes that they have sufficient time to perform these roles as moral agent and fail-safe, which will depend on the specific situation. *Id*. at 42. The same pressures that drove the desire to automate the essential operation of the weapon system could complicate the human's ability to act as moral agent or fail-safe. *See id*. at 9–10 ("Some guided munitions have the ability to be controlled, aborted or retargeted in flight . . . [but] [s]ome guided munitions cannot be retargeted in flight and are considered 'fire and forget' weapons; once launched, they cannot be recalled.").

[29] *See* Clyde Haberman, *From Private Ordeal to National Fight: The Case of Teri Schiavo*, N.Y. TIMES (Apr. 20, 2014), http://www.nytimes.com/2014/04/21/us/from-private-ordeal-to-national-fight-the-case-of-terri-schiavo.html?_r=0 (highlighting the human moral

have automation to perform the essential task of flying the aircraft, with human pilots largely in a fail-safe role, able to intervene in the event the automation fails.[30] As automation becomes more advanced across a range of applications, it will become technically possible to remove the human from the role of essential operator in many circumstances.[31] In fact, automating the weapon system's operation may result in far greater accuracy, precision, and reliability than relying on a human operator.[32] Just as autonomous systems can land airplanes, manage subway repair schedules, play chess, answer trivia questions, and arrive at complex medical diagnoses more accurately than humans, they may also be capable of performing many tasks in war better than humans.[33] Automating the human's role as moral agent or fail-safe, however, may be far harder. Humans have moral and legal judgment, responsibility, and accountability, making their role as moral agents important for many tasks in war.[34] Humans also have great value as fail-safes, with the ability to flexibly respond to a range of unplanned scenarios.[35]

C. The role of the human as moral agent and fail-safe

It is possible to design systems that incorporate both automation *and* human decision-making, using automation to perform essential tasks with greater precision and accuracy, while retaining humans in the role of moral agents and

struggle around end of life decisions).

[30] *See* Mary Cummings & Alexander Stimpson, Opinion, *Full Auto Pilot: Is it Really Necessary to Have a Human in the Cockpit?*, JAPANTODAY (May 20, 2015, 6:53 AM), http://www.japantoday.com/category/opinions/view/full-auto-pilot-is-it-really-necessary-to-have-a-human-in-the-cockpit (describing the increasingly limited role human aircraft pilots play in an age of increased automation).

[31] *See id*. ("With the development of new automated technologies, the workload in the cockpit has been dramatically decreasing—so much so that pilots self-report only touching the controls for about three to seven minutes during a typical flight.").

[32] *See id*. ("Human pilot error is responsible for 80 percent of all accidents in military and commercial flights.").

[33] *See* John Markoff, *Planes Without Pilots*, N.Y. TIMES (Apr. 6, 2015), http://www.nytimes.com/2015/04/07/science/planes-without-pilots.html?_r=0 ("Advances in sensor technology, computing and artificial intelligence are making human pilots less necessary than ever in the cockpit."); Hal Hodson, *The AI Boss that Deploys Hong Kong's Subway Engineers*, NEW SCIENTIST, http://www.newscientist.com/article/mg22329764.000-the-ai-boss-that-deploys-hong-kongs-subway-engineers.html#.VRB7jELVt0c (last updated July 7, 2014) (describing the benefits of artificial intelligence and automation in the realm of managing engineering functions for Hong Kong's mass transit system); IBM, *What is Watson?*, *supra* note 5 (describing Watson's ability to process complex information in order to provide answers to questions); *IBM Watson Health*, IBM, http://www.ibm.com/smarterplanet/us/en/ibmwatson /health/ (last visited Feb. 19, 2016) (describing Watson's capabilities in the healthcare marketplace).

[34] Christof Heyns, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns*, U.N. Doc. A/HRC/23/47 (Apr. 9, 2013); HRC, 23rd Sess., www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf.

[35] *See* Pietrucha, *supra* note 21 ("Meeting the challenges of tactical execution and weapons employment, and maintaining the ability to learn and improve the fighter aviation enterprise are essential ingredients and remain entirely human endeavors.").

fail-safes. The U.S. counter-rocket, artillery, and mortar system (C-RAM) is an example of this approach, automating much of the engagement, resulting in more precise and accurate engagements, while keeping a human in the loop as a fail-safe.[36]

The C-RAM is designed to protect U.S. bases where it is installed from rocket, artillery, and mortar attacks, using a network of radars to automatically identify and track incoming rounds.[37] Because the C-RAM is frequently used at U.S. military bases where there are friendly aircraft in the sky, the system autonomously creates a "Do Not Engage Sector" around friendly aircraft to prevent fratricide.[38] The result is a highly automated system that, in theory, would be capable of safely and lawfully completing engagements entirely on its own.[39] However, humans are still kept "in the loop" for final verification of each individual target before engagement.[40] One C-RAM operator described the role the automation and human operators play:

> The human operators do not aim or execute any sort of direct control over the firing of the C-RAM system. The role of the human operators is to act as a final fail-safe in the process by verifying that the target is in fact a rocket or mortar, and that there are no friendly aircraft in the engagement zone. A [h]uman operator just presses the button that gives the authorization to the weapon to track, target, and destroy the incoming projectile.[41]

Thus, the C-RAM employs overlapping safeties, both automated and human.[42] The autonomous safety tracks friendly aircraft in the sky with greater precision and reliability than human operators could.[43] But a human is still retained in the loop to react to unforeseen circumstances.[44] This model also has the virtue of ensuring that

[36] *Counter Rocket, Artillery, and Mortar (C-RAM)*, GLOBALSECURITY.ORG, http://www.globalsecurity.org/military/systems/ground/cram.htm (last modified Apr. 3, 2015, 7:31 PM).

[37] *Id*.

[38] *See* Sam Wallace, *The Proposed Ban on Offensive Autonomous Weapons is Unrealistic and Dangerous*, KURZWEILAI (Aug. 5, 2015), http://www.kurzweilai.net/the-proposed-ban-on-offensive-autonomous-weapons-is-unrealistic-and-dangerous (explaining that humans make the final decision to fire in an engagement zone).

[39] *Id*.

[40] *Id*.

[41] *Id*.

[42] *See id*. (indicating that within the C-RAM, the machine does the majority of the work, but there is still limited human involvement).

[43] *See id*. ("Machines excel at making split-second tactical decisions that humans have trouble with.").

[44] *See* Wallace, *supra* note 38 (indicating that there is still limited human involvement within the C-RAM system). The C-RAM was designed post-2003, after the Patriot fratricides. *See* Josh Davidson, *Countering Rockets, Artillery, and Mortars*, 12 MIL. INFO. TECH. (Nov. 14, 2008), http://www.kmimediagroup.com/military-information-technology/magazines/9-mit-2008-volume-12-issue-9/44-countering-rockets-artillery-and-mortars (describing the advent of C-RAM and its ability to avoid fratricides). One can understand why this dual-safety approach was desirable, given the Patriot's record in Operation Iraqi Freedom. *See* JOHN K. HAWLEY, ARMY RESEARCH LABORATORY, ARL-SR-0158, LOOKING BACK AT 20 YEARS OF MANPRINT ON

human operators must take a positive action before each engagement,[45] helping to ensure clear human responsibility for engagements.[46]

In principle, an approach along the lines of C-RAM's blended use of automation and human decision-making is optimal, leveraging the advantages of each. This allows militaries to add automation to increase precision and accuracy without giving up the role of the human as moral agent and fail-safe.[47] The human may not be able to necessarily prevent all accidents from occurring (after all, humans make mistakes), but the inclusion of a human in the loop allows the human to adapt to unanticipated situations and undertake corrective action, if required, between engagements.[48] This dramatically reduces the potential for multiple erroneous engagements.[49]

In order for the human operators to actually perform the roles of moral agent and fail-safe, the operators must be trained for and supported by a culture of active participation in the weapon system's operation.[50] Humans were "in the loop" in two fratricide incidents in 2003 when the U.S. Patriot air defense system shot down two friendly aircraft.[51] Afterward, a lengthy internal Army investigation criticized the Patriot community culture for "trusting the system without question."[52] According to John K. Hawley, an engineering psychologist at the Army Research Laboratory Human Research and Engineering Directorate, Patriot operators, while nominally in control, exhibited automation bias—an "unwarranted and uncritical trust in automation. In essence, control responsibility is ceded to the machine."[53] The type of "unwarranted trust" in automation that led to the Patriot

PATRIOT: OBSERVATIONS AND LESSONS 1 (2007) ("During the combat operations phase of Operation Iraqi Freedom (OIF), Army Patriot units were involved in two fratricide incidents.").

[45] *See* Wallace, *supra* note 38 ("A [h]uman operator . . . presses the button that gives the authorization to the weapon to track, target, and destroy the incoming projectile.").

[46] *See* Michael N. Schmitt, *Autonomous Weapons Systems and International Humanitarian Law: A Reply to the Critics*, 4 HARV. NAT'L SEC. J. 201, 277 (2013) (describing situations in which the engagement of an autonomous weapon may result in human accountability).

[47] *See* Wallace, *supra* note 38 (showing that an essential role of the human operators is to act as a final fail-safe in the process).

[48] *See id*. ("[H]uman operators [] act as a final fail-safe in the process by verifying that the target is in fact a rocket or mortar, and that there are no friendly aircraft in the engagement zone.").

[49] *See id*. ("This system can be automated, but . . . human authority is still required to authorize the weapon system to fire"). If you have a human that can provide a kill switch at crucial moments, the machine will not make a mistake that then is a result of human responses to the incident, in addition to mistaken attacks by the autonomous weapon. *Id*.

[50] *See* Schmitt, *supra* note 46 (indicating that the U.S. Department of Defense requires certain high-level military officials to review training and doctrine for AWSs to ensure operators understand the functioning, capabilities, and limitations of a system's autonomy in realistic operational conditions).

[51] *See* HAWLEY, supra note 44 ("During the combat operations phase of Operation Iraqi Freedom (OIF), Patriot air and missile defense units were involved in two fratricide incidents.").

[52] *Id*. at 4.

[53] John K. Hawley, *Not by Widgets Alone: The Human Challenge of Technology-Intensive Military Systems*, ARMED FORCES J. (Feb. 1, 2011), http://www.armedforcesjournal.com/not-by-

fratricides would result in a human in the loop in name only.[54] Training that requires human operators to exercise judgment and a culture that emphasizes human responsibility are essential to ensuring that the human's role remains meaningful.

D. The limits of centaur warfighting: speed and communications

There are some situations in which this hybrid human-machine cognitive model begins to break down, including: when high-speed reactions are required faster than human reaction times; and when communications are denied between the human and machine.[55]

Once again, chess is a useful analogy. While centaur human-machine teams generally result in better decision-making in chess, it is not an optimal model in timed games where the player has a limited amount of time to make a move.[56] When the time to decide is compressed, the human does not add any value compared to the computer alone, and may even be harmful by introducing errors.[57] This is clearly the case today for high-speed chess games where a player has only thirty to sixty seconds to make a move.[58] Over time, as computers advance, one would anticipate this time horizon to expand until humans no longer add any value regardless of how much time is allowed.[59] This situation is analogous to the role human-supervised autonomous weapons play today in defending against saturation attacks from missiles and rockets, where the speed of engagements could easily overwhelm human operators' ability to respond quickly enough. While humans remain in the loop for C-RAM, at least thirty countries, including the United States, employ automated defensive systems similar to C-RAM human-supervised "on the loop" modes of operation.[60] Once these modes are activated, human

widgets-alone/. Patriot operators now train on this and other similar scenarios to avoid this problem of unwarranted trust in the automation. *Id*.

[54] *Id*.

[55] *See* Wallace, *supra* note 38 (describing scenarios where there is a high-speed attack and where there is a loss of communications).

[56] *See* Max Nisen, *Humans Are on the Verge of Losing One of Their Last Big Advantages over Computers*, BUS. INSIDER (Nov. 5, 2013, 12:36 PM), http://www.businessinsider.com/computers-beating-humans-at-advanced-chess-2013-11 ("You have to seriously slow down the game for a centaur to compare programs at a deep enough level that they can add anything.").

[57] *See* Cowen, *supra* note 7 ("[A]t sufficiently fast time controls the human attempts to improve on the computer may simply amount to noise or may even be harmful, given the possibility of human error.").

[58] *See id*. ("[A]t, say, thirty or sixty seconds a game the human hasn't been able to add value to the computer for some time now.").

[59] *See id*. (noting that as computer programs improve, centaur advantages become harder to exploit).

[60] *See* Scharre & Horowitz, *An Introduction to Autonomy in Weapons Systems*, *supra* note 16, at 12 ("At least 30 nations use human-supervised defensive systems with greater autonomy, where humans are 'on the loop' for selecting and engaging specific targets.").

[One] way in which the word autonomy is used refers to the relationship between the person and the machine. Machines that perform a function for some period of time, then stop and wait for operators can observe the weapon system's operation and can intervene, if necessary, but the weapon will not wait for human authorization before firing.[61] Over time as adversaries employ more advanced missiles, including potentially with cooperative swarming behavior, these defensive human-supervised autonomous weapons are likely to become even more important.

Human-supervised autonomous weapons entail a higher degree of risk than semi-autonomous systems, such as C-RAM. While human operators nominally have the ability to intervene and reassert control over the system, in practice, this may be difficult.[62] The May 2010 stock market "flash crash," where the Dow Jones Industrial Average lost nearly ten percent of its value in a matter of minutes, illustrates some of the challenges in maintaining effective control over high-speed human-supervised autonomous systems.[63] While humans maintain supervisory control over the stock market in principle, the speed of interactions means that the potential damage high frequency trading algorithms can cause before humans take corrective action may be quite high. Human control is more akin to that of an inattentive human driver on an autonomous car speeding down the highway. Steering wheel or no, the driver is a *de facto* passenger along for the ride. From a risk perspective, the damage potential of a human-supervised autonomous weapon depends on how quickly human operators can identify that the system is failing and take corrective action. In some situations, multiple unintended engagements could occur.[64]

Shifting to a human-supervised control model also complicates the role of the human as moral agent. Rather than requiring the human to take a positive action to cause an engagement, in a human-supervised control model the engagement will occur if the human does nothing.[65] The human has to take a positive action to halt

human input before continuing, are often referred to as "semiautonomous" or as having a "human in the loop." Machines that can perform a function entirely on their own but have a human in a monitoring role, with the ability to intervene if the machine fails or malfunctions, are often referred to as "human-supervised autonomous" or "human on the loop." Machines that can perform a function entirely on their own with humans unable to intervene are often referred to as "fully autonomous" or "human out of the loop." In this sense, "autonomy" is not about the intelligence of the machine, but rather its relationship to a human controller.

Id. at 6.

[61] *Id*.

[62] *See id*. at 12–13 (explaining that human controllers can supervise the operation of a system in real-time and can intervene, but may not be able to intervene due to the short amount of time required for engagements).

[63] U.S. COMMODITY FUTURES TRADING COMM'N AND U.S. SEC. AND EXCH. COMM'N, FINDINGS REGARDING THE MARKET EVENTS OF MAY 6, 2010, 1–2 (Sept. 30, 2010), http://www.sec.gov/news/studies/2010/marketevents-report.pdf.

[64] *See Review of the 2012 US Policy on Autonomy in Weapons Systems*, HUM. RTS. WATCH (Apr. 15, 2013), https://www.hrw.org/news/2013/04/15/review-2012-us-policy-autonomy-weapons-systems ("An unintended engagement is defined as '[t]he use of force resulting in damage to persons or objects that human operators did not intend to be the targets of U.S. military operations.'").

[65] Michael C. Horowitz & Paul Scharre, *Meaningful Human Control in Weapon Systems: A Primer* 12–13 (Ctr. for New Am. Sec., Working Paper No. 031315, 2015), http://www.cnas.org/sites/default/files/publications-pdf/Ethical_Autonomy_Working_Paper_031315.pdf.

the weapon system from operating.[66] This significantly complicates the automation bias problem that arose in the Patriot fratricides, and its significance from a human psychology perspective should not be overlooked.[67] As with semi-autonomous weapons, rigorous training where human operators are forced to exercise judgment and intervene to halt the weapon system's operation is critical, buttressed by a culture that emphasizes human responsibility.[68]

The human-machine teaming model changes even more significantly under conditions of degraded or denied communications. Human supervision is only possible if the weapon system has reliable, real-time or near-real-time communications with human operators.[69] This may not always be the case, however.[70] Communications are challenging in some environments, such as underwater, and adversaries will seek to jam or disrupt communications in contested areas.[71]

[66] *Id*. at 13.

[67] *See Report of the Defense Science Board Task Force on Patriot System Performance*, U.S. DEP'T OF DEF. 2 (Jan. 2005), http://www.acq.osd.mil/dsb/reports/ADA435837.pdf (stating the human trust of the Patriot missile is only appropriate for heavy missile attacks, but not when friendlies significantly outnumber enemies, and the solution is for more operator involvement).

[68] *See Q&A on Fully Autonomous Weapons*, HUM. RTS. WATCH (Oct. 21, 2013), https://www.hrw.org/news/2013/10/21/qa-fully-autonomous-weapons#1 (arguing human accountability is important because it can deter people from committing war crimes or being negligent and accountability dignifies victims by recognizing they were wronged and can see someone punished).

[69] *See Unmanned Systems Integrated Roadmap FY2011-2036*, U.S. DEP'T OF DEF. vi (2011), http://www.acq.osd.mil/sts/docs/Unmanned%20Systems%20Integrated%20Roadmap%20FY2011-2036.pdf (stating today's unmanned systems require a high degree of human interaction and a reliance on full-time high-speed communication links).

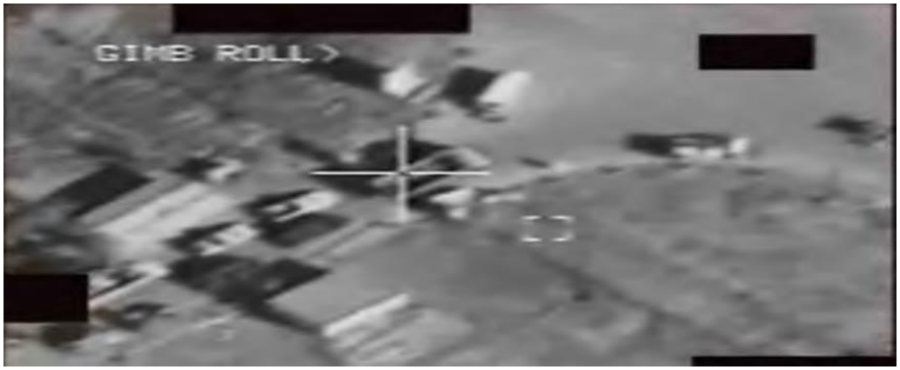
[70] *See id*. at 5 (explaining the goal of creating unmanned systems to enhance effectiveness, efficiency, speed, and close war-fighting gaps).

[71] One conceivable argument for fully autonomous weapons might be to deliberately sever the communications link with human controllers to minimize vulnerability to hacking. *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects*, INT'L COMM. OF THE RED CROSS, 69–70 (Nov. 2014), https://www.icrc.org/en/download/file/1707/4221-002-autonomous-weapons-systems-full-report.pdf. It is worth pointing out that this would eliminate one potential vector for hackers to gain access to the system, but would not render the system hacker-proof. *Id*. Any system with a computer is susceptible to malware, which could still be introduced through other means, such as when the system is connected during maintenance. *Id*. USB flash drives, for example, are notorious for spreading malware across military computer systems. *See* Elinor Mills, *Bad flash drive caused worst U.S. military breach*, CNET (Aug. 25, 2010, 3:37 PM), http://www.cnet.com/news/bad-flash-drive-caused-worst-u-s-military-breach/ (describing a 2008 incident where a flash drive with malware led to a severe military computer breach). While severing communications would take away one potential vector for attacks, doing so would come at a high cost: forgoing any ability to retake control of the system or re-task it once it is launched, even if it began performing inappropriately. Paul Scharre, *Autonomous Weapons and Operational Risk*, *supra* note 7, at 46–47. Computer security, protected communications, and a process for authenticating valid authorization for commands are essential to all networked, computerized military systems, autonomous or not. *Id*. However, opting for fully autonomous weapons because of fears about hacking would be a strange choice. *Id*. at 9. While the probability of an adversary gaining access might be somewhat reduced, the

Communications in contested areas is not an all-or-nothing proposition, however. Capable militaries will be able to employ jam-resistant communications, although they will be limited in bandwidth and range.[72] This could allow a human in a nearby vehicle to remotely remain in the loop to authorize engagements.[73] The type of high-bandwidth, high-definition full motion video that is used from uninhabited aircraft today would not be possible in contested environments, but some communications are likely possible.[74] This raises a critical question: how much bandwidth is required to keep a human in the loop?

Not much. As one example, the below screen grab from a video of an F-15 strike in Iraq is a mere 12 kilobytes in size.[75] While grainy, it clearly shows sufficient resolution to make out individual vehicles, and would allow a trained operator to discriminate military-specific vehicles, such as a tank or mobile missile launcher, from dual-use vehicles such as buses or trucks.[76]

Targeting Image from F-15 strike in Iraq (12KB)



To give a sense of scale, connections on par with a 56K modem from the 1990s could transmit two such frames per second and still have some bandwidth left over for vehicle command-and-control.[77] This would be far from full-motion video, but would allow half-second updates for human operators to respond to

potential consequences if an adversary were to gain access could be far more severe. *Id*. The net balance of risk is likely to favor opting for increased opportunity for human control, where possible. *Id*. at 8–10.

[72] Paul Scharre, *Yes, Unmanned Combat Aircraft are the Future*, WAR ON THE ROCKS (Aug. 11, 2015), http://warontherocks.com/2015/08/yes-unmanned-combat-aircraft-are-the-future/; *see also* Albert Muller, *The Future of Naval Communications,* NAVAL-TECHNOLOGY.COM (Jun. 16, 2010) (discussing the benefits of surface communication links, which offer higher bandwidth, work in multinational environments and are secure and jam resistant).

[73] Int'l Comm. of the Red Cross, *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian* Aspects, *supra* note 73, at 70.

[74] Paul Scharre, *Yes, Unmanned Combat Aircraft are the Future*, *supra* note 72.

[75] *Id*.

[76] *Id*.

[77] *Id*.

changing events on the ground.[78] While this resolution is not sufficient for finding and tracking people in counterinsurgency conflicts, militaries are not likely to face communications challenges in those environments.[79] In high-end anti-access/area denial (A2/AD) environments where communications will be contested, militaries would primarily be targeting an enemy's military equipment (radars, tanks, missile launchers, trucks, airplanes, airfields, ships, etc.), where this lower level of resolution would likely be sufficient for targeting.[80]

Including other sensor data, command-and-control data, and overhead for encryption, total bandwidth requirements might be on the order of hundreds of kilobytes per second to keep a human "in the loop" for targeting decisions, but certainly megabytes or gigabytes per second.[81] The current "remotely piloted" model where high-definition full motion video is streamed from uninhabited systems forward in the battlespace back to remote human controllers is both impossible in contested environments and unnecessary.[82] Autonomy can be used to control uninhabited systems, search for targets, track potential targets, and cue images of the target and surrounding area to human operators who determine final authorization.[83] Human operators could be forward in the battlespace in human-inhabited vehicles exercising command-and-control over uninhabited systems with shorter-range (and higher bandwidth) line of sight communications.[84] Uninhabited vehicles could operate at the vanguard of the formation, with human controllers quarterbacking the fight from nearby vehicles further removed from enemy threats.[85] In this paradigm of human-machine combat teaming, militaries could exploit many of the advantages of uninhabited systems in contested environments, including their ability to take greater risk, while still keeping a human in the loop for engagement decisions.[86]

Certainly, even this reduced-bandwidth approach would not work in areas

[78] *Id*.

[79] Scharre, *Yes, Unmanned Combat Aircraft are the Future*, *supra* note 72.

[80] *Id*.

[81] *See generally id*.

[82] *See* CHRISTOPHER H. STERLING, MILITARY COMMUNICATIONS: FROM ANCIENT TIMES TO THE 21ST CENTURY 241 (2008) (noting the 2003 U.S.–led coalition forces used a bandwidth of 783 megabytes per second); *see also How much bandwidth does Skype need?*, SKYPE (2016), https://support.skype.com/en/faq/FA1417/how-much-bandwidth-does-skype-need (stating the recommended bandwidth for high quality video calling is 500 kilobytes per second).

[83] *See* Robert P. Haffa, Jr. & Anand Datla, *Joint Intelligence, Surveillance, and Reconnaissance in Contested Airspace*, 28 AIR & SPACE POWER J. 29, 38 (May 2014) (explaining the obstacles that intelligence, surveillance, and reconnaissance networks face in contested environments).

[84] Scharre & Horowitz, *An Introduction to Autonomy in Weapons Systems*, *supra* note 16, at 10.

[85] *Id*. at 11.

[86] *See* IBM, *What is Watson?*, *supra* note 5 ("Hybrid human-machine cognition can leverage the precision and reliability of automation, without sacrificing the robustness and flexibility of human intelligence.").

where communications were denied entirely.[87] In such environments, semi-autonomous weapons could only engage specific targets that had been pre-authorized by human controllers, much as cruise missiles do today.[88] Militaries might desire fully autonomous weapons, however, either offensively to seek out and destroy emerging targets of opportunity that had not been preauthorized or defensively to give uninhabited systems the ability to defend themselves in the event that they come under attack.[89]

The military value of fully autonomous weapons for operations in communications-denied environments cannot be easily dismissed, particularly the role of limited, proportional defensive measures to protect costly uninhabited systems. However, these situations are far more limited than much of the literature on autonomous weapons would suggest.[90] Advanced militaries will have the ability to retain some communications, even if it is limited in bandwidth and range, even in contested areas.[91] When possible, a centaur model that retains human decision-making, both as a fail-safe and as a moral agent, will nearly always be preferable.[92] In fact, the trend in next-generation weapons is toward both greater autonomy *and* greater connectivity to human controllers.[93] "Net enabled" weapons allow dynamic re-tasking in flight by human controllers, resulting in more militarily-effective weapons.[94]

[87] *See* M.L. Cummings, *The Human Role in Autonomous Weapon Design and Deployment*, CTR. FOR ETHICS & THE RULE OF L., 2–3 (Nov. 2014), https://www.law.upenn.edu/live/files/3884-cummings-the-human-role-in-autonomous-weapons (explaining how autonomous vehicles can still function without human intervention in the case of communication failure).

[88] *See id*. at 5–6 (explaining the system behind an autonomous weapon and how it would behave without human control).

[89] *See id*. at 2 (noting a Department of Defense directive stating autonomous weapons can be used effectively in certain military situations, both offensively and defensively).

[90] *See* Scharre & Horowitz, *An Introduction to Autonomy in Weapons Systems*, *supra* note 16, at 18 (discussing the many situations autonomous weapons are already used where communication is not an issue).

[91] Int'l Comm. of the Red Cross, *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects*, *supra* note 73, at 70.

[92] *See id*. (arguing decisions in war do not have objectively right answers, and humans are necessary for moral judgments). *But see* Scharre & Horowitz, *An Introduction to Autonomy in Weapons Systems*, *supra* note 16, at 13–15 (explaining how militaries already allow autonomous decisions based on types of targets they should engage).

[93] *See* Peter W. Singer, *In the Loop? Armed Robots and the Future of War*, BROOKINGS (Jan. 28, 2009), http://www.brookings.edu/research/articles/2009/01/28-robots-singer (discussing the trend towards growing autonomy in weapons).

[94] *See* Bob Work, *Deputy Secretary of Defense Speech*, U.S. DEP'T OF DEF. (Dec. 14, 2015), http://www.defense.gov/News/Speeches/Speech-View/Article/634214/cnas-defense-forum (commenting on the military effectiveness of autonomous weapons); *see also Tomahawk*, NAVAL AIR SYS. COMMAND, http://www.navair.navy.mil/index.cfm?fuseaction=home.display&key=F4E98B0F-33F5-413B-9FAE-8B8F7C5F0766 ("The Tomahawk Block IV . . . [has] the capability to reprogram the missile while in-flight via two-way satellite communications to strike any of 15 pre-programmed alternate targets or redirect the missile to any Global Positioning System (GPS) target coordinates.").

E. Recalibrating the debate on autonomous weapons

This suggests the need to recalibrate some of the debate over autonomous weapons. In many situations, militaries will be able to use the advantages of automation without sacrificing the role of the human operator as a responsible moral agent or a fail-safe.[95] Arguments for using automation to increase precision and reduce civilian casualties are arguments for incorporating greater autonomy into weapons. They are not necessarily arguments for removing human control and creating fully autonomous weapons. Rather, there are more narrow circumstances, where engagement decisions must occur faster than human reaction times or where communications with human controllers are not possible, that human-supervised and fully autonomous weapons may have utility, respectively. Considerations of the potential benefits and risks of autonomous weapons should therefore take into account the likely circumstances for use.

[95] *See* Singer, *supra* note 93 (discussing the situations in which humans can be kept in the loop).