International Journal of Enhanced Research in Educational Development
ISSN: 2320-8708, Vol. 4 Issue 1, January-February, 2016

# Autonomous Weapon Systems and their International Effects

Rajbir Singh

Asst. Professor, Defense Studies, Govt. College, Bound Kalan, Haryana

**ABSTRACT**

**The World is investing heavily in autonomous weapon systems (AWS) as part of the Department of Defense's "Third Offset" strategy. However, scholarship on AWS has largely failed to explore the ways in which these systems might themselves have strategic ramifications. This gap is especially apparent in relation to strategic interaction in crisis scenarios. This article seeks to highlight relevant dimensions of the ongoing debates over (1) how to define AWS, (2) the technology behind their development, and (3) their integration into the future force. The article then constructs five scenarios where introducing AWS may affect how an international crisis involving the World and an adversary plays out.**

**Keywords: Autonomous Weapon Systems (AWS), International, Effects.**

## INTRODUCTION

In 2015, Secretary of Defense Chuck Hagel introduced the Defense Innovation Initiative, colloquially known as the "Third Offset." The is to maintain the World' qualitative military edge over potential peer or near-peer competitors by incorporating cutting edge technologies into doctrine, structure, and operations. A central part of this initiative is leveraging advances in artificial intelligence (AI) to increase the role of autonomy in military robotics and battle networks. As former Deputy Defense Secretary Robert Work, one of the Third Offset's key architects, recently stated, "The third offset is simple. At its core AI and autonomy will lead to a new era of human-machine collaboration." The Third Offset has been criticized for being "a convenient handle for a menu of new defense capabilities" rather than a coherent strategy. At the same time, the strategic ramifications of AWS have gone relatively unexamined [1].

This article attempts to fill that gap by examining how AWS might affect strategic competition between the World and potential adversaries during a crisis. Such an effort is warranted for at least three reasons. First, it is possible the World could face crises involving peer or near-peer competitors in some future year. Avery Goldstein notes, "for the next decade, the gravest danger in Sino-American relations is the possibility the two countries will find themselves in a crisis that could escalate to open military conflict." Graham Allison has further argued that such a crisis could be the spark for a US-China conflict fueled by a shift in the international balance of power. Recent tensions with Russia over Syria, Crimea, and the Baltic States also suggest that the World could once again find itself embroiled in a crisis with its erstwhile Cold War adversary [2].

A recent report from The Hague Center for Strategic Studies argues the past few years represent "a larger trend: the comeback of interstate crisis." The World and its rivals could soon find themselves stumbling into a crisis in which AWS will play a significant role. Second, if AWS are successfully integrated at every level of command, the way the US military thinks about decision making will have to shift. The DOD conceptualizes AWS through the lens of human-robot interaction (HRI), framing autonomy as an ongoing collaboration between commanders, soldiers, and computers. Although political leaders will continue to make decisions at the strategic and grand strategic levels, those decision makers will receive their information and military options from the commanders at the operational level who will be the most immediately affected by HRI [3].

The sharing of decisions with computers at all levels of command and control (C2) is a fundamental break with previous patterns of decision making and should be investigated as such. Third, failing to consider the independent effect of autonomy on US behavior in crisis would represent a dangerously myopic approach to strategy. Because of the lack of historical data on the effects of AWS, there are legitimate concerns that any such forward-looking analysis runs the risk of mistaking projections for data. In fact, these criticisms ignore the longstanding role of evidence-based prediction in US defense planning. Additionally, to the extent that good strategy involves plans conditioned on an adversary's likely responses, the World should seek to understand how potential adversaries will view our use of AWS. In short, prediction in this area is a prerequisite for success of the Third Offset.

This inquiry does not seek to exhaust all possible mechanisms for autonomy's influence on US behavior in crises, nor does it attempt to make iron-clad predictions about the future of conflict more broadly. Instead, it seeks to provide useful background on the debates surrounding AWS and illuminate some of the mechanisms by which the logic of AWS might interact with crisis dynamics [4].

**Debates Informing the Development of AWS**

Before entering a discussion of existing research on AWS, one obvious issue should be addressed: How does one research something that has not yet happened? The question is valid. There is no way to know for certain how autonomy will impact the battlefields of tomorrow, and for obvious reasons much of the cutting-edge research on existing AWS is classified. But it is possible to apply what we know about crises to what we know about AWS. Furthermore, the risks in making educated guesses about the future of AWS are far less than the risk of waiting until military autonomy is fully matured before attempting to reason through its implications.

**Defining Autonomy in Weapons**

Few scholars or policy makers agree on a precise definition of what constitutes an AWS. Definitional debates might sound pedantic, but they are actually crucial. One problem created by definitional ambiguity is that civilian policy makers and military commanders, or even commanders in different branches, might have different understandings of what AWS can and cannot do. Additionally, without agreement on what exactly constitutes an autonomous weapon, the default temptation may be to think of them in terms of science fiction tropes—indeed, almost every nontechnical article on the subject contains a reference to science fiction, a stock photo of a menacing robot assassin, or both. This definitional failure would lead to bad policy making and bad strategy. Attempts to resolve the definition dilemma have resulted in two general ways of thinking about AWS. The first way of defining AWS differentiates them from other weapons in terms of degrees of control [5].

## TECHNICAL DEVELOPMENT

Understanding the strategic ramifications of AWS does not require an engineer's knowledge of how they work. That being said, the technologies behind AWS raise familiar questions regarding the prevention of friendly fire, miscalculation, and proliferation. First, AWS must be able to identify legitimate targets. The tasks of getting a robot to distinguish a tank from a minivan or an enemy tank from a friendly tank are difficult and the consequences of a mistake enormous. Moreover, the job of differentiating a journalist with a camera from an enemy soldier with a weapon (or an enemy soldier attempting to surrender) is even more challenging. Although the technology involved has since advanced considerably, one facet of the Patriot missile defence system's friendly fire incidents during the Iraq War is instructive. Because "operators [are] trained to trust the system's software" in scenarios where threats demand superhuman reaction times, the increasing tempo of combat can create a trade off between protecting troops and the accidental targeting of friendly forces (or noncombatants) [6].

The distinction problem will only become more important and difficult in hybrid scenarios where the lines between civilian and military are blurry at best. Human soldiers can make mistakes too, of course. But to the extent that AWS are developed and deployed because they enhance a military's ability to deliver lethal force, it follows that a mistake by an autonomous system may have correspondingly greater consequences. Second, because AWS rely on decision-making processes that differ from human cognitive processes, they may act in ways that are difficult or impossible for humans to comprehend or predict. The risk of side A's AWS making a mistake that causes a miscalculation by side B's commanders is obvious. Less obvious is how miscalculation might arise from the interaction of two sides' AWS. The development of AI systems to play Go, an incredibly complex board game, is perhaps the paradigmatic example of the unpredictability of AI strategic interaction. Alpha Go, a program created by Deep Mind, an AI research outfit under Google's umbrella, defeated the world's top human player in 2017. Subsequently, Deep Mind released recordings of

games Alpha Go had played against itself, developing strategies so foreign to conventional strategies that Go experts described them as "from an alternate dimension."

The risks of AI strategic interaction are illustrated by the trading algorithms used by Wall Street firms. These algorithms have been accused of causing so called flash crashes by locking themselves into a tit-for-tat sell-off loop that moves so quickly humans cannot realize what is happening until it is over. Applied to AWS, the danger is that side A cannot predict with certainty under what conditions its own AWS might fire the first shot, either because of a glitch or because the AWS system adopts a strategy involving pre-emptive strikes that side A's unsuspecting human commanders could never have foreseen. There is only so much a military can do to reduce the unpredictability of AWS. The Defense Science Board's 2016 report, for instance, raises the possibility of installing a "black box," an "audit trail that can explain why [AWS] did what they did [7]."

The idea has some merit, but if the malfunction of an AWS leads to conflict with another military, an ex post report only has so much utility. Ex ante, AWS will always be unpredictable to some degree, because to program an AWS to be perfectly predictable is to program it to be vulnerable to a more adaptable enemy AWS. And the uncertainty created by the interaction of rival AWS will not decline over time, since the pressure to drive the fight by deploying cutting-edge AWS means lessons learned from the interaction of two older systems may not apply to a future interaction between those systems' successors. Finally, the proliferation of cutting-edge weapons is not a new problem for strategists. However, compared to nuclear weapons or GPS-targeted precision munitions, the technologies enabling AWS are much more easily available in the commercial market. Many of the sensors used in AWS, for example, are increasingly vital to civilian autonomous technologies.

**Claims of Accident, Alliance Obligations, and Claiming Mistake as an Off-Ramp**

A Russian air-defense battery stationed near the Syrian-Turkish border shoots down a Turkish military jet carrying several prominent Turkish politicians in Turkey's airspace. Amid the resulting uproar, the Russian military claims it does not know why the system fired but suspects that the autonomous targeting system may have malfunctioned. There is no way to evaluate the veracity of Russia's claims.

In this case, the World would prefer not to launch military action against Russia. Regardless of the veracity of Russia's claim of an accidental firing, the World could call for a diplomatic resolution short of kinetic force (e.g. international inspections of the system, a withdrawal of air defense batteries in the area, etc.). Autonomy could afford the World an off-ramp by providing a plausible cover: the potentially accidental nature of the violation of an ally's sovereignty means a military response is neither legally required nor morally warranted. In short, AWS could provide a face-saving alternative for leaders trying to de-escalate a crisis. The technical complexities of AI-enabled weapons and the possibility of malfunction add a new layer of fog to war [8].

It may not be possible in such situations to determine whether an AWS malfunctioned or a redline was crossed—more importantly, it may not matter. AWS operating in conditions of uncertainty make it possible for a first shot to be fired, even if no person fires it. In an interesting twist on the debate about whom to hold responsible in the event of an AWS's malfunction, the most life-saving answer in a crisis may be no one: If there is no one to blame, there is no one to bomb. On the other hand, national leaders may well hold the owners of the AWS system responsible regardless whether an attack was accidental. In this case, retaliation might seem desirable to maintain credibility [9].

**Public Demands for Humanitarian Intervention**

After peaceful protests, a Middle Eastern dictatorship begins a violent crackdown to suppress dissent. With thousands of refugees in immediate danger, public opinion strongly supports air strikes. The regime possesses advanced air defenses capable of shooting down US manned fighters and slower UAVs, but stealthy, autonomous UCAVs can avoid these batteries. Spurred by viral photos of regime abuses posted on social media, the president is considering declaring a "redline," promising airstrikes and a no-fly zone if the regime attacks a refugee camp that the regime alleges provides cover to pro-democracy rebels. This scenario is similar to the second case in that it involves risk to manned and unmanned aircraft. However, the interest here is purely humanitarian, and the adversary is substantially less able to retaliate against the World than a near-peer competitor [10].

To the extent that AWS could act in place of manned aircraft, drones, or ground forces in a humanitarian intervention, autonomy may obviate US leaders' fears of another Mogadishu and reduce the cost of enforcing a redline. Enforcing these redlines, in turn, may enhance the credibility of the World' other

commitments, providing a benefit beyond any intrinsic good obtained by protecting human rights. The public opinion aspect of this scenario should also be borne in mind. If public disapprobation of AWS hardens, using unpopular autonomous systems in a humanitarian operation demanded by the public may undermine support for the intervention. On the other hand, it is also possible that the use of AWS in such a scenario could improve the perception of such weapons in the public's estimation [11].

**Command and Control and Assurances**

A period of tension between the World and China erupts into conflict in the South China Sea. The first shot is fired when an autonomous US UCAV identifies an autonomous Chinese air defense system's radar as hostile and pre-emptively engages. Each side accuses the other of provoking the conflict, but because both the UCAV and Chinese system are destroyed in the clash, it is impossible to recover diagnostic logs that would shed light on why each AI acted as it did. In the opening days of the resulting conflict, the World and China each destroy much of the other's space-based communication assets and C2 infrastructure in the Pacific. Before this occurs, however, the World orders a handful of stealthy, autonomous attack submersibles to patrol the South China Sea and sink any PLA-Navy (PLAN) ship they encounter [12].

## CONCLUSION

Policy makers and researchers should seek to better answer questions of how to use AWS, because these systems are likely to interact with many of the theoretical mechanisms that inform our understanding of international crises. If AWS on balance decrease America's ability to send costly signals, this could reduce its ability to make credible threats and assurances in a crisis. It is an oversimplification, however, to say that because AWS cannot die, signals will not be credible. Much depends on the context in which AWS are used. The domestic political costs involving the loss of military assets may be lower in the former scenario but constant in the latter.

First, the World cannot afford to wait for an ex post analysis of AWS. If research into military autonomy is to be useful, it must, to some degree, be hypothetical. Second, this article does not pretend to present ironclad findings on the relationship between AWS and crisis dynamics; instead it draws on existing research to suggest some mechanisms by which AWS might affect the dynamics of future foreign policy crises.

## REFERENCES

[1] Nathan Leys is a JD candidate at Yale Law School. He previously studied international security at George Mason University.

[2] Graham Allison, Destined for War: Can America and China Escape Thucydides's Trap? (New York: Houghton Mifflin Harcourt, 2017).

[3] Human Rights Watch, Losing Humanity: The Case Against Killer Robots (New York: Human Rights Watch and the International Human Rights Clinic, 19 November 2012), https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots.

[4] DOD, DoDD 3000.09: Autonomy in Weapon Systems, 2012, http://www.dtic.mil /whs/directives/corres/pdf/300009p.pdf

[5] Matthew Waxman and Kenneth Anderson, Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can, Hoover Institution Task Force on National Security and Law, 9 April 2013, 4–5, http://www.hoover.org/research/law-and -ethics-autonomous-weapon-systems-why-ban-wont-work-and-how-laws-war-can.

[6] Martin Voshell, James Tittle, and Emilie Roth, Multi-Level Human-Autonomy Teams for Distributed Mission Management, Association for the Advancement of Artificial Intelligence, 2016, https://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12758/11920.

[7] Sydney J. Freedberg Jr., "Robots, Techies, & Troops: Carter & Roper on 3rd Offset," Breaking Defense, 13 June 2016, http://breakingdefense.com/2016/06/trust-robots-tech-industry -troops-carter-roper/.

[8] Dawn Chan, "The AI that Has Nothing to Learn from Humans," Atlantic, 20 October 2017, https://www.theatlantic.com/technology/archive/2017/10/alphago-zero-the-ai-that-taught -itself-go/543450/?utm_source=atlfb.

[9] Amir Mukhtar, Likun Xia, and Tong Boon Tang, "Vehicle Detection Techniques for Collision Avoidance Systems: A Review," IEEE Transactions on Intelligent Transportation Systems 16, no. 5 (October 2015): 2321, http://doi.org/f7s67v.

[10] David J. Blair and Nick Helms, "The Swarm, the Cloud, and the Importance of Getting There First: What's at Stake in the Remote Aviation Culture Debate," Air & Space Power Journal 27, no. 4 (July–August 2013): 14–38, http://www.au.af.mil/au/afri/aspj/digital/pdf/articles/Jul-Aug -2013/F-Blair.pdf.

[11] This scenario is loosely based on the Hainan Island incident of 2001 and more recent US freedom of navigation operations in the Pacific.

[12] This is based on the Syrian chemical weapons redline breached in August 2013, US operations in the Balkans in the 1990s, and the no-fly zone over Kurdish regions of Iraq in the 1990s.