Marine Corps Lance Corporal Donte Mathews, rifleman with 1st Battalion, 8th Marine Regiment, 2nd Marine Division, flys unmanned aircraft system on Camp Lejeune, North Carolina, January 17, 2023 (U.S. Marine Corps/Michael Virtue)

A Framework for Lethal Autonomous Weapons Systems Deterrence

By Steven D. Sacks

A s the United States and the People's Republic of China (PRC) continue down a path of increasing rivalry, both nations are investing heavily in emerging and dis-

Captain Steven D. Sacks, USMCR, is a Private-Sector Security and Risk Advisor based out of Washington, DC.

ruptive technologies in search of competitive military advantage. Artificial intelligence (AI) is a major component of this race. By leveraging the speed of computers, the interconnectedness of the Internet of Things, and big-data algorithms, the United States and the PRC are racing to make the next leading discovery in the field. Both nations endeavor to incorporate AI

into weapons systems and platforms to form *lethal autonomous weapons systems* (LAWS), which are defined as weapons platforms with the ability to select, target, and engage an adversary autonomously, with minimal human inputs into their processes.[1] Without a clear framework through which to assess interactions between LAWS of different nations, the likelihood of accidental

or inadvertent escalation to military crisis increases. *Accidental escalation* is an unintended consequence of events that were not originally intentional, whereas *inadvertent escalation* is a situation in which an actor's intended actions are unintentionally escalatory toward another.[2] This article explores how LAWS affect deterrence among Great Powers, developing a framework to better understand various theories' applicability in a competition or crisis scenario between nations employing these novel lethal platforms.[3]

Deterrence

In *Deterrence in American Foreign Policy*, Alexander L. George and Richard Smoke define *deterrence* as "simply the persuasion of one's opponent that the costs and/or risks of a given course of action he might take outweigh the benefits."[4] The act of persuasion relies on psychological characteristics of the actors in a potential conflict scenario. By leveraging an understanding of an opponent's motivations to generate signals of allegedly guaranteed reactions the sending nation will take if provoked, that sending nation is signaling both its capability and its will to fight.[5] Deterrence can be further broken down into *direct deterrence*, a state's dissuading an adversary from attacking its sovereign territory, and *extended deterrence*, the act of dissuading an aggressor from attacking a third party, usually a partner or ally.[6] This article focuses on the latter, specifically looking at concepts that would be applicable to the U.S. attempt to deter the PRC from conducting aggressive military operations against a partner or ally in the Indo-Pacific region. The proffered framework also applies to scenarios in which the PRC attempts to deter the United States from third-party intervention subsequent to a fait accompli aggressive action against an American partner or ally.

According to George and Smoke's definition, to increase the effectiveness of deterrence a state must either increase the cost of the aggressor state's escalation or expand the overall risk of increased aggression within the relationship. James

Fearon's "tying hands" and "sinking costs" are two methods by which a country can signal to another its level of resolve if attacked. Tying hands links the credibility of political leadership to a response to foreign aggression; sinking costs involves deploying forces overseas, incurring *ex ante* costs that signal military resolve.[7] Glenn Snyder further expounds on the sunk cost theory, introducing the idea of a "plate-glass window" of deployed troops that an aggressor must shatter to attempt any offensive action against a third country.[8] The shattering of the plate-glass window is understood as an assured trigger for third-party intervention, exemplified historically by the U.S. decision in 1961 to deploy an Army brigade to West Berlin meant to deter a Soviet invasion of the city.[9]

The Department of Defense has defined the *endstate of deterrence* as the ability to "decisively influence the adversary's decisionmaking calculus in order to prevent hostile action against U.S. vital interests."[10] To achieve this end, the U.S. military conducts global operations and activities that affect the ways adversaries view threats and risks to their own national security. More recently, American military leadership has emphasized deterrence as the desired endstate of a defending country's military strategy, separate and distinct from compellence.[11] Chinese scholars, in contrast, discuss deterrence as more analogous to Thomas Schelling's overall characterization of coercion, melding the concept of deterrence with that of compellence.[12] These scholars view deterrence in a similar manner to Maria Sperandei's "'Blurring the Boundaries': The Third Approach," acknowledging the often overlapping relationship between deterrence and compellence, in which one can easily be framed in the context of the other.[13] Additionally, Chinese authors see deterrence as a milestone that supports setting conditions, which then enable the achievement of more strategic political endstates, rather than an endstate itself.[14] Chinese military scholars have written about the use of limited kinetic force as a deterrent, showing the adversary an example of PRC military capabilities to

dissuade the potential aggressor from taking any actions.[15] The use of kinetic weapons platforms as a deterrent likely increases the risk of *inadvertent escalation*, defined as when "one party deliberately takes actions that it does not believe are escalatory but [that] another party to the conflict interprets as such," thereby making the competition more volatile.[16] Leaders within the People's Liberation Army (PLA) almost certainly view their introduction of AI and LAWS as contributing to competitive military advantage while simultaneously setting favorable conditions for conflict should the relationship escalate by deploying and employing these capabilities among PLA units.[17] One concern with Chinese writings on deterrence is the yet-unreconciled tension between the dual goals of deterring escalation and simultaneously preparing the battlefield; they lack assessments regarding which deterrent activities risk interpretation as escalatory by their adversaries.[18]

Even as PLA writers look to the military application of AI to generate control, the lack of available scholarly work on how the United States will interpret its introduction is cause for concern.[19] The PLA's theory of military victory is based on its ability to effectively control the escalation of the conflict, employing both deterrence and compellence principles to achieve strategic political goals in a predictable manner that leaves Beijing in the driver's seat of conflict.[20] Although a 2021 RAND report on deciphering Chinese deterrence signals establishes a framework by which the United States can better understand PLA military deterrence signals, a more comprehensive understanding of effective deterrent signaling between the United States and China remains elusive.[21] As long as this gap persists, there remains a high risk of inadvertent escalation to major conflict due to misunderstanding as new technologies and capabilities are phased in to the militaries of both nations.

Employment of Autonomy in Warfare

AI is the employment of computers to enable or wholly execute tasks and/or decisions to generate faster, better,

or less resource-intensive results than if a human were completing the task. AI applies across disciplines, from conducting light-speed stock market trades to performing supply chain risk analysis. AI brings speed-of-machine decisionmaking that often frees human resources to focus on more complex tasks, making it a useful means within the current Great Power competitive dynamic to gain advantages against adversaries in a resource-constrained environment.[22] The Chinese government has allocated increasing resources to the development of disruptive capabilities such as AI as a key pillar of its national strategy, leveraging science and technology as part of the PRC's pursuit of Great Power status.[23]

AI encompasses a spectrum of capabilities that leverages computers to increase speed, reduce costs, and limit the requirement for human involvement in task and decision processes. Within AI, there are two concepts that play critical roles in understanding how LAWS affect conventional deterrence theory: machine learning (ML) and autonomy. ML employs techniques that often rely on large amounts of data to train computer systems to identify trends and analyze best courses of action.[24] An AI system's ability to learn depends on the quality and quantity of data. More pertinent data available across a wide spectrum of relevant scenarios allow the ML algorithms to train to handle a wider range of situations. The better the ML code training, the more autonomous a system can become. Regarding the second concept, autonomy, there exists a spectrum, from "human-AI co-evolving teams," in which both parties mature together on the basis of mutual interactions over long periods of time, to "human-biased AI executing effects," in which the autonomous platform reacts rapidly to its environment in a manner informed by human input and set parameters.[25] From enhancing logistics operations through predictive supply chain modifications to reducing commanders' uncertainty through sensor proliferation and programmed analysis, autonomous systems can provide significant benefits to militaries able and willing to incorporate them into emerging concepts of operations.[26]

The use of LAWS in combat affects the application of deterrence through the manipulation of cost-benefit analyses conducted by the actors in conflict. Replacing human assets with unmanned equivalents diminishes the risk of human losses from military engagements, potentially changing the escalation calculus for militaries that place a high value on human life.[27] By decreasing the risk of human casualties, the introduction of LAWS may reduce the political barriers hindering a decision to launch escalatory military operations, thereby increasing the potential for large-scale conflict.[28] Reducing these barriers to escalation further increases the risk of inadvertent or accidental escalation, in combination with the uncertainties brought about by relegating increasing amounts of decisionmaking authority from humans to weaponized battlefield platforms. The effects of emerging and disruptive technologies and operations in the United States and the Soviet Union during the Cold War were counterbalanced and the situation stabilized through a mutually understood framework of deterrence. The advent of emerging and disruptive LAWS, combined with a lack of established messaging and signaling norms, is destabilizing to the future of the U.S.-PRC relationship.

One aspect of the introduction of autonomy to the battlefield that does not deal directly with deterrence but remains relevant is the potential for increased autonomy to result in degraded control of systems by human military commanders and leaders. Both Washington and Beijing have made it clear that human involvement in weapons systems engagement decisions remains a priority. In 2012, the Pentagon released a directive mandating that autonomous and semi-autonomous weapons systems be designed to allow humans appropriate oversight and management of the use of force employed by those systems.[29] These decisionmaking processes will also remain squarely within the legal boundaries of the codified rules of engagement and law of war. China's military has remained more ambiguous as to its stance on the use of autonomy in lethal warfare.

Beijing has both called for the prohibition of autonomous weapons, through a United Nations binding protocol in 2016, and issued its New Generation of AI Development Plan, in 2017—which served as the foundation for its development of autonomous weapons.[30] Both nations have shown hesitance to deploy fully autonomous lethal weapons systems to the battlefield; however, with emerging technologies and innovations, that reluctance may change.

Brinkmanship and Signaling

In *Arms and Influence*, Schelling describes brinkmanship as a subset of deterrence theory defined by two actors pushing the escalation envelope closer to total war; brinkmanship must include elements of "uncertainty or anticipated irrationality or it won't work."[31] In the Cold War era, uncertainty was driven by human psychology and external actors—would a military leader take it upon him- or herself to make aggressive moves that might initiate a limited conflict, or would a third party take action that would force one of the belligerents to respond offensively? In the era of AI, ML, and LAWS, uncertainty is also derived from the unpredictability of the system code itself.[32] The amount of trust practitioners can place in their LAWS is limited to the breadth, depth, and quality of the data and scenarios in which the platform is tested and evaluated—a concern because real-world combat often lies outside of training estimates.[33] The problem of amassing sufficient quantities of data with necessary fidelity and relevancy to future operations is compounded by the pace of the change to the character of warfare brought about by the implementation of AI and ML on the battlefield.[34] All of these factors challenge the ability to generate human trust in LAWS, given the increased levels of uncertainty about their predictable performance across a spectrum of military operations.[35]

This uncertainty in the reliability of autonomous weapons presents a security dilemma among Great Powers because the side with more lethal platforms gains a greater first-strike advantage over time.[36]

MQ-8B Fire Scout unmanned aircraft system from Helicopter Maritime Strike Squadron (HSM) 35 performs ground turns aboard littoral combat ship USS *Fort Worth*, South China Sea, May 1, 2015 (U.S. Navy/Conor Minto)

The speed at which computers make decisions also enhances the effect of autonomous unpredictability on brinkmanship.[37] Additionally, adversaries can hack LAWS code to degrade or deny operational capability, introducing further uncertainty into autonomous warfare.[38] An early example of autonomous unpredictability occurred in 2017, when the Chinese Communist Party developed automated Internet chatbots to amplify party messaging; the bots gradually began to stray off message, culminating in posts criticizing the party as "corrupt and incompetent" before officials took the software offline.[39]

The concept of private information also contributes to uncertainty and brinkmanship. Private information is privileged knowledge about capabilities and intentions known only to the originating country. Nations have an incentive to keep private information hidden from adversaries to generate a tailored external perception favorable to the owner of the information.[40] But countries can deliberately reveal private information to external actors through signaling—the sending of a calculated message to a target audience to convey specific information for a desired effect. To be successful, a signal must be received and interpreted as intended by the sender. State leaders and administrations, however, are prone to misperception because of inherent biases that influence their reception of signals.[41] The ability to successfully signal capabilities and intentions regarding LAWS is complicated

by the uncertainty introduced by the employment of autonomous algorithms. There remains a dearth of research exploring how emerging robotics will potentially affect the successful conveyance of deterrent signals.[42]

Separate but not necessarily distinct from the ability to signal capability while retaining the advantage of private information is the ability to signal intent. Experts including Robert Jervis have explored the ability of states to increase national security without falling victim to the security dilemma by developing overt distinctions between weapons systems with offensive versus defensive intents. Jervis writes, "When defensive weapons differ from offensive ones, it is possible for a state to make itself more secure

Digital recreation of Holloman High-Speed Test Track with ejection seat test setup created as part of Project Zero, in effort to train drones through machine learning to conduct automated, artificial-intelligent driven operations and data analysis, Holloman Air Force Base, New Mexico, June 6, 2022 (U.S. Air Force)

without making others less secure."[43] Table 1 depicts the two variables Jervis assessed, offense-defense distinguishability and offense-defense advantage in conflict, to create quadrants describing "worlds" of risk conditions. This framework is especially applicable to overlay with current concepts of deterrence by punishment, where offense has the advantage, and deterrence by denial, where defense has the advantage.

Decisions by Washington and Beijing to prioritize private information and operations security surrounding the development and testing of LAWS inhibit the diffusion of technology to the private or commercial sector or to other national militaries, even when those external entities may have technological advantages over national military capabilities. By compartmenting the technology at the foundation of AI-enabled warfighting platforms, these decisions make it difficult to distinguish the military intent of these capabilities—whether they are for offensive or defensive posturing. Additionally, proprietary and

classified LAWS enhance first-mover advantage as each Great Power is racing to develop measures and countermeasures to provide its military a battlefield advantage.[44] This effect is further highlighted in Chinese military strategy, which stresses the importance of seizing and maintaining the initiative in conflict, often through rapid escalation across domains, before an adversary has a chance to react or respond—a fait accompli campaign.[45] An inability to distinguish defensive systems from offensive ones employed in a world where offensive first movers have the advantage places the situation in Jervis's "doubly dangerous" world.

The inability to trace autonomous decision processes further challenges the ability to predict and understand the effectiveness of signaling through LAWS. Neural networks at the core of AI decisionmaking are characterized as "black boxes," offering minimal insight into the impetus behind their autonomous assessments or decisions.[46] Without the ability to analyze how these

algorithms make decisions, engineers struggle to make reliable cause-to-effect assessments to determine how the autonomous systems can be expected to act in specific situations. Recent wargames have demonstrated that autonomous systems are less capable of understanding signals and therefore are more prone to unpredictable decision-making than humans. These systems are often programmed to maximize decision speed and to seek out perceived exploitable opportunities to capitalize on rapidly. These priorities make them more likely to escalate battlefield engagements in situations where a human would be reluctant to deviate from the status quo.[47] Deploying LAWS into the competition domain thus introduces novel signaling opportunities: the ability to overtly switch a weapons system to autonomous operation, unswayed by outside factors or emotions, can indicate military determination, taking the decision to initiate aggressive defensive

Table 1. Jervis's Four Possible Worlds

	Offense Has the Advantage	Defense Has the Advantage
Offensive Posture Not Distinguishable from Defensive One	1 Doubly dangerous	2 Security dilemma, but security requirements may be compatible
Offensive Posture Distinguishable from Defensive One	3 No security dilemma, but aggression possible. Status-quo states can follow different policy than aggressors. Warning given.	4 Doubly stable

Source: Robert Jervis, "Cooperation Under the Security Dilemma," *World Politics* 30, no. 2 (January 1978), 211.

actions out of human hands, should a preprogrammed red line be crossed.[48]

There is the potential that the unpredictability in the LAWS decisionmaking process constitutes its own deterrent. In a scenario where the adversary cannot assess with confidence how an autonomous weapons system will act in a specific battlefield situation, there is the potential that the adversary will be dissuaded from initiating an attack for fear of an unknown ability that eclipses the adversary's own. However, a more effective use of unpredictability resides at the operational rather than the tactical level of warfare. By reliably revealing a new lethal autonomous capability during a large-scale demonstration or exercise, the United States can show that it has more operational options for military forces at its disposal.[49] There is a likelihood that the PRC will observe a new demonstrated capability and infer that the United States is concealing even more capable and lethal proficiencies.[50] Both of these effects would lend themselves to the conclusion that revealing a novel LAWS capability may have more deterrent impact than concealing it.

A Framework for Deterrence With Autonomous Weapons Systems

Two critical factors determine how LAWS affect deterrence in future warfare: predictable lethality of the weapons systems and effective signaling of that lethality to adversaries. Table 2 describes four possible permutations of deterrence through the use of LAWS in a naval blockade scenario. In these scenarios, a defending nation has established a naval blockade using LAWS deployed in permanent autonomous modes of operations by their human users and coded to engage any foreign platform that approaches within a set distance from the blockade. The aggressor state is advancing toward the blockade with manned platforms, threatening offensive action against the defender. The defending nation has attempted to signal to the aggressor that the unmanned blockade has been switched to autonomous mode and will attack the advancing adversary if it crosses the red line of proximity.

In the table's Tripwire Deterrence quadrant, the defending nation possesses predictability in the lethal autonomous weapons systems' ability to execute their decisionmaking processes as intended, and it has effectively signaled this capability to the advancing force. In this scenario, uncertainty is minimized; both sides understand the red line and how the autonomous blockade will react to a crossing. Because the role of humans is minimized in the decision loop of AI systems operating on the "human-biased" side of the autonomous spectrum, individual psychology and emotions do not inject unpredictability into the engagement, resulting in what Schelling describes as a defensive tripwire.[51] In Tripwire Bluff, the defenders have effectively signaled to adversaries the lethal autonomous weapons systems' predictable lethality; however, the purported predictability is not manifest in reality. Either the autonomous systems in the blockade are untested, or they have been tested with inconsistent results. In this scenario, the defender is successfully bluffing a tripwire defense to the adversary.

In Single-Side Uncertainty, the defender has confirmed predictable lethality from its blockade but has failed to effectively signal this capability to the advancing aggressor. In this scenario, the aggressor is unsure whether to believe that the blockade will operate as intended and is subsequently faced with making a decision handicapped by the uncertainty about the defender's true capabilities. In Brinkmanship Deterrence, the defending blockade does not possess predictable lethality, nor has the defender effectively communicated that capability to the adversary; both sides are uncertain how the blockade will react to aggressor action.

Of the scenarios described above, Tripwire Deterrence brought about by LAWS is the most stable because private information is minimized. In this context, both the sender and receiver of the deterrence signals understand the capabilities of the autonomous weapons platforms and know under what conditions these platforms will initiate action against an adversary. Tripwire Bluff situations are stable only so long as the nation receiving the deterrence signal does not become privy to the unpredictability of the autonomous

Table 2. Effect of Lethal Autonomous Weapons Systems in Brinksmanship Deterrence

		PREDICTABLE CAPABILITY	
		YES	**NO**
EFFECTIVE SIGNALING	**YES**	**Tripwire Deterrence** ■ Reduction of uncertainty reduces applicability of brinksmanship ■ Red lines are known by both sides ■ LAWS capabilities are demonstrated as consistent across scenarios	**Tripwire Bluff** ■ Algorithm unpredictability increases applicability of brinksmanship ■ Red lines are known by both sides ■ LAWS capabilities are undemonstrated to the adversary force
	NO	**Single-Side Uncertainty** ■ Reduction of uncertainty reduces applicability of brinksmanship ■ Red lines are left ambiguous, opening door for aggressor escalation ■ LAWS capabilities are demonstrated as consistent across scenarios	**Brinksmanship Deterrence** ■ Algorithm unpredictability increases applicability of brinksmanship ■ Red lines are left ambiguous, opening door for aggressor escalation ■ LAWS capabilities are undemonstrated or demonstrated as inconsistent

systems being employed by the signal-sending nation. This scenario may arise through deceptive practices, whereby the signaling nation projects a level of autonomous predictability in operations that it has yet to achieve in reality. The danger of this environment is that the signal-receiving nation may begin to doubt the true abilities of the signal-sending nation, incentivizing it to call the signaling nation's bluff and escalate to seize a competitive military advantage.

In a Brinkmanship Deterrence scenario, autonomous systems are not mature enough to produce predictable results across a wide array of situations, possibly because of a lack of sufficient quantity or quality of data with which to train. As the data increase in both amount and relevance, LAWS are more likely to operate in a realiable manner,

transitioning to a Single-Side Uncertainty environment. In Single-Side Uncertainty, the signal-sending nation knows its autonomous systems perform predictably, but the receiving nation is unaware of this fact. This scenario might be brought about because the signaling nation has kept the testing and experimentation of its autonomous weapons platform secret, denying the receiving nation the ability to observe and assess the reliability of its performance. This scenario may also be driven by a perception by the signal-receiving nation that the autonomous system has not been sufficiently tested in a realistic environment representative of the future battlefield. If provided an opportunity to confirm the reliable performace of the LAWS, the signal-receiving nation ideally becomes aware of the circumstances under which the

autonomous system will perform its intended functions, driving the competitive dynamic into stable Tripwire Deterrence.

Why the Framework Is Relevant Today

The above framework highlights the critical role signaling plays in the effectiveness of the LAWS contribution to deterrence. Systems with an AI core introduce unpredictability for both the employer of the system and adversaries. States will be faced with the tension between needing to openly test their algorithms in the most realistic scenarios and simultaneously protecting proprietary information from foreign collection and exploitation, resulting in deliberate ambiguity. The overt testing of the LAWS capabilities reduces uncertainty for the LAWS user and

Corporal Thomas Rexrode, reconnaissance Marine with Company A, 1st Reconnaissance Battalion, 1st Marine Division, launches RQ-20B Puma small unmanned aircraft system from rigid-hull inflatable boat at Camp Pendleton, California, September 30, 2021 (U.S. Marine Corps/Connor Hancock)

Army paratrooper uses Dronebuster 3B to disrupt enemy drones as part of Exercise Shield 23 in Pula, Croatia, April 20, 2023 (U.S. Army/Mariah Y. Gonzalez)

signals capability to potential aggressors; however, the protection and deliberate obfuscation of such experiments help retain the exclusivity of capabilities and reduce the risk of an AI-fueled security dilemma between Great Powers.[52] The above framework promotes the argument that deterrence is better served through open testing and evaluation, contributing to more effective signaling of the LAWS capabilities. Recent studies have shown that under conditions of incomplete information the initial messaging of capability and intent is the most effective in deferring conflict; lack of clarity in that signal invites adversaries to pursue opportunistic aggression.[53] Effective signaling is only made more complex once autonomous systems are tasked with receiving and interpreting the messages and signals originating from other autonomous platforms.

PLA strategists expect that the future of combat lies in the employment of unmanned systems, manned-unmanned teaming, and ML-enabled decisionmaking processes designed to outpace the adversary's military cycles of operations. These advances should reduce identified shortfalls in the ability of PLA leadership to make complex decisions in uncertain

situations.[54] In 2013, the PLA's Academy of Military Science released a report arguing that strategic military deterrence is enhanced by not only cutting-edge technology but also the injection of unpredictability and uncertainty in adversary assessments through new military concepts and doctrine.[55] The advent of LAWS contributes new uncertainty to China's ability to predict the actions of its own forces and challenges the PLA's ability to achieve effective control over the behavior of adversary autonomous systems on the battlefield—both of which have the potential to raise the risk of accidental escalation and thus major conflict.

The attractiveness of unmanned replacements can be observed in China's current AI military research prioritizing autonomous hardware solutions, ranging from robotic tanks and autonomous drone swarms to remote-controlled submarines.[56] Some in the PRC quickly recognized the disruptive potential of LAWS coupled with swarm tactics, defining a concept of "intelligentized warfare" as the next revolution in military affairs, which would dramatically affect traditional military operational models.[57] Intelligentized warfare is defined by AI at its core, employing cutting-edge technologies within operational command,

equipment, tactics, and decisionmaking across the tactical, operational, and strategic levels of conflict.[58] But intelligentized warfare also expands beyond solely AI-enabled platforms, incorporating new concepts of employment of human-machine integrated units where autonomous systems and software play dominant roles.[59] One example of a new concept of employment for PLA autonomous systems is "latent warfare," in which LAWS are deployed to critical locations in anticipation of future conflict, loitering in those locations and programmed to be activated to conduct offensive operations against the adversary's forces or critical infrastructure.[60]

The U.S. military, too, is looking to AI and LAWS as a key pillar of achieving its desired endstates on current and future battlefields. American military leaders see autonomous systems as presenting a wide array of protection and lethality possibilities, while concurrently providing commanders an ability to make faster and better-informed decisions in both competition and crisis.[61] As both the PRC and United States pursue disruptive capabilities and concepts of military operations with LAWS, the lack of a mutually understood framework through which to interpret each other's actions in competition significantly

increases the risk of inadvertent escalation to crisis and conflict. Additionally, the criticality of quality adversary data in sufficient quantity to ensure predictable LAWS performance in conflict has the potential to drive an increase in military deception as a means to deny an adversary trust in the data and therefore trust in the platforms' performance against a real enemy.

Conclusion

As nations around the world continue to pursue lethal autonomous platforms for use on the battlefield, the lack of a commonly understood framework for their employment increases the risk of inadvertent or accidental escalation due to miscommunication or misinterpretation of deterrent signals in competition and crisis. A desire to gain and maintain a competitive edge in the military domain often creates incentive for the compartmentalization of information about emerging and disruptive battlefield technologies. However, if the desired endstate of the U.S. military is to achieve effective deterrence, and the future battlefield is anticipated to include myriad LAWS, then the framework proffered here recommends limiting private information in the process of acquisitions and development. Once the predictability of an autonomous platform has been established by a nation, the ability for an adversary to observe and assess that predictability enhances the stability of deterrence through effective signaling. Additionally, relevant data of both friendly and adversary information will become a premium as nations attempt to develop LAWS that can operate across the widest spectrum of scenarios, potentially driving an increase in military deceptive activities in steady state.

As the implementation of LAWS expands from a situation where autonomous systems serve as deterrent signals to a world where autonomous systems are tasked with interpreting and responding to deterrent signals, additional research will be required to help refine the above framework. Such research would likely benefit from a focus on the willingness of governments to delegate decisionmaking authority to LAWS. The Chinese Communist Party prizes centralized control over the military, which makes delegation less likely. However, Beijing also remains distrustful of the decisionmaking capabilities of its officer corps, making delegation more appealing as a means to mitigate observed shortfalls in PLA decisionmaking abilities.[62] Both policymakers and scholars could also explore the effectiveness of signaling and deterrence across variations of intermixed manned and unmanned networked systems because the increased risk of loss of human life coupled with the introduction of psychology and emotions to decisionmaking processes could affect the escalatory dynamic.[63] **JFQ**

Notes

[1] Alex S. Wilner, "Artificial Intelligence and Deterrence: Science, Theory and Practice," in *Deterrence and Assurance Within an Alliance Framework*, STO-MP-SAS-141 (Brussels: NATO Science and Technology Organization, January 18, 2019), 6.

[2] Forrest E. Morgan et al., *Dangerous Thresholds: Managing Escalation in the 21st Century* (Santa Monica, CA: RAND, 2008), 23–26, https://www.rand.org/pubs/monographs/MG614.html.

[3] Joint Doctrine Note 1-19, *Competition Continuum* (Washington, DC: The Joint Staff, June 3, 2019), 2, https://www.jcs.mil/Portals/36/Documents/Doctrine/jdn_jg/jdn1_19.pdf.

[4] Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy: Theory and Practice* (New York: Columbia University Press, 1974), 11.

[5] Michael J. Mazarr, "Understanding Deterrence," in *NL ARMS: Netherlands Annual Review of Military Studies 2020*, ed. Frans Osinga and Tim Sweijs (The Hague: T.M.C. Asser Press, 2020), 14–15, https://library.oapen.org/bitstream/handle/20.500.12657/47298/9789462654198.pdf?sequence=1&isAllowed=y.

[6] Michael C. Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power," *Texas National Security Review* 1, no. 3 (May 2018), 8–9, https://repositories.lib.utexas.edu/bitstream/handle/2152/65638/TNSR-Vol-1-Iss-3_Horowitz.pdf?sequence=2&isAllowed=y.

[7] James D. Fearon, "Signaling Foreign Policy Interests: Tying Hands Versus Sinking Costs," *The Journal of Conflict Resolution* 41, no. 1 (February 1997), 68, https://web.stanford.edu/group/fearon-research/cgi-bin/wordpress/wp-content/uploads/2013/10/Signaling-Foreign-Policy-Interests-Tying-Hands-versus-Sinking-Costs.pdf.

[8] Glenn H. Snyder, *Deterrence and Defense: Toward a Theory of National Security* (Princeton: Princeton University Press, 1961), 7.

[9] Adam Lockyer, "The Real Reasons for Positioning U.S. Forces Here," *Sydney Morning Herald*, November 24, 2011, 1, https://www.smh.com.au/politics/federal/the-real-reasons-for-positioning-us-forces-here-20111124-1v1ik.html.

[10] *Deterrence Operations Joint Operating Concept, Version 2.0* (Washington, DC: The Joint Staff, December 2006), 19, https://apps.dtic.mil/sti/pdfs/ADA490279.pdf.

[11] C. Todd Lopez, "Defense Secretary Says 'Integrated Deterrence' Is Cornerstone of U.S. Defense," Department of Defense, April 30, 2021, https://www.defense.gov/Explore/News/Article/Article/2592149/defense-secretary-says-integrated-deterrence-is-cornerstone-of-us-defense/.

[12] Thomas C. Schelling, *Arms and Influence* (New Haven, CT: Yale University Press, 1966); Dean Cheng, "An Overview of Chinese Thinking About Deterrence," in *NL ARMS: Netherlands Annual Review of Military Studies 2020*, 178.

[13] Maria Sperandei, "Bridging Deterrence and Compellence: An Alternative Approach to the Study of Coercive Diplomacy," *International Studies Review* 8, no. 2 (June 2006), 259.

[14] Cheng, "An Overview of Chinese Thinking About Deterrence," 179.

[15] Alison A. Kaufman and Daniel M. Hartnett, *Managing Conflict: Examining Recent PLA Writings on Escalation Control* (Arlington, VA: CNA, February 2016), 53, https://www.cna.org/reports/2016/drm-2015-u-009963-final3.pdf.

[16] Herbert Lin, "Escalation Risks in an AI-Infused World," in *AI, China, Russia, and the Global Order: Technological, Political, Global, and Creative Perspectives*, ed. Nicholas D. Wright (Washington, DC: Department of Defense, December 2018), 136, https://nsiteam.com/social/wp-content/uploads/2018/12/AI-China-Russia-Global-WP_FINAL.pdf.

[17] Ryan Fedasiuk, *Chinese Perspectives on AI and Future Military Capabilities*, CSET Policy Brief (Washington, DC: Center for Security and Emerging Technology, 2020), 13, https://cset.georgetown.edu/publication/chinese-perspectives-on-ai-and-future-military-capabilities/.

[18] Burgess Laird, *War Control: Chinese Writings on the Control of Escalation in Crisis and Conflict* (Washington, DC: Center for a New American Security, March 30, 2017), 9–10, https://www.cnas.org/publications/reports/war-control.

[19] Ibid., 6.

[20] John Dotson and Howard Wang, "The 'Algorithm Game' and Its Implications for Chinese War Control," *China Brief* 19, no. 7

(April 9, 2019), 4, https://jamestown.org/program/the-algorithm-game-and-its-implications-for-chinese-war-control/.

21 Nathan Beauchamp-Mustafaga et al., *Deciphering Chinese Deterrence Signalling in the New Era: An Analytic Framework and Seven Case Studies* (Canberra: RAND Australia, 2021), https://doi.org/10.7249/RRA1074-1.

22 Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power," 53.

23 Elsa B. Kania, *"AI Weapons" in China's Military Innovation* (Washington, DC: The Brookings Institution, April 2020), 2, https://www.brookings.edu/wp-content/uploads/2020/04/FP_20200427_ai_weapons_kania_v2.pdf.

24 Lin, "Escalation Risks in an AI-Infused World," 134.

25 Jason S. Metcalfe et al., "Systemic Oversimplification Limits the Potential for Human-AI Partnership," *IEEE Access* 9 (2021), 70242–70260, https://ieeexplore.ieee.org/document/9425540.

26 Wilner, "Artificial Intelligence and Deterrence," 9.

27 Ibid., 2.

28 Erik Lin-Greenberg, "Allies and Artificial Intelligence: Obstacles to Operations and Decision-Making," *Texas National Security Review* 3, no. 2 (Spring 2020), 61, https://repositories.lib.utexas.edu/bitstream/handle/2152/81858/TNSRVol3Issue2Lin-Greenberg.pdf?sequence=2&isAllowed=y.

29 Department of Defense Directive 3000.09, *Autonomy in Weapons Systems* (Washington, DC: Office of the Under Secretary of Defense for Policy, January 25, 2012), 15, https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf.

30 Putu Shangrina Pramudia, "China's Strategic Ambiguity on the Issue of Autonomous Weapons Systems," *Global: Jurnal Politik Internasional* 24, no. 1 (July 2022), 1, https://scholarhub.ui.ac.id/cgi/viewcontent.cgi?article=1229&context=global.

31 Schelling, *Arms and Influence*, 99.

32 Michael C. Horowitz, "When Speed Kills: Lethal Autonomous Weapon Systems, Deterrence, and Stability," *Journal of Strategic Studies* 42, no. 6 (August 22, 2019), 774.

33 Lin, "Escalation Risks in an AI-Infused World," 143.

34 Meredith Roaten, "More Data Needed to Build Trust in Autonomous Systems," *National Defense*, April 13, 2021, https://www.nationaldefensemagazine.org/articles/2021/4/13/more-data-needed-to-build-trust-in-autonomous-systems.

35 S. Kate Devitt, "Trustworthiness of Autonomous Systems," in *Foundations of Trusted Autonomy*, ed. Hussein A. Abbass, Jason Scholz, and Darryn J. Reid (Cham, Switzerland: Springer, 2018), 172, https://link.springer.com/chapter/10.1007/978-3-319-64816-3_9.

36 Elsa B. Kania, *Chinese Military Innovation in Artificial Intelligence*, Testimony Before the U.S.-China Economic and Security Review Commission, Hearing on Trade, Technology, and Military-Civil Fusion, June 7, 2019, https://www.uscc.gov/sites/default/files/June 7 Hearing_Panel 1_Elsa Kania_Chinese Military Innovation in Artificial Intelligence_0.pdf; Caitlin Talmadge, "Emerging Technology and Intra-War Escalation Risks: Evidence From the Cold War, Implications for Today," *Journal of Strategic Studies* 42, no. 6 (2019), 879.

37 Horowitz, "When Speed Kills," 766.

38 Lin-Greenberg, "Allies and Artificial Intelligence," 65.

39 Louise Lucas, Nicolle Liu, and Yingzhi Yang, "China Chatbot Goes Rogue: 'Do You Love the Communist Party?' 'No,'" *Financial Times*, August 2, 2017.

40 James D. Fearon, "Rationalist Explanations for War," *International Organization* 49, no. 3 (Summer 1995), 381, https://web.stanford.edu/group/fearon-research/cgi-bin/wordpress/wp-content/uploads/2013/10/Rationalist-Explanations-for-War.pdf.

41 Robert Jervis, "Deterrence and Perception," *International Security* 7, no. 3 (Winter 1982–1983), 4, https://academiccommons.columbia.edu/doi/10.7916/D8PR7TT5; Richard Ned Lebow and Janice Gross Stein, "Rational Deterrence Theory: I Think, Therefore I Deter," *World Politics* 41, no. 2 (January 1989), 215–216.

42 Shawn Brimley, Ben FitzGerald, and Kelley Sayler, *Game Changers: Disruptive Technology and U.S. Defense Strategy* (Washington, DC: Center for a New American Security, September 2013), 20, https://www.files.ethz.ch/isn/170630/CNAS_Gamechangers_BrimleyFitzGeraldSayler_0.pdf.

43 Robert Jervis, "Cooperation Under the Security Dilemma," *World Politics* 30, no. 2 (January 1978), 187.

44 James Johnson, "The End of Military-Techno *Pax Americana*? Washington's Strategic Responses to Chinese AI-Enabled Military Technology," *The Pacific Review* 34, no. 3 (2021), 371–372.

45 Morgan et al., *Dangerous Thresholds*, 57.

46 Lin-Greenberg, "Allies and Artificial Intelligence," 69.

47 Yuna Huh Wong et al., *Deterrence in the Age of Thinking Machines* (Santa Monica, CA: RAND, 2020), 66, https://www.rand.org/pubs/research_reports/RR2797.html.

48 Ibid., 52.

49 Miranda Priebe et al., *Operational Unpredictability and Deterrence: Evaluating Options for Complicating Adversary Decision-making* (Santa Monica, CA: RAND, 2021), 28, https://www.rand.org/pubs/research_reports/RRA448-1.html.

50 Brendan Rittenhouse Green and Austin Long, "Conceal or Reveal? Managing Clandestine Military Capabilities in Peacetime Competition," *International Security* 44, no. 3 (Winter 2019–2020), 48.

51 Schelling, *Arms and Influence*, 99.

52 Chris Meserole, "Artificial Intelligence and the Security Dilemma," *Lawfare*, November 4, 2018, https://www.lawfareblog.com/artificial-intelligence-and-security-dilemma.

53 Bahar Leventoğlu and Ahmer Tarar, "Does Private Information Lead to Delay or War in Crisis Bargaining?" *International Studies Quarterly* 52, no. 3 (September 2008), 533, https://people.duke.edu/~bl38/articles/warinfoisq2008.pdf; Michael J. Mazarr et al., *What Deters and Why: Exploring Requirements for Effective Deterrence of Interstate Aggression* (Santa Monica, CA: RAND, 2018), 88–89, https://www.rand.org/pubs/research_reports/RR2451.html.

54 Elsa B. Kania, "Artificial Intelligence in Future Chinese Command Decision-Making," in Wright, *AI, China, Russia, and the Global Order*, 141–143.

55 Academy of Military Science Military Strategy Studies Department, *Science of Strategy (2013 ed.)* (Beijing: Military Science Press, December 2013), trans. and pub. Maxwell Air Force Base, China Aerospace Studies Institute, February 2021, https://www.airuniversity.af.edu/CASI/Display/Article/2485204/plas-science-of-military-strategy-2013/.

56 Brent M. Eastwood, "A Smarter Battlefield? PLA Concepts for 'Intelligent Operations' Begin to Take Shape," *China Brief* 19, no. 4 (February 15, 2019), https://jamestown.org/program/a-smarter-battlefield-pla-concepts-for-intelligent-operations-begin-to-take-shape/.

57 Elsa B. Kania, "Swarms at War: Chinese Advances in Swarm Intelligence," *China Brief* 17, no. 9 (July 6, 2017), 13, https://jamestown.org/program/swarms-war-chinese-advances-swarm-intelligence/.

58 Eastwood, "A Smarter Battlefield?" 3.

59 Kania, *Chinese Military Innovation in Artificial Intelligence*; and Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power," 47.

60 Kania, "Artificial Intelligence in Future Chinese Command Decision-Making," 144.

61 Brian David Ray, Jeanne F. Forgey, and Benjamin N. Mathias, "Harnessing Artificial Intelligence and Autonomous Systems Across the Seven Joint Functions," *Joint Force Quarterly* 96 (1st Quarter 2020), 115–128, https://ndupress.ndu.edu/Portals/68/Documents/jfq/jfq-96/JFQ-96_115-128_Ray-Forgey-Mathias.pdf?ver=2020-02-07-150502-693.

62 Kania, *"AI Weapons" in China's Military Innovation*, 6; Kania, "Artificial Intelligence in Future Chinese Command Decision-Making," 146.

63 Wong et al., *Deterrence in the Age of Thinking Machines*, 63.