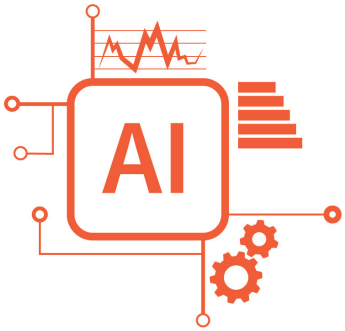

5 | Tabletop Exercise Takeaways

An analysis of strategic SA capabilities according to the attributes and risk factors they could introduce in a crisis suggests some of the ways these technologies could pose escalatory risk, complicate decisionmaking, and challenge traditional notions of information dominance in the strategic SA ecosystem. And yet, real-world case studies or other experiential sources of information to evaluate these assessments are highly limited or overly dated. To evaluate some of the risk assessments identified in research and explore the decisionmaking process of policymakers and technical experts in the throes of crises under a nuclear shadow, the Project on Nuclear Issues (PONI) developed and conducted a series of tabletop exercises on two fictitious regional scenarios. These exercises provided insight regarding both the decisionmaking calculus involved in deploying emerging SA technologies and how their use could potentially impact strategic stability.

Conducted eight times over the last year, with nearly 150 people overall, the tabletop exercises involved a wide range of participants, from senior policy experts with significant government decisionmaking experience to several next-generation nuclear scholars, researchers, and operators. The scenarios sought to inform the policy implications of the theoretical analysis, understand how sensitive U.S. decisionmakers might be to the risks associated with these technologies, and draw conclusions on potential ways to improve crisis decisionmaking and escalation management. The tabletops were not designed to emphasize highly uniform and consistent variables and generate replicable, quantifiable data results but rather to inform a discussion and serve as a learning experience for both participants and observers. What this series of tabletop exercises offers is not concrete facts or indisputable knowledge but a deeper understanding of the human aspect of decisionmaking in nuclear crises.[1] This process provided unique insights irretrievable through traditional academic approaches, raised awareness about strategic SA risk and complexity among both technical and policy participants, and highlighted areas where extant high levels of escalatory anxiety may complicate and even increase escalatory risk—a set of outcomes not fully anticipated in the research phase.

Using two different scenarios across eight different exercises, the study team examined the variation in potential decisionmaker reactions according to the level of intensity of the crisis and different military capabilities, both conventional and nuclear.

Using two different scenarios across eight different exercises, the study team examined the variation in potential decisionmaker

reactions according to the level of intensity of the crisis and different military capabilities, both conventional and nuclear. The China scenario represented a potential "near-peer" in a comparatively early crisis, and the North Korea scenario represented a far more asymmetrical adversary in a more advanced crisis where the initial stages of military conflict are already underway. In both cases, the scenarios took place approximately five years in the future under geopolitical circumstances roughly similar to the present. The SA capabilities discussed and evaluated were all deemed to be technically feasible in the five-year time frame and operationally available for the purposes of the exercise.

Figure 5.1 List of Tabletop Exercises Conducted

LOCATION	DATE	SCENARIO
1. University of California, Berkeley	*February 4, 2019*	*China*
2. Lawrence Livermore National Laboratory	*February 6, 2019*	*China*
3. National Defense University	*April 17, 2019*	*China*
4. CSIS Nuclear Scholars	*June 26, 2019*	*China*
4. CSIS Nuclear Scholars	*June 27, 2019*	*North Korea*
6. Kings Bay Naval Base	*October 17 ,2019*	*China*
6. Kings Bay Naval Base	*October 17 ,2019*	*North Korea*
8. CSIS Senior Experts	*October 28, 2019*	*North Korea*

The first scenario, "Blind Spot," presented a political crisis in the Taiwan Straits precipitated by Chinese escalation in the region and focused on competition between near-peer adversaries. The scenario takes place in 2024 at a time of increasing Chinese pressure to assert regional dominance primarily through economic and grey zone tactics, with reunification with Taiwan an increasing priority. Following a close-approach incident between the Chinese and Taiwanese navies in the Taiwan Strait, China demands the withdrawal of Taiwan's naval assets from the strait, accelerates the timeline for its yearly live-fire exercise, and extends its air defense identification zone (ADIZ) beyond the first island chain while dialing up its rhetoric regarding reunification. Twelve hours before the simulation exercise, U.S. Navy ships near Taiwan have reported significant satellite navigation errors preventing them from conducting regular operations. Several U.S. remote sensing satellites, which provide critical intelligence, surveillance, and reconnaissance (ISR) of Taiwan and the surrounding region, are no longer providing imagery. Facing a growing regional outcry, participants in the exercise are given presidential guidance and objectives to shape their crisis decisions, such as protecting U.S. forces and vital interests in the region, limiting China's expanding influence, and assuring U.S. allies of its commitment to defend their security while avoiding escalation.

The second scenario, "Risky Business," explores the exacerbation of an inter-Korean crisis on the Korean Peninsula. In this scenario, the U.S.-North Korea relationship has reverted to an uneasy deterrent relationship following the breakdown of denuclearization talks, the return of a conservative coalition government in South Korea, and continued economic decline in North Korea. The crisis unfolds when North Korea attacks Baengnyeong Island following a shipping vessel dispute, takes 50 South Korean marines hostage, and issues a series of demands for economic relief and political accommodation. When immediate demands are not met, North Korean forces cross the demilitarized zone (DMZ) on the far-

Republic of Korea Air Force F-16 Fighting Falcon aircraft pilots prepare to take off during Red Flag-Alaska 15-1 at Eielson Air Force Base, Alaska, Oct. 9, 2014.

DoD photo by Tech. Sgt. Joseph Swafford Jr., U.S. Air Force/Released

east side of the peninsula and establish a position on a ridge 20 kilometers into South Korean territory. Presidential guidance includes insistence on restoration of the status quo ante while preventing North Korea's use of nuclear or other weapons of mass destruction against the United States or its allies and avoiding wide-scale conventional war on the peninsula.

DESIGN AND EXECUTION OF THE EXERCISE

In each exercise, participants were split into two groups—a technology team and a policy team. Representing the technical collection communities of the U.S. Intelligence Community and Department of Defense, the technology groups evaluated the utility of a set of strategic SA options (a "collection plan"), which they then briefed to a group of policy decisionmakers for approval. The technology team was tasked with developing a series of options (capabilities and targets) to improve U.S. SA (a "collection plan") and then present the plan to the policy team.

The policy group represented a high-level group of interagency decisionmakers (a notional Deputies Committee) charged with providing advice to the president and implementing presidential guidance. In some of the tabletop exercises, technical groups met contemporaneously with the policy groups; at other times, in order to reduce the time and administrative burden of the exercise, the technical group met virtually in advance to come up with the proposed collection plan which was then briefed to the policy group during the in-person exercise. The policy group was tasked with evaluating the crisis and associated priorities, interpreting presidential guidance, and approving or disapproving the collection plan following discussion of each of the proposed actions. In addition, the policy group would provide additional guidance and limitations, or "guardrails," designed to limit the escalatory risks they identified with some of the approved options. Ultimately, the policy team was responsible for deciding whether to approve each option in the collection plan developed by the technology team. During the collection plan approval process, the technology team contributed to the discussion and answered questions about the collection options. However, the technology team was not allowed to vote to approve/disapprove specific options. Figure 5.2 offers a top line summary of the types of technologies that were offered to

Figure 5.2 Voting Results Across Exercises

VOTING TABLE

DOMAIN	CAPABILITY	APPROVALS OUT OF TIMES OFFERED, CHINA	APPROVALS OUT OF TIMES OFFERED, NORTH KOREA	GUARDRAILS
SPACE	Small Sat	6 out of 8	6 out of 6	*Deployment must be accompanied with diplomatic message; approved with order of preference for use of smallsats to be firstly to monitor maritime forces, then conventional ground forces, and lastly nuclear forces*
AIR	Manned Stealth Aircraft	2 out of 5	1 out of 3	*Approved only for missions that did not violate Chinese airspace*
	UAV	15 out of 25	16 out of 25	*National territory off-limits; Launch facilities only; Only deployed in allied airspace; safeguard this asset for eventual future use*
				In allied littorals only if sufficient information was exchanged with the allies and if the United States properly signaled to the adversary that the swarm was unarmed
				Approved only for missions that did not violate adversarial airspace
				Approved only for missions that did not violate adversarial airspace
SEA	UUV	9 out of 15	1 out of 3	*Deploy only in the contested areas outside of adversary's territorial waters*
				Only to be deployed at chokepoints located within international waters
	Unmanned Surface Vehicles	3 out of 6	1 out of 3	*Only to international and contested waters*
CYBER	Zero Day Exploit	4 out of 7	3 out of 3	*Cyber must be overt and reversible; purely passive collection amd not offesnive or degradatory; safeguard this asset for eventual future use*
	AI Analysis Application	5 out of 6	3 out of 3	*Operators must have established high confidence in this technology prior to deployment; AI must be tested pre-crisis; Don't share methods with allies, just the results*
LAND/ DIRECT PLACEMENT	Compact multi-sensor devices	1 out of 4	2 out of 2	*Allied SOF insertion; inform ally before deploying*

decisionmakers, and how often they chose to utilize them to close critical information gaps. In addition, this chart includes examples of the types of guardrails/conditions that the policy groups levied for using the capabilities, if approved at all.

Analysis

TECH VERSUS POLICY: TWO ROADS DIVERGED

Technology groups were consistently surprised by policy decisionmaking they believed to be "irrational" or unduly conservative given the state of related technology, its broad acceptance and utility in conventional conflicts, and the value they believed it could provide. Technology groups consistently underestimated the level of caution that policymakers might bring to a crisis between nuclear-armed adversaries.

By contrast, policymakers were highly attuned to the escalatory risk associated with intrusive technologies, often weighing their concerns about the potential provocation risks to be more important than the SA benefit that capabilities may provide. Even when such capabilities were approved, policymakers routinely placed guardrails—geographic, target-based, or other—to limit the use of intrusive technologies. Generally, policy participants were so concerned about using any collection options that seemed to be intrusive that they were reluctant to intrude on sovereign territory, waters, or airspace. Such caution was evident even in the North Korea scenario, during which the crisis was presented as severe, the informational benefits potentially significant, and the U.S. asymmetric advantages quite substantial.

INTRUSIVENESS AND SOVEREIGNTY

U.S. policymakers placed high value on internationally recognized borders and Western legal interpretations of "sovereignty." In other words, crossings of internationally recognized "sovereign" borders were interpreted as legally provocative and not just escalatory from a crisis management perspective. When confronted with adversary territorial claims (such as an expanded and enforced ADIZ), policymakers had fewer concerns with placing collection assets in these disputed areas but remained highly cautious and preferred overt modes of collection that could be used for signaling purposes as well as information collection. This remained true even when the adversary in the scenario was engaging in aggressive enforcement of the expanded claim (as in harassing Taiwanese or other ships' aircraft or in the case of North Korean forces establishing de facto control of the island). In these cases, however, policy groups focused on the signaling value of these collection platforms as much, and sometimes more, than their information collection value. Covert or stealthy intrusive capabilities were generally met with skepticism and concern that the risks of escalation by surprise and misunderstanding outweighed the benefits of secrecy.

DOMAIN-BASED PERCEPTION AND MISPERCEPTION

While perceived thresholds associated with sovereignty were highly valued by policymakers, they were not equally valued in all domains. Assets in the air domain were consistently seen as riskier and requiring higher guardrails than those in maritime or cyber domains. Sometimes the use of air-based assets was met with even more skepticism than use of capabilities that required covert emplacement within adversary territory. Some of this caution stemmed from the worry that escalatory risks associated with discovery of an air-based collection capability by the adversary could be provocative but also that the public destruction or shoot down of an air asset could force the United States into an escalatory response. Violating adversary airspace was a noteworthy concern: UAVs were deployed at

surprisingly similar levels across both scenarios, being approved in the China scenario approximately 60 percent of the time and in the North Korea scenario 64 percent of the time. All approvals were conditioned upon extensive use of guardrails to limit the territory in which the assets could be used.

> *While perceived thresholds associated with sovereignty were highly valued by policymakers, they were not equally valued in all domains. Assets in the air domain were consistently seen as riskier and requiring higher guardrails than those in maritime or cyber domains.*

Relatedly, policy players also frequently discounted the value and efficacy of stealth. They accepted it might make it easier to avoid loss but not to avoid detection, and therefore stealth on an air asset generally did not make the asset more likely to be deployed. Policy groups also engaged in robust (and sometimes counterintuitive) debates on the escalatory risks associated with manned versus unmanned aircraft. Technical groups almost always discouraged manned aircraft options for collection, even with advanced stealth, given almost all collection needs could be met with unmanned aircraft at lower operational risk. At times, this disagreement reflected the policy teams' unwillingness to differentiate intelligence collection and signaling, such as when some groups sought to deploy manned aircraft as a signal of determined resolve. In other cases, policy groups sought to raise the escalatory stakes for the adversary while reducing the risk of surprise or misunderstanding as to U.S. intentions by preferring overt and, in some cases, manned aircraft over unmanned and highly vulnerable aircraft like HALE UAVs. Overall, manned aircraft were approved only 40 percent of the time in the China scenario and not at all in the North Korea scenario; these choices were guided almost entirely by policymakers' perceptions of escalation management and signaling rather than informational demands or benefits.

Discussions along these lines became much more pronounced following the Iranian shoot down of a U.S. Global Hawk.[2] The session held after the Global Hawk shoot down involved an extensive discussion of the risk of shoot down of unmanned assets as too easy or appealing for China. That group determined that it was essential to assert U.S. willingness to put manned, non-stealthy assets into the contested area (but not over internationally recognized Chinese territory) before using unmanned assets and that clear deterrence-oriented, declaratory statements are needed regarding the targeting of surveillance assets. This was strongly considered as a means of rejecting Chinese claims of an expanded ADIZ while simultaneously collecting information in the China scenario. In many ways, these decisions may have represented "recency bias" in action, given proximity to the Iranian shootdown. In the North Korea scenario, policy groups remained reluctant to fly unmanned platforms over DPRK territory given the shoot down risk, and only authorized their use over the ROK or international waters or territories. Even in cases where the UAV platforms were approved, approvals were contentious and involved longer debates among participants than other aspects of the collection plan.

TWO IF BY SEA

Sea-based assets, both surface and subsurface, generally receive similar guardrails, but policymakers showed greater willingness both to risk these assets, in terms of discovery and loss, and see them as either more easily hidden (subsurface) or somewhat less provocative. Overall, unmanned underwater

The Navy's most technologically advanced surface ship USS Zumwalt (DDG 1000) steams in formation with USS Independence (LCS 2) and USS Bunker Hill (CG 52) on the final leg of her three-month journey to her new homeport in San Diego.

U.S. Navy Combat Camera photo by Petty Officer 1st Class Ace Rheaume/Released

vehicles (UUVs) were approved 60 percent of the time in the China scenario and one out of three times during the North Korea scenario—ratios roughly similar to the UAV approvals. However, the discrepancy is clear when more detailed options are considered. For example, static UUV nets deployed at key choke points were approved all four of the times offered in the China scenario and three of four times in the North Korea scenario. On the other hand, the more intrusive autonomous UUVs with advanced sensors that would provide far more actionable information were approved only 33 percent of the time in both scenarios.

During a discussion after the China exercise, one participant suggested that perhaps they had regarded naval assets as less escalatory because the crisis had begun in the naval domain and increased naval surveillance therefore seemed proportional. However, the deployment of aerial assets overall was consistently perceived as riskier than the use of underwater assets. The policy team often argued that underwater assets gave leeway for plausible deniability and the loss of an asset was less likely to prompt a public response or go viral on social media the way a more visible shootdown of an air asset might. For instance, should an adversary sink a U.S. underwater asset, it would be more difficult for an adversary to retrieve that asset, thus protecting U.S. technology from falling into adversary hands and allowing the United States the option to deny involvement. At least implicitly, the comparatively more public and visible nature of targeting and destroying an air asset in ways that could "force the hands" of policymakers seemed to weigh heavily on policy groups in ways that similar capabilities and sensors did not when used in the maritime domain.

While of little signaling value, subsurface capabilities did risk surprising an adversary, which could have difficulty distinguishing between armed and unarmed capabilities. Hence policy groups generally

rejected placing such collection platforms in proximity to sensitive targets. The utility of surface vessels as collection platforms were evaluated largely independent of their informational value; instead, approval decisions largely depended on whether a group weighed positive signaling benefits more than the risk of attack or loss.

In sum, policy groups remained very cautious with any intrusions into an adversary's airspace or territorial waters and in all cases approved these collection capabilities only with clear guardrails denying approval to enter sovereign territory, airspace, or waters and generally adjudicated the use of these platforms according to how they perceived their value in shaping the crisis overall.

SPACE AND CYBERSPACE

Even supplemental space assets raised interesting domain and sovereignty questions. What constitutes sovereign airspace? What about capabilities such as pseudosatellites or smallsats that are deployed from aircraft or exist in the region between outer space and airspace?[3] Smallsats represented a consistent point of divergence between technology and policy groups, particularly in the China scenario, which involved Chinese dazzling of U.S. naval navigational assets as part of the initial crisis. Tech groups consistently recommended the deployment of smallsats as providing targeted coverage and vital redundancy with relative safety. Policy groups were far more skeptical and sometimes dismissive, questioning the additional value-added to existing space systems, fearing additional targeting and disablement of vulnerable systems and expressing concern about how the launching and deployment of the constellation would be seen and perceived by the adversary during the crisis. Concerns were often assuaged with a back and forth between policy and tech teams. Thus, despite trepidation, co-orbital reconnaissance small satellites were approved in the China scenario 50 percent of the time, and smallsat constellations were approved 75 percent of the time. Policy teams playing the North Korea scenario approved these capabilities every single time they were offered.

Pseudosatellites, which are multi-payload, high-altitude air vehicles or airships able to maintain a fixed position over a single area of interest for extended periods of time, were initially met with skepticism from the policy groups. While tech groups saw the platform as providing persistent surveillance with impressive sensor capacity at safer distances, policy groups focused on the challenges of the high visibility deployment, concerns about intrusions into national airspace even at very high altitude, and vulnerability to attack and shootdown, among other concerns. After dialogue between the groups, voting patterns demonstrate greater trust in the capabilities—pseudosatellites were approved three out of three times offered in the China scenario and two out of three times for North Korea.

Cyberspace is one area where groups tended to diverge, with some participants treating cyberspace as highly intrusive and escalatory. Such participants were particularly concerned with any action that appeared to target adversary C2 and decisionmaking, typically out of fear that such action could escalate the crisis. In the China scenario, cyber espionage was only approved 57 percent of the time, as participants expressed wariness of inciting aggression. Others felt that it was less escalatory ("states do cyber intrusions all the time and it doesn't start wars"). Cyber espionage was approved every time it was offered on a North Korean collection plan. Several participants voiced the perspective that aggressive moves made by North Korea indicated a resumption of hostilities, and much of the discussion around capabilities focused not on if they should be deployed but rather when during the unfolding crisis they would be most effective. Accordingly, even though cyber espionage capabilities were approved for all missions, policy participants voiced concerns that the zero-day

> *Policymakers routinely expressed concerns about anything that appeared to target C2 assets, especially in the cyber domain.*

vulnerabilities were so valuable that it may be prudent to hold them for when they would have the most impact in the event of a military conflict (e.g., targeting North Korean leadership, tracking nuclear weapon deployment). In cyber-related options, discussion turned more to targets than to domain as areas of concern or potential constraints, but isolating targets in ways that would be demonstrable or transparent (and therefore presumably less escalatory) was very difficult.

Policymakers routinely expressed concerns about anything that appeared to target C2 assets, especially in the cyber domain. Groups could not articulate effective ways to differentiate between nuclear and conventional C2 assets (even just for information collection, not degradation) and tended to disapprove of these options even when critical gaps on adversary decisionmaking significantly impeded crisis management. This tended to lead to very expansive guardrails and often included that any cyber actions taken to degrade an adversary's SA must be reversible and overt to prevent misinterpretation of the purpose of the attack.

"NICE TO HAVES" VERSUS "GOTTA HAVES"

Policy teams expressed frustration with the inability of technical collectors to clearly articulate detailed value propositions associated with each collection capability. They posed questions such as, what information will I gain from capability X that I cannot get from a less risky option like Y? What will it cost? What are the trade-offs? Some of the questioning betrayed the bounds of scenario-based discussions or exercises in which the policy group had to make decisions based on the limited information available, but the interrogative nature of the exchange and repeated requests for more "homework" appeared to replicate potential real-world crises in which decisionmakers seek higher confidence information at lower levels of risk and fear slippery slopes and unintended consequences that could lurk behind information collection choices they do not fully understand.

> *Generally, policy groups viewed new and unfamiliar technical capabilities with higher levels of mistrust and with keen attention to perceived escalation risks.*

Generally, policy groups viewed new and unfamiliar technical capabilities with higher levels of mistrust and with keen attention to perceived escalation risks. Policy teams often epitomized generational and experiential gaps compared to tech collectors and hence a slightly lower "technology IQ" that manifested in higher concerns about the utility and risks of these capabilities. Due to their different knowledge base, they demonstrated a subsequent lesser comfort level with deploying emerging technologies. Policy groups also tended to assume a higher likelihood of technology failure (worst-case scenario decisionmaking), while technical groups generally held high confidence in the capability to perform as intended and approached information collection from an optimization perspective.

"BUT WHAT DOES THIS SIGNAL?"

Juxtaposed with policy teams, tech group participants were largely indifferent to the signaling that accompanies the deployment of certain technologies. In stark contrast, signaling was often a primary subject of discussion for policy teams, whose comments often underscored the perceived

Juxtaposed with policy teams, tech group participants were largely indifferent to the signaling that accompanies the deployment of certain technologies.

inextricability between collection and signaling. Policy teams recognized that SA capabilities are primarily for information collection but clarified that if they were to be employed in a signaling capacity, this would have to be made clear in order to prevent inadvertent escalation. That is why caveats were often added to approved capabilities. For example, even when smallsats were overwhelmingly approved, policymakers advocated that their deployment be accompanied by a diplomatic message.

Not only were policy participants concerned with the signals that their actions conveyed to the adversary and allies, but they also repeatedly attempted to decipher the signal that an adversary was relating to them. For example, what was China signaling when it spoofed and jammed U.S. SA capabilities? For some individuals and policy groups that believed in a more assertive military posture, SA capabilities were to be deployed to signal resolve. This was perhaps most evident when participants played the North Korea scenario: some participants expressed the belief that recent developments already signaled the resumption of military hostilities, with some even going so far as to consider whether these SA capabilities should be offensive as opposed to merely intended for increasing SA (for example, weaponizing cyber espionage to introduce malicious code into North Korean networks).

Related to the topic of signaling, concerns about North Korea's nuclear capabilities were acute, but participants showed less regard for establishing guardrails around North Korea's nuclear command, control, and communication (NC3) systems. While there was discussion around the sensitivity of these systems, as well as monitoring launch sites and nuclear warhead facilities, several participants diminished the threat of miscalculation or misinterpretation. One participant argued it would be "deeply irresponsible" to avoid gathering as much information as possible about North Korean nuclear capabilities considering the risks, and several others commented that given their provocative actions, North Korean leadership likely already assumed that the United States would be targeting their NC3 systems. Since the North Korea scenario reflected a fairly advanced crisis with a much more inferior adversary, policy groups seemed willing to take higher levels of risk, with one participant explaining their logic as a concern not over the riskiness about any one capability but rather the usefulness of the information currently. That said, these discussions were among the most contentious and the approval decisions were far from unanimous.

Endnotes

1 Ed McGrady, "Getting the Story Right About Wargaming," War on the Rocks, November 8, 2019, https://warontherocks.com/2019/11/getting-the-story-right-about-wargaming/.

2 Tara Law, "Iran Shot Down a $176 Million U.S. Drone. Here's What to Know About the RQ-4 Global Hawk," *Time*, June 21, 2019, https://time.com/5611222/rq-4-global-hawk-iran-shot-down/.

3 On pseudosatellites, see Pedro Vicente Valde and Paulina Wheeler, "High Altitude Pseudosatellites," On the Radar, CSIS, July 29, 2019, https://ontheradar.csis.org/issue-briefs/high-altitude-pseudosatellites/; and on smallsats, see Goldblum and Reddie, "Smallsats."