SOCIAL MEDIA STORMS AND NUCLEAR EARLY WARNING SYSTEMS

A DEEP DIVE AND SPEED SCENARIOS WORKSHOP

Technology for Global Security

Preventive Defense Project, Stanford University

Nautilus Institute

N2 Collaborative

January 8, 2019

Authorship: This report was prepared by staff of Nautilus Institute, Preventive Defense Project Stanford University, and Technology for Global Security with assistance from N Square Collaborative.

Acknowledgments: The workshop was funded by the John D. and Catherine T. MacArthur Foundation "X-grant" program. The Hewlett Foundation kindly provided use of its conference room for the workshop. The workshop was held on October 10, 2018 under the Chatham House Rule.

Publication: This report is published simultaneously by Technology for Global Security here and by Nautilus Institute here.

TABLE OF CONTENTS

I. EXECUTIVE SUMMARY AND INTRODUCTION

I.1 SOCIAL MEDIA EMERGES AS A POSSIBLE TRIGGER OF NUCLEAR EARLY WARNING SYSTEMS

This workshop was convened because social media burst onto the stage of nuclear warfare in 2018. In the Asia-Pacific region alone, six instances of social media playing a role in nuclear-prone conflicts occurred between August 2017 and January 2018. For decades, strategists have worried about the possibility that states armed with nuclear weapons might mistakenly launch a nuclear strike due to a false alarm originating in its early warning system or due to degraded decision-making.

Why does this matter? The primary reason is the terrifying combination of speed and unique scale of violence when it comes to nuclear weapons that continues to set them apart from all other means of coercion. The simple problem is that nuclear commanders must make decisions to use nuclear weapons for mass destruction in time measured in minutes and seconds, not hours and days.

How should a nuclear weapons state treat the vast amounts of social media, including content that may be factually accurate that is transmitted almost instantly, but may also be created by malevolent parties that aim to deceive, manipulate, and mislead its early warning system and pollute its nuclear command decision-making process with fake information?

In addition to "traditional" sources of NC3 error such as accidents, hardware failure, and human error, new technologies such as including artificial intelligence, autonomous vehicles, quantum computing and sensing superimposed on the legacy US NC3 system may create new types of coincident error involving social media.

In the past, the insulation of nuclear commanders from unclassified data and their near total dependence on official, classified information systems in the midst of crisis might have served to reduce the influence of erroneous or deceptive information. In today's world where the President of the United States makes public declaratory policy on Twitter, one cannot be sure that social media will not be influential.

I.2 LESSONS FROM SOCIAL MEDIA IN NON-NUCLEAR DOMAINS

To investigate how social media might play out in the world of nuclear early warning the workshop considered the use of social media to promote extremist views and behavior in promoting anti-vaccination, anti-Semitism, gang, ethnic, and terrorist violence in cities. From this evidence, two sets of lessons were drawn. The first focused on the issue of false data and false alarms leading to conflict as well as conflict escalation. The second lesson primarily focused on what antidotes exist for false alarms on social media or via other ways of creating an authoritative and credible reference knowledge.

Citing detailed case studies and data, presenters showed that social media is proving capable of quickly mobilizing fear and channeling aggregated animosity against real-world targets. Therefore, in these domains there are precursors of how activists might attempt to mobilize mass sentiment online to generate fear and alarm in the face of a nuclear threat. Additionally, the targeting of nuclear commanders or key individuals in a nuclear command and control system may occur in an effort to launch a nuclear attack against a third party presented as worthy only of nuclear annihilation.

The conclusion is unavoidable that individuals, organizations, and even states may start to use social media to try to provoke nuclear attacks against their adversaries; or for other political-ideological or religious reasons; that they will be effective in terms of reaching some highly influential people as well as large numbers of people; and included in these two types of readers are likely to be some people making nuclear early warning assessments, and nuclear command decisions.

A variety of strategies were described that partly ameliorated the problem, although generally the response by social media platforms themselves was slow, often only at the behest of outsiders, and inhibited by many problems of attribution and tracking of online aggressors.

Where social media platforms have attempted to curtail manipulative and dangerous use of their services, they have found that one of the best ways to do so is to simply slow down the ability of users to run their campaigns. This can be achieved by automated systems that shut down sites found to be fake sites and by identifying core behaviors used by social media activists that become markers that can be used to create traction and slow down their ability to game the rules of social media by using bots, anonymity, etc. One problem in such regulation of online behavior is that context is critical to determining what is and is not dangerous, and a human must be in the loop. However, once deep fakes become widespread, even having a human involved may not suffice to determine the truth content of a specific post or site.

Given the speed of nuclear early warning and nuclear command decisions, following these pathways is likely too little, too late. Indeed, as one social media practitioner said, “If we find something [really bad] in the nuclear community, then who do we really tell is the question!” One possibility is that first responders, especially at the city level, might use their information systems to provide credible information to reporters and mass media; and to fold carefully evaluated social media reports into their information after first validating the early reports with a variety of real-time sensors. First responders noted that it is essential to ensure that the messages across cities and counties are aligned and consistent, to avoid public confusion and disenchantment with official sources of early warning. As one first responder noted, “You need to have a battery of circuit breakers to preempt and steer false information away - one message is not going to solve it.”

Thus, even in the case of threats with extremely short timelines—minutes in the case of bombing, seconds in the case of earthquakes, it is possible to use social media to inform and guide humans to respond in ways that reduce risk, without the risk of the alert system being

hijacked by malevolent actors. The critical element is that the users trust the reliability and authenticity of the information sent out.

I.3 CIRCUIT BREAKERS & SHORT CIRCUITS SPEED SCENARIOS

Participants developed four "short circuit" scenarios that explored how and what circuit breakers might be created that avoid or overcome the destabilizing effect of social media on nuclear early warning systems and nuclear command decisions.

The four scenarios and circuit breakers include:

Korea A Sweltering Crisis—The United States and North: Multiple circuit breakers were prefigured including launching fact-finding missions, improving official communications with mass media, and leveraging trusted third parties willing to put themselves on the line as exchange-hostages to push the international community to a peaceful solution, reestablish trust, and mitigate panic.

Fast and Furious—India, Pakistan, and a Non-State Actor: The circuit breaker in this firestorm scenario aimed to create the necessary political space to deescalate the situation, starting with mitigation strategies set in motion by social media luminaries and companies to slow down communication combined with inter-state backchannel conversations and more traditional mutual hostage taking to stop conflict from spiraling out of control in conditions of extreme nuclear provocation.

Mutual Miscalculations: NATO, United States, and the Russian Federation: The circuit breaker set out to answer the question: Are there ICBMs in the air and if so from where? While dealing with increasing panic. The circuit breaker is to create a new international organization with only one mission--to provide sensor-based information on international incidents that may related to nuclear war, to validate data events; and impartially to send data to all parties to a nuclear prone conflict.

Embrace Tiger, Retreat to Mainland—China, Taiwan, and the United States: Similar to the last scenario, the key question in this scenario is: What's the nature of the Chinese missiles, nuclear or conventional, that are in the air; and will they hit Guam and Okinawa or splash down nearby? In this case, China initiates a last minute, last second concerted, all-out *diplomatic* effort to stop a US retaliatory strike by calling on trusted persons to act as intermediaries. They activate personal backchannels including businessmen, politicians, and even religious leaders. Their first act (because it is the fastest) is to turn off aggressive social media in China itself.

I.4 CONCLUSIONS

Some themes recurred across all the scenarios, suggesting that these elements might lend robustness to many de-escalatory strategies. For example, whatever the role of social media in

in creating or amplifying nuclear-prone conflicts, high-level communication between trusted individuals was an ingredient in resolving conflict. Other elements included high and low-level hostage exchange; doing whatever it took to slow escalatory spirals; and anticipating the loss of control induced by social media and other drivers by establishing hot lines, market and civil society-based communication channels, and trusted, third party, and impartial sources of authoritative information on the status of forces.

The workshop participants were convinced that social media platforms and social media users can shift the center of gravity away from the current, celebrity-driven and conflict-amplifying social media conflict amplifying dynamic that degrades the quality of much information and towards more reliable, authenticated information while preserving the ability of users to free speech and near-instantaneous networking. In this regard, cities and civil society emerged as a set of actors and networks that may be positioned to create new forms of governance and public information goods that restrains the aggressive use of social media that may contribute to false alarms and poor decision-making at the national level, while contributing to independent, impartial and validated information that is useful to nuclear early warning systems and nuclear commanders who may be relatively poorly served by traditional sensors, early warning systems, and conflict resolution mechanisms at the level of inter-state conflict.

II. SOCIAL MEDIA STORMS AND NUCLEAR EARLY WARNING SYSTEMS—OPERATIONAL CONTEXT

Social media burst onto the stage of nuclear warfare in 2018. In the Asia-Pacific region alone, six instances of social media playing a role in nuclear-prone conflicts occurred between August 2017 and January 2018. Three (September, November, December 2017) related to indicators that the United States might be readying to attack North Korea with nuclear weapons.

TABLE 1: 20170-18 FALSE ALARMS AND SOCIAL MEDIA STORMS ABOUT NUCLEAR AND MISSILE ATTACKS

GUAM: E. Adamczyk, "Guam residents unnerved by accidental 'civil danger warning,' UPI, August 15, 2017.

SEOUL: D. Lamothe, "U.S. families got fake orders to leave South Korea. Now counterintelligence is involved," *Washington Post*, September 22, 2017.

"The U.S.S. Kentucky is part of what is called the nuclear triad." The triad are the three components of a nuclear defense system: "land-based missiles fired from secret silos, B-1 bombers that can drop them from the air, and submarine-launched ballistic missiles." STRATCOM tweet, C. Perez, "Military tweets out error-filled story about US nukes," *New York Post*, November 15, 2017.

SOUTHERN CA: 'Update: Minuteman III ICBM test launch from Vandenberg canceled," December 6, 2017 Postponed" Minuteman missile launch supplanted by unannounced Trident missile launch same day
(April 7 2013 launch postponed to avoid provoking DPRK.)

HAWAII: A. Wang, B. Lyte, 'Ballistic Missile Threat Inbound To Hawaii,' the alert screamed. It was a false alarm," *Washington Post,* January 13, 2018 .

TOKYO: AP, "Japan public TV sends mistaken North Korean missile alert," January 16, 2018.

Three (August 2017 and January 2018) led to social media storms that amplified false alarms of pending nuclear attack involving millions of people hiding under tables and waiting for their world to end.

For decades, strategists have worried about the possibility that states armed with nuclear weapons might mistakenly launch a nuclear strike due to a false alarm originating in its early warning system or due to degraded decision-making. (The flip side of that concern is that a nuclear weapons state might not notice that it is under nuclear attack because of errors in its early warning system and might not respond appropriately due to degraded nuclear decision-making).

Nine states now have nuclear weapons. Each of them has a nuclear command and control system to maintain their nuclear forces, and to operate them in peacetime. Those who command these forces rely on information about the status of their potential nuclear adversaries. This information is obtained from many sources culled together in strategic intelligence; and on sensors and other sources of real-time or immediately available information that monitor the status of potential attacking that considered together, suggest that nuclear weapons are—or are not—about to attack a given nuclear weapons state.

Thus, each nuclear weapons state maintains an early warning system that evaluates threat data, and if it receives information that might suggest an attack is underway, assesses the significance of the threat. Some states use physical sensors such as infrared detectors on satellites and long-range radars to provide "dual" warning from physically distinct and separate systems; and rely on one to cross-check and confirm the readings from the other. In this regard, the United States has the most mature early warning and nuclear decision-making system, supported by a global network of sensors linked by communication systems to early warning assessment centers that in turn report to nuclear commanders. (See Figure 1).

Other nuclear weapons states have much less capable early warning systems, using a few satellites with limited coverage, supplemented by a few other long-distance sensors such as radars. Some have no long-range sensor systems at all, such as North Korea.

Yet all these states have social media available to their officials, even in North Korea. Thus, the first warning that nuclear capable missiles or bombers might be heading for Pyongyang might be on Twitter or Facebook. In addition to providing possible early warning of the physical status of nuclear weapons, social media may also provide unprecedented and unique access to the intentions—and the state of mind—of a nuclear commander.

Even more worrisome, much of the information found on social media is factually incorrect; and some of it is purposely posted to manipulate users as "fake media" or to fan the flames of conflict by manipulating readers *en masse*. Even worse, it is becoming difficult and even impossible to distinguish between actual videos, photographs, and voices and "deep fakes."

Why does this matter? The primary reason is the terrifying combination of speed and the unique scale of violence when it comes to nuclear weapons that continues to set them apart from all other means of coercion. The simple problem is that nuclear commanders must make decisions to use nuclear weapons for mass destruction in time measured in minutes and seconds, not hours, days and minutes. This is due to the compression of decision-making time by the deployment of long-range delivery systems that can take as little as 10-12 minutes to arrive from firing point, as shown in Table 2. (Here, the decision-time is shown for POTUS or the President of the United States; similar or worse timelines confront all nuclear weapons commanders in all nuclear weapons states).

FIGURE 1: US NC3I SYSTEM

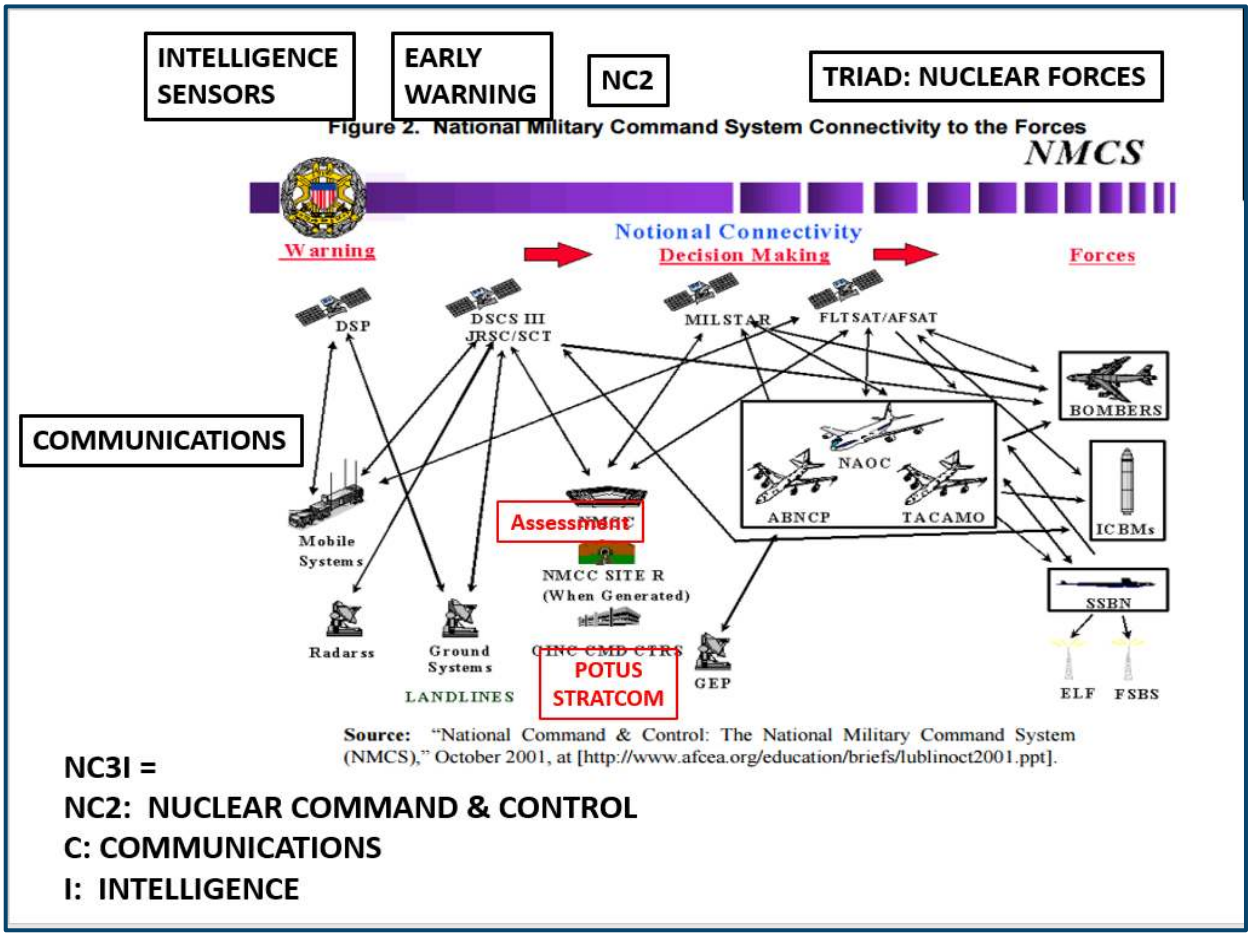

TABLE 2: MINUTES TO RESPOND TO LONG-RANGE NUCLEAR MISSILE ATTACK

		Low	Cumulative	High	Cumulative
Satellites Sense Missile Infrared Signature		1	1	1	0
Space object/OTH radar sight payloads*		2	3	2	1
Missile Assessment Conference*		3	6	8	3
Alert to Commanders		1	7	1	11
POTUS briefing		0.5	7.5	2	12
POTUS decision time		6	13.5	12	14
Silo missile forces to receive authenticate fire order		2	15.5	4	26
Silo missile launch procedures		1	16.5	3	30
Silo missile liftoff to safe distance		1	17.5	1	33
Total Minutes		**17.5**		**34**	
Minutes early/late relative to ICBM NUDET time	30	**12.5**	**mins early**	**-4**	**mins late**
Minutes early/late relative to depressed trajectory SLBM	15	**-2.5**	**mins late**	**-19**	**mins late**
Minutes early/late relative to IRBM NUDET time	10	**-7.5**	**mins late**	**-24**	**mins late**

Notes:

* two independent information sources using different physical principles, such as radar and infrared satellite sensors associated with the same event

** Actual time to assess in two serious alerts was 8 minutes

Depending on the nature of the attack (long-range land-based missiles or offshore submarine or ground-launched intermediate range missiles fired at the United States or its allies, the first nuclear explosions could occur in as little as 10-15 minutes, and as long as 30 minutes. But the time the liftoff is noticed by satellites, cross-checked by radar, assessed, alerted to the US national military command, the President briefed and decisions made, at minimum, 13 or more minutes and more likely, as much as a quarter of an hour will have elapsed. Given the additional time to order and launch a responding strike (four to 8 minutes, at minimum), the US president will be under tremendous pressure to fire or risk losing most of his force if the incoming attack is massive; or making an enormous error of launching a nuclear counter-strike if the warning is for "only" a few incoming missiles of unknown arming, origin, or intention. This use-or-lose imperative is even more extreme in other nuclear-prone conflicts such as India-Pakistan, or for North Korea staring down the US nuclear barrel with no long-range sensors of its own.

What are the operational implications of this trend? How should a nuclear weapons state treat the vast amounts of social media, including content that may be factually accurate that is transmitted almost instantly, but may also be created by malevolent parties that aim to deceive, manipulate, and mislead its early warning system and pollute its nuclear command decision-making process with fake information?

The participants in the Social Media Storms and Nuclear Early Warning Systems workshop set out to answer this question. Social media is a huge phenomenon which affects almost every aspect of globalization today. However, we were not interested in social media in general, but only how it might affect and mislead a nuclear early warning system into declaring a false alarm, or might degrade nuclear decision-making in ways that increase the risk of the mistaken use of nuclear weapons.

At the workshop, we explored the impact of three types of social media: a) celebrity posts that demand attention due to the gravitas or the perceived credibility of the source (for example, a head of state; or a person known to be reputable and impartial from the viewpoint of the reader); b) field reports on the status of nuclear weapons and fire orders (for example, social media reports of missiles on the move, live streaming of launches, deployment of delivery vehicles, or possible detonations); c) false alarms originating in or amplified by social media that may indicate that a state's early warning system is already "primed for war" and jittery. Each of these types of social media in turn may inform the threat sensed by an early warning assessment center in a nuclear weapons state; how such a threat is assessed and the significance attributed to the perceived threat indicators; and how an alert sent to nuclear commanders may be treated by them given that the commanders have independent access to social media posting, and may be susceptible to the effects of social media.

In this early warning-assessment-alerting-decision making sequence, social media posts may provide supplementary data or influence that—when combined with other sensor data and indicators—"tip" the assessment from "no attack underway" to "possible attack" or "attack underway." We posit that social media's influence is always at the margin, supplementing information from strategic intelligence (such as communications or electronic intelligence) and

real-time sensor data, or tactical warning; and that no nuclear commander would ever make a decision solely based on social media posts by an adversary, nor by a friendly social media source that is valued by a nuclear commander. (This is a working hypothesis that may not be sound, but we used it to simplify the issues examined at the workshop).

One of the critical difficulties in tackling these issues is that errors in early warning systems and decision- making take place in contested information environments characterized more by lack of data, ambiguous indicators, mixed signals, and conflicting sensor data. In this information milieu, false signals, are routine and are even expected—not least because sensor systems may not be cross-calibrated to provide cross-checking confirmation. Use of many and redundant sensors may overwhelm the assessment system with so much data that it has to prioritize and sequence assessment that may in turn clash with the decision-making time available, as noted above—and more sensors may create more technical errors in the first place, depending on how many independent confirmations are required to assess that a threat is imminent and significant. Complete lack of data—or sudden sensor silence or unexplained disappearance—is itself significant, because such conditions may denote early attack and disablement of the sensor or communication systems, or an insider attack on the data fusion and display system; or the system may simply have failed for purely technical reasons that coincide with a period of high tension and alertness. In short, situational awareness is always imperfect, and even in the United States, there are limits on how much a command and control system can rest on strategic and tactical intelligence to provide reliable support to nuclear commanders.

During the Cold War, "Missile Display Conferences to Evaluate Possible Threats" (MDCs) were called as soon as a possible launch was detected, or unusual information appeared from warning sensors. 1,152 moderately serious false alarms occurred during the period 1977 to 1984, an average of almost three false alarms per week. (Current data is not available, but for reasons adduced below, it is likely to have increased over the decades). It is possible for two such alarms to happen simultaneously and take time to resolve. More time may be needed if the systems reporting them are independent (for example, space and radar), resulting in sequencing and loss of decision-making time while the reports are resolved—more time than may be available for decision-making in some circumstances, thereby neutralizing “dual phenomenology.”

Today, the US early warning system not only assesses missile launches all over the world almost every day due to the intensity of wars involving missiles, but also other natural and technological events that may appear similar to missile launches (sun glinting off clouds, technological accidents and explosions) or to incoming re-entry vehicles (space objects regularly falling out of orbit). Thus, early warning personnel have to discriminate possible nuclear strike signals from background noise that is constant and increasing—a hubbub to which social media is now added. Although the role it plays in threat assessment is unknown, Twitter feeds are streamed onto the monitors at STRATCOM’s battle-deck alongside imagery of missile trajectories and other information sources.

As noted above, we consider the marginal role of social media not because we expect it to be the only or critical source of error in nuclear early warning or decision-making, but because it

may influence assessments at critical moments when other factors are driving the system to an erroneous assessment and false alarm that a nuclear attack is underway. Historically, these other factors include technical failure of computer chips and communication hardware in shared systems that eliminated dual-sensor system cross-checking of apparent attacks, procedural error such as the use of exercise tapes that displayed missile attacks leading to alerting of national commanders that the United States was under attack, the coincidence of procedural error such as mistaken use of exercise tapes in a radar leading to a perceived missile threat from Cuba that coincided with a real sensor reading of an actual Soviet satellite passing overhead at the height of the Cuban Missile Crisis, and to outright organizational pathology, such as "Operation Doom 99," the unauthorized transfer of six nuclear warheads across the United States in August 2007.

As new technologies such as including artificial intelligence, autonomous vehicles, quantum computing and sensing are introduced and superimposed on the legacy US NC3 system, new types of coincident error involving social media input seem more likely than not in at least one if not more of the national NC3 systems that exist today.

Finally, humans under stress of time and high stakes introduce all sorts of perceptual, psychological, and institutional biases, distortions, and errors into all information processing at all steps in the system, including fully automated states (that incorporate the biases of their designers). In this messy process, therefore, social media may seem less unreasonable and more reliable as a source of early warning when compared with other strategic intelligence or tactical warning available to assessors and commanders—especially if it reinforces other erroneous data or interpretations by the early warning system, or if it reinforces the prejudices of nuclear commanders in ways that incline them to early response.

In the past, the insulation of nuclear commanders from unclassified data and their near total dependence on official, classified information systems in the midst of crisis might have served to reduce the influence of erroneous or deceptive information. In today's world where the President of the United States makes public declaratory policy on Twitter, one cannot be sure that social media will not be influential.

III. LESSONS FROM SOCIAL MEDIA IN NON-NUCLEAR DOMAINS

The workshop considered the use of social media to promote extremist views and behavior to ascertain if lessons might be learned for how social media might play out in the world of nuclear early warning.

Presented papers and the discussion included the role of social media in promoting anti-vaccination, anti-Semitism on the one hand, and gang, ethnic and terrorist violence in cities on the other.

The first set of lessons concerned the question of false alarms generated by social media, and the resultant creation and amplification of conflict where little or none existed before.

In the case of the "anti-vaccers" or social media campaigners who aim to stop vaccination, a case study showed that a virtual social network is vulnerable to cross-platform manipulation that develops a large standing audience for the anti-vaccination perspective. The anti-vaccers used many sophisticated techniques to drive vulnerable readers away from vaccination such as the use of gameable algorithms. The published paper by Renee Diresta argues that in the anti-vaccination case, a confluence of three factors - mass consolidation of audiences onto a handful of social networks; the adoption of curatorial algorithms as a primary means of disseminating and engaging with content; and the ease of precision targeting of users via the leveraging of proprietary profiles built from their own media consumption signals - has resulted in an information ecosystem that can be manipulated by a variety of actors with relative ease.

Concludes Diresta:

The anti-vaccine movement is well-funded and technically savvy. They followed the best practices of internet marketers, writing blogs and cross-promoting content and sharing material across all of the new platforms. Social network design choices meant that popularity determined what people saw; even nuanced policy issues began to be run as digital marketing campaigns.

In effect, the anti-vaccers used social media on a massive scale to short circuit the traditional flows of medical knowledge and expert, evidence-based advice available to individuals (in particular, parents) to reduce the rate of vaccination, in some cases, to the point that public health was threatened by revived outbreaks of contagious diseases.

In the case of anti-Semitic online activism, a case study by Brittan Heller found that the trolling and attacks on minority journalists, especially of Jewish ethnicity, used social media-based aggression to spark violence, including off-line violence. This study found that the online attacks were highly targeted—83 percent of 2.6 million anti-Semitic tweets, for example, were targeted at only 10 people; and that the attackers were professionalized, Moreover, stated Heller:

Overall, the study found that a comparatively small group of attackers drove most of the anti-Semitic hate and harassment on Twitter, but these individuals had an outsized impact. More than two-thirds of the anti-Semitic tweets directed at journalists were sent by 1,631 Twitter accounts, out of 313 million total Twitter accounts at the time of the attack. While this is a small proportion of Twitter users, the comparative impact of this abuse was widespread. The reach was tantamount to the spread covered by a $20 million-dollar Superbowl ad.

As with the anti-vaccers, a relatively tiny number of social media activists were able to manipulate readers to change their views and their behavior in the real world, by not vaccinating or by attacking target persons not only virtually but in the real world. Other examples from other domains were cited at the workshop along the same lines.

In cities, social media has been employed to motivate and then to orchestrate mob violence against minority populations, and even to coordinate large-scale terrorist attacks (as in the case

of Mumbai in 2008. Cities are natural targets for such manipulative campaigns because they aggregate so many people who can be networked into flash mobs and riotous behavior that traditional policing has no answer short of total shutdown of civilian communications such as occurred in London in 2012. Thus, in his presentation to the workshop, Sunil Dubey concluded that:

By 2030, over 65% of total world population will live in cities. Cities confront the rising influence and penetration of social media platforms on all aspects of urban life. Although this virtual urban life makes cities smarter, more efficient, and more sustainable in many respects, it also subverts the safety, security and resilience of our cities."

In discussion, a study was cited that found that 80 percent of gang violence in Chicago starts or is facilitated online. Thus, time and again, social media is proving capable of quickly mobilizing fear and channeling that aggregated animosity against real-world targets. In these domains, therefore, there are precursors of how activists might attempt to mobilize mass online to generate fear and alarm in the face of nuclear threat; and to target key individuals either in a nuclear command and control system, or nuclear commanders themselves, to launch a nuclear attack against a third party presented as worthy only of nuclear annihilation. Such campaigns might also involve the use of fake news propagated over social media to justify the campaign—as occurred on September 11, 2104 in an online hoax involving many fake twitter accounts that posed about the attack and targeted celebrities in order to maximize the attention. Not only did the hoaxers present edited CNN screenshots; they even posted functioning clones of TV stations that purported to cover the event.

This creation of a virtual attack is similar in nature to the fake post of a non-existent non-combatant evacuation in Korea in September 2017. It suggests that it is almost inevitable that individuals, organizations, and even states may start to use social media to try to provoke nuclear attacks against their adversaries; or for other political-ideological or religious reasons; that they will be effective in terms of reaching some highly influential people as well as large numbers of people; and included in these two types of readers are likely to be some people making nuclear early warning assessments, and nuclear command decisions.

The second set of lessons concerned the question of what antidotes exist for these types of false alarms either on social media, or via other ways of creating authoritative and credible reference knowledge. A variety of strategies were described that partly ameliorated the problem, although generally the response by social media platforms themselves was slow, often only at the behest of outsiders, and inhibited by many problems of attribution and tracking of online aggressors.

In one real world circumstance involving crisis management with North Korea, it emerged that social media plays out differently in the Korean context than in the west (there being relatively far less Twitter users in South Korea). In general, social media widens the pre-existing political and ideological divide that characterizes almost all public discourse in South Korea, and after a land mine exploded in 2015 injuring South Korean soldiers, social media sought to punish the

North. Even the fake non-combatant social media report was quickly dampened by countervailing messaging by US Forces Korea.

Where social media platforms have attempted to curtail manipulative and dangerous use of their services, they have found that one of the best ways to do so is to simply slow down the ability of users to run their campaigns. This can be achieved by automated systems that shut down sites found to be fake sites and by identifying core behaviors used by social media activists that become markers that can be used to create traction and slow down their ability to game the rules of social media by using bots, anonymity, etc. One problem in such regulation of online behavior is that context is critical to determining what is and is not dangerous, and a human must be in the loop. However, once deep fakes become widespread, even having a human involved may not suffice to determine the truth content of a specific post or site.

Given the speed of nuclear early warning and nuclear command decisions, following these pathways is likely too little, too late. Indeed, as one social media practitioner said, "If we find something [really bad] in the nuclear community, then who do we really tell is the question!"

This comment implies the need for what another participant called a "truth infrastructure" trusted by users as able to provide legitimate and authoritative review of what's real versus what is fake. As the number of points of governance in the nuclear weapons field increases—in part due to the proliferation of the weapons, and in part due to the involvement of more actors lower in the governance system, especially of cities—it seems clear that a pre-existing source of authoritative information on the status of nuclear forces that is judged to be credible by nuclear weapons states and independent of their own and their adversary's early warning systems is the only sure-fire way of overcoming the pernicious effect of social media on early warning assessment and command decisions.

In this regard, the comment was made: "Let's back away from what's true or not. What we see is the weaponization of uncertainty. It's hard to know if it's true or false. We see false statement but see more opinions that are hard to determine if they're false." As nuclear command and control is defined by the imperative to make decisions in the context of inevitable uncertainty, what becomes important in a world saturated with social media posting is not just finding an answer—of which there are any number of competing claims—but having to search or the reliable answer. "This," it was said, "is a big shift. In an information-abundant world, you can't browse for information; you have to search. If you rely on social media or the Internet, you get to answers faster. This makes people go to the first answer you find."

One possibility is that first responders, especially at the city level, might use their information systems to provide credible information to reporters and mass media; and to fold carefully evaluated social media reports into their information after first validating the early reports with a variety of real-time sensors. First responders noted that it is essential to ensure that the messages across cities and counties are aligned and consistent, to avoid public confusion and disenchantment with official sources of early warning. As one first responder noted, "You need

to have a battery of circuit breakers to preempt and steer false information away - one message is not going to solve it."

Participants referred to the case of Hala Systems that provides early warning of pending bomb attacks to civilians trapped in rebel-held areas of Syria. Hala observe aircraft using a variety of data-mining techniques to predict attacks and sends Take Cover alerts via via Telegram and Facebook platforms. In the case of ShakeAlert, an earthquake early warning social media App that uses smart phones to collect seismic data by detecting ground motion, and then collates and interprets the data to predict location and intensity of the pending earthquake sufficiently quickly to allow critical infrastructure and individuals to take immediate protective steps (that is, within 15 seconds), the system relies solely on physical sensors and automated interpretation and communication.

Thus, even in the case of threats with extremely short timelines—minutes in the case of bombing, seconds in the case of earthquakes, it is possible to use social media to inform and guide humans to respond in ways that reduce risk, without the risk of the alert system being hijacked by malevolent actors. The critical element is that the users trust the reliability and authenticity of the information sent out—just as it would be for an independent, impartial third party nuclear early warning system.

This discussion left the workshop with some key questions for exploration in the speed scenarios.

For nuclear armed states:

- Should nuclear early warning systems include social media in their threat assessments?
- Should nuclear early warning systems ignore social media reports of attacks to avoid increasing frequency of assessment and of eventual assessment error?
- Should they rely on the maturity, competence and professionalism of their adversaries to assess the status of their own nuclear forces, or is it time to start building collaborative information systems that reduce the risk that they may make errors, including errors that might arise from social media reports?

For everyone, including non-nuclear states, and non-state sectors such as social media:

- Is there a third party that can provide real-time status of nuclear forces that would serve as an independent reference for nuclear weapon states early warning systems and commanders and everyone else?
- If so, who should take the lead in creating it?

These and other questions formed the basis for the Speed Scenarios that are described in the next section of this report.

IV. SPEED SCENARIOS: "CIRCUIT BREAKERS & SHORT CIRCUITS: SOCIAL MEDIA AND NUCLEAR EARLY WARNING SYSTEMS"

This section presents four "short circuit" scenarios that explore how social media might short circuit nuclear early warning systems and nuclear command decisions; and how in turn circuit breakers might be created that avoid or overcome the destabilizing effect of social media on nuclear early warning systems and nuclear command decisions. First we describe how scenarios work. Second, we define some terms used in the scenarios developed by the participants. Third, we offer the four narratives sketched at the workshop and refined afterwards in an iterative process. Fourth and finally, we present some implications for policy makers at every level on some ways that the hazard created by social media for nuclear early warning systems and nuclear commanders might be ameliorated.

a. The Scenaric Process

Scenario *(sce·nar·i·o /səˈnerē¸ō/)*

Noun. A postulated sequence or development of events.

Scenarios are stories created and used to explore provocative yet plausible ways in which the future might play out. Each of the four short scenarios that follow begins with a conflict that increases tensions between nuclear-armed states, making them susceptible to the triggering effects of social media. Each include:

- **A timeframe:** Each scenario is set in the year 2021. That's not far into the future—yet we know that the form and content of social media *will* have changed radically by then. Political and military conditions may also change significantly within this timeframe, shifting the *driving forces* for war, peace, and conflict. Although demographic and economic factors are less likely to change much in this timeframe, some biological, epidemiological, political, and ecological events—like pandemics, deadly temperature shifts, earthquakes, or election upsets—can occur almost without warning, jolting entire social and political systems.
- **A baseline conflict:** In each scenario, a crisis is emerging or is already full blown, and key stakes and issues are defined—including confounding circumstances, precursors, and premonitions.
- **Social media trigger or stand-down effects:** These are social media events that lead to the alerting of early warning systems, or to the de-alerting or non-use of nuclear weapons after an alert has been sent to top-level nuclear commanders.

Working in small groups, each team was given a scenario to adapt, making it more robust and increasingly plausible, adding elements such as:

White noise: The hubbub (and even fever pitch) conversations that define the conflict, drowning out alternatives just when they may be most important to examine.

Wild cards: Unanticipated events that shock the entire system. Some wildcards are large scale, while others consist of multiple, small, geographically distributed, low-probability, or previously unknown trends or events that occur at the same time. Wildcards can happen through sheer coincidence, or they can spring from previously unrecognized common causal factors.

Crossroad: Moments when leaders are obligated to make tough choices or decisions that set or change outcomes. These can also be moments when a third party or "small" agent makes a difference, often at considerable risk to themselves.

Each team then developed a "circuit breaker"—a solution that interrupts the social media information cascade and changes the trajectory of the conflict away from nuclear war—explaining which actions must be taken by which global actors to break the circuit.

IV.2. DEFINITION OF TERMS

Short circuit (short cir·cuit /SHôrt ˈsərkət/)
Noun. In a device, an electrical circuit of lower resistance than that of a normal circuit, typically resulting from the unintended contact of components and consequent accidental diversion of the current. *Verb.* Shorten (a process or activity) by using a more direct (but often improper) method.

In this context, the term "short circuit" refers to the ways in which social media could—intentionally or inadvertently—trigger nuclear early warning systems with the potential to cause nuclear war. A social media-induced short circuit could introduce extraneous, unreliable, or insignificant information into nuclear early warning systems, falsely indicating that nuclear weapons are "on the move" or casting doubt on the reliability of what standard indicators—monitored and reported to sensors, then assessed before being passed up the nuclear weapons chain of command—are signaling.

Possible results of a social media-induced short circuit include:

- Political leaders connected to social media outside decision support systems misinterpreting or overruling reliable nuclear early warning systems
- Warning systems failing altogether
- Social media reports providing additional "evidence" that confirms a false-positive report by an early warning system that a nuclear weapons state is under attack

Social media could *directly* or *indirectly* spark a short circuit:

- **Directly:** People monitoring multiple sensors in nuclear early warning systems are also paying attention to social media feeds. Such monitors could judge social media posts to be significant enough to require a nation to evaluate whether it is under attack.
- **Indirectly:** Social media information streams could change a nation's propensity to attack by ramping up perceived conflict and raising the domestic political stakes of not responding aggressively—ultimately leading to provocative actions that feed an escalation spiral.

Circuit breaker (cir·cuit break·er /ˈsərkət ˌbrākər/)
Noun. An automatic device for stopping the flow of current in an electric circuit as a safety measure.

In this context, "circuit breaker" refers to a mechanism that interrupts the flow of inflammatory or inaccurate social media to early warning systems, either through blocking, offsetting, or neutralizing information that would otherwise lead to degraded decision-making. It can also refer to the role played by social media in providing corrective information that offsets false positives flowing through the early warning system, or otherwise compels decision-makers *not* to fire their nuclear weapons.

Circuit breakers use competing information flows—other social media, hotlines, independent, credible sensors, etc.—either to prevent alerts altogether or to signal leaders and/or forces that they should stand down after an escalation spiral has begun. Circuit breakers potentially offer other benefits: enhanced ability to communicate and to negotiate termination of nuclear war once hostilities have begun.

IV.3. SHORT CIRCUIT 1: A SWELTERING CRISIS—THE UNITED STATES AND NORTH KOREA

It's August 2021, and the planet is on fire.

For the second summer in a row, northern hemispheric cities are enduring a protracted, multi-week heat wave, with daytime temperatures reaching 50C (122F) in the shade. Forests are ablaze across Europe, Eurasia, and North America, as far north as the Arctic Circle creating a pall of smoke that circles the entire planet. The death toll has already topped 100 million and there's nothing but more heat in the forecast. Nobody wants to hear about the cause—"air conditioning thermal exhaust piled on top of preexisting heat islands, boosted by a sudden and possibly irreversible phase-shift in global heat circulation from the equator to the poles driven by climate change," as one lofty scientist, one of the privileged few with access to electricity, put it. China, Japan, and Russia all report casualties in the many hundreds of thousands, but the worst-hit country is North Korea, where already poor water systems and military-security controls have millions realizing that if they stay put, they'll die.

Climate refugees from Pyongyang, Sinuiju, Nampo, and elsewhere in North Korea—along with a more controlled flow from China, Japan, and South Korea—are flocking to borders in hopes of crossing into less deadly climate zones. In total, at least half a billion people are on the move throughout the region, migrating *en masse* from central cities to highlands and coastal settlements. Will they ever move back? And where will they stay in the interim?

The heat is up in other ways, too.

In the United States, President Trump, who was re-elected in November 2020 by a popular vote majority (due in part to his having negotiated the end of the Korean War, for which he won the

Nobel Peace Prize, and in part to an inexplicable and possibly malevolent election-day electrical grid failure in several battleground states) has just staged a triumphal military parade in Washington, DC. He's feeling particularly bullish thanks to his most recent political win: successfully pulling the licenses of major mass media companies opposed to his policies and reelection. He has also wrested control of the public internet away from its former governance structure, putting it under DOD control, citing a national security emergency.

Following Trump's move, Twitter shut down immediately, followed by a slew of other social media platforms that chose not to cooperate. Others went into exile, setting up shop on offshore islands and operating from international waters via satellite connectivity. Many internet renegades (including Tim Berners Lee and his Solid network) ended up in New Zealand and Australia, where they began building a new non-state internet called OOD.net (featuring a new social media channel, Twitler).

It's an understatement to say that global tensions were running high. So maybe it shouldn't have been a shock when everything came unraveled.

On the day after Trump staged his parade with troops squelching through melting tarmac, South Korean marines mistakenly fire upon and down an Asiana civilian airliner arriving at Incheon airport from San Francisco. As reports of the aircraft exploding in mid-air flood social media, some declare it a United Airlines plane (which flies the same route, and has a Star Alliance codeshare with Asiana, leading to confused mass media reports). Immediate and inaccurate intelligence place the blame for the attack on an increasingly chaotic, desperate North Korea, leading South Korea—the party actually responsible for the incident—to launch a knee-jerk retaliatory attack against North Korean aircraft flying on the western perimeter but north of the demilitarized zone.

And that's what blows the lid right off the Korean pressure cooker.

Remember that kumbaya moment in December 2018, when North Korea declared it would dismantle most of its key warheads and missile facilities? All that ceremonial press coverage showing the North Koreans handing over fissile material to China and destroying warhead parts under the supervision of joint US-China-Russia teams working with the IAEA, all of it complete by June 2019? Oops. Guess who still has nuclear weapons? Within minutes of the North Korean aircraft plummeting to ground, Kim Jong Un defiantly announces that he will deploy 10 previously undisclosed, nuclear-armed, long-range missile launchers. The message to the United States and South Korea is crystal clear: Stand down, or we will unleash the largest nuclear strike in history.

The United States is already on red alert, a fact communicated that evening by a Presidential Alert message that hits every American mobile phone. So are the South Koreans, and so is Japan. Having veered sharply from Article 9 of its Constitution (which prohibits waging war) toward a much more assertive posture throughout East Asia, Japan now has at least 20 nuclear weapons of its very own, although none have been tested except on supercomputers.

As the world contemplates disaster, tempers explode, and temperatures soar, millions of Twitler accounts suddenly start lighting up with the same seemingly official message, sent and re-Twitled so quickly that it's hard to figure out its origins: *All US and allied non-combatants on the Korean Peninsula are to report for immediate evacuation.*

WILDCARD (Sidebar)

President Trump decided to shut down or install political appointees to control all remaining social media platforms and news organizations (including digital media outlets like BuzzFeed) in the US under the guise of national security. Thanks to a newly expanded and strengthened Patriot Act this action is legal. Americans now have no access to information outside of the administration's state-run news channels.

CIRCUIT BREAKER

How might we determine what's true and convey the best information in this scenario? The team assigned to this scenario thought of several ways to engage the problem.

- Launch a trilateral fact-finding mission, with joint press conferences to communicate one unified view of what happened. Both diplomatic and military officials would work together quickly to uncover the facts.
- Pre-establish protocols for government collaboration with the media in the face of confusing and escalating international incidents. NGOs, media companies, relevant businesses would all be engaged and aware of these protocols.
- Invite local press and social media into a joint press conference. Communication will be a key aspect of the response.
- Send civilians ("human shield") into North Korea as both a political gesture and also to deliver much-needed humanitarian aid—in other words, leverage the relative perceived impartiality of third parties, like humanitarian groups, to push the international community to a peaceful solution, reestablish trust, and mitigate panic.

The team focused mostly on the question of communications—a problem they could solve for more easily than tackling the complex humanitarian challenges found in the DPRK. How might we best convince the United States, South Korea, and North Korea that they all have common interest in the crisis not escalating further? There would need to be a mechanism for them to communicate with one another that was trusted and understood each by all parties.

IV.4. SHORT CIRCUIT 2: FAST AND FURIOUS—INDIA, PAKISTAN, AND A NON-STATE ACTOR

It's December 2021, and we are in the midst of a news cycle like nothing we've ever seen.

December 2

From their new Caliphate in the lawless Afghanistan-Pakistan border region, leaders from the offshoots of Daesh—the Islamic State of Iraq and the Levant in Afghanistan—send simultaneous posts to multiple social media channels: *We have successfully commandeered two Pakistani nuclear weapons with the assistance of insiders from the Pakistani Air Force.* The weapons have been strategically moved to two global capitals—but they do not specify which cities. *If Western powers do not leave the Middle East and Central Asia immediately*, they announce, *we will detonate.*

The United Nations Security Council (UNSC) convenes an emergency session, as each nuclear weapons state also attempts to assess the credibility of this threat on their own. Citizens panic in many large metropolitan areas in the United States, India, and the UK, and begin to evacuate of their own volition, motivated by social media speculation about which cities might be most at-risk.

December 3

Terrorists apparently affiliated, funded, and controlled by the Caliphate commandeer a fully loaded liquefied natural gas (LNG) tanker in the Straits of Malacca, setting its course toward Chennai, a major Indian port city adjacent to a nuclear power plant, broadcasting their efforts on Facebook live for long enough to make the story real before getting taken off the air. Such a tanker has enormous explosive potential, on the order of a small atomic bomb. Combined with fear that nuclear warheads may be *en route* from Pakistan aboard ships, ports have been shut down to incoming vessels on the US West Coast and in Hawaii, and Guam, as well as in Japan, South Korea, China, and Russia.

December 4

Hackers connected to the same terrorist group disable traffic controls and surveillance cameras in Chennai, including the power grid, grinding the region to a halt. A heavily armed terrorist commando team shoots its way into the nuclear plant's control room, threatening to detonate a shaped charge (which would cause it to de-flood and catch fire) if their demands are not met. They rely in part on crowd reporting on social media to ascertain the location of security forces and time their operation to exploit the attacks on the transport systems. Chief among their demands is that the LNG tanker be allowed to anchor opposite the reactor and the Indira Gandhi Centre for Atomic Research, the birthplace of the Indian nuclear power and bomb industry.

December 5

Isolated Kashmiri terrorist attacks in northern India, to include New Delhi, precipitate increasingly significant military blows between Pakistan and India, including India attacking Pakistani terrorist bases along the border inside Pakistan. The UNSC resolves that the two states should negotiate a ceasefire but does nothing more because China opposes the other great powers backing India, given its massive investments in the Belt and Road Initiative and its

related investments in Pakistan, and its suspicions of India's intention to push China out of the South Asian region.

December 6

Reports—but no evidence—of an assassination attempt on the Indian prime minister are traced via social media to a Pakistani controller based in Kashmir. Soon after, the new Caliphate issues a new threat via social media, claiming they've placed a nuclear warhead in an Indian city and will detonate it unless India withdraws its forces from Kashmir immediately. India rejects the threats, accusing Pakistan of state-sponsored terrorism, and calls on the UNSC and other states to lend assistance. The UNSC resolves to do so in an ambiguous manner in part aimed at deflecting the residual nuclear threat from themselves.

December 7

The terrorist group detonates an unknown weapon in the hijacked city of Kalpakkam, with widespread damage at the atomic reactor facilities. While details are unclear, it appears that the tanker was used and the explosion was immense. The role and likely spread of radiological materials also remain unclear, but the attack is live-tweeted by terrorists, victims, and global intelligence services, effectively turning the catastrophe into a global media event. Pakistani social media report that a Pakistani nuclear-armed submarine is preparing to put to sea, based on posts by its crew to their families. India puts its air force and missile force on high alert, begins massing significant ground forces, and puts Pakistan on warning. Pakistan reciprocates by putting its military forces on their highest level of warning.

December 8

The Pakistani prime minister responds with a Tweet that threatens India and others with nuclear reprisal if any military action is taken. Without mentioning the United States by name, he also asserts that any attempt to intervene with special forces to attack Pakistan-based terrorists will be met with an "instant and devastating response". He calls on all Pakistanis to monitor the skies day and night, and to use social media to instantly report any incursions by "outsiders."

No one knows where this set of spiraling threats and attacks is heading. What is clear is that non-state actors have skillfully executed nuclear threats and attacks in ways that have caused multiple nuclear-armed states to threaten one another—and that their plot has launched nuclear escalation spirals across the global landscape.

WILDCARD (sidebar)

National Geographic's Blue Terra Project is on a mission to create a detailed topographic map the entire sea floor using open-source underwater drones. Vetted volunteers are able to remotely pilot the drones via live-streamed video and analyze data in real time. During a standard mapping session, volunteers discovered nuclear submarines from India hidden just off the coast of Pakistan in the Arabian Sea. The accidental discovery is immediately reported

across media outlets, despite the Indian and Pakistani government's best efforts to squash the story.

CIRCUIT BREAKER

The group assigned to this scenario identified several key action questions: How might we support and protect the communities affected by the explosion? How might we understand what has happened? Most importantly, how might we create the necessary political space to deescalate the situation?

- Tech luminaries/social media companies have an opportunity to limit the amount of social media traffic happening and this could be used to slow down all the communication. They could whitelist a series of national and international policy and scientific users so those specific social media messages would still move forward and reach a broader audience. In other words, social media companies would have the opportunity to limit the amount of negative/inflammatory social media traffic happening, and this could be used to slow down the frenzied pace of communications and war planning. With the noise lessened somewhat, scientists and organizations could go in, verify what happened (and what didn't), and communicate their findings in a coordinated way. This was all predicated on social media companies being willing to take these steps, that they could actually control the virality of stories, and that they would know who to call/speak with in government so as to make sure things were being done beneficially.

  This might not work—it is just a possible mitigating tool, and it could have unintended consequences. But the necessary action ahead of any type of crisis like this would be to coordinate among different social media platforms as well as with federal and local authorities to be aware of potential crisis situations. If we only get one social media platform to do this, it doesn't prevent the others from amplifying the negative/inflammatory messages. We would need practice for emergencies and what to do to reduce panic, and forethought about who to "whitelist" during such a crisis, as well as agreement from these companies that these are steps they would be willing to take. Preparing/informing the population about methods of communications during a scenario like this would also be key.

- Meanwhile, states would still be having backchannel conversations with one another. There would also likely be movement of human shields, with US and maybe even Chinese personnel moving to where conflict could otherwise begin to spiral out of control (in keeping with historical practice in South Asian crisis management).

IV.5. SHORT CIRCUIT 3: MUTUAL MISCALCULATIONS: NATO, THE UNITED STATES, AND THE RUSSIAN FEDERATION

The veins in US President Michael Avenatti's temples are throbbing so hard he wonders, fleetingly, if he's having a stroke. He has just minutes to decide how to respond to global media

hysteria —emanating from unverified social media reports in Russia and Eastern Europe—that nuclear-tipped missiles are in the air and headed to several US cities.

As appalling as it was, few would have predicted that the global financial downturn of 2019 would lead to nuclear war. First the Chinese financial sector softened, and along with it trade financing became more difficult. Despite having relatively small direct impact, these challenges were enough to push already struggling European Union economies over the edge, and each Organization for Economic Cooperation and Development (OECD) nation enacted fortress-like fiscal measures to stave off economic catastrophe. When banks shut down, domestic business finance was unavailable, leading to massive layoffs and plummeting consumer demand. Stocks and property prices plunged.

Seizing an opportunity created by most nations' preoccupation with domestic affairs, Russian forces moved into Belarus, escalating tensions with NATO. Concern about Russia's intent on its western border with Europe intensified. When Russia mobilized a large flotilla off the coast of Syria the distrust was amplified even more. And Vostok-2018 in the Pacific, involving 300,000 troops and 900 tanks—the largest military drill since the cold war--- with China also participating, significantly increased the West's concerns about the warming relationship (and economic collaboration) between these two great powers, leading the US to strengthen ties with Poland and to speed up deployment of missile defenses in Poland.

Latvia, Lithuania and Estonia were profoundly worried about being Russia's next target of aggression. Small but unmistakable signs appeared that NATO was preparing to defend the Baltics. The Russian enclave of Kaliningrad, situated as it was between Poland, Lithuania and the Baltic Sea, was particularly concerned it would become a battleground as distrust and fear simmered in the US/NATO/Russia relationship. Chatter proliferated on social media: Were we suddenly on the verge of World War III?

Meanwhile, Vladimir Putin was confident he could rely on strong economic growth due to stable and rising oil prices—and Russia's investments in new military technologies were really paying off. The warming relationship with China would be beneficial to the economy as he was sure China's financial outlook would return to earlier growth figures. His confidence was tempered by his wariness of the United States, however, because of the US' increased willingness to use cyber weapons offensively. That concern notwithstanding, and seemingly bolstered by confusion and dissent sown in the West by a social media offensive orchestrated by the Kremlin, Putin ordered the Baltic Navy Fleet (headquartered in Kaliningrad) out to sea. At the same time the Russian flotilla sailed west from Syria, its destination unknown.

In Turkey, President Erdogan's criticism of US foreign policy became openly hostile and threatening. While most NATO allies aligned along traditional Cold War lines, some (like Turkey) demonstrated skepticism about US policy and commitment given its recent history of erratic decision making. Loyalties shifted; even old NATO allies such as France and Germany indicated mistrust of the United States.

While still in its infancy, the US-backed Polish missile defense system was trumpeted as a deterrent to the Russians making any further moves beyond Belarus. This message reached a

fever pitch on social media networks through posts on verified and fake accounts alike. The info wars began in earnest when the Islamic State used social media to start rumors that the Russians were preparing to invade the Baltics and at the same time spread the rumor that the Americans were preparing to attack the Russian Baltic force in Kaliningrad.

All over the world, citizens posted impossible demands on their leaders: Avoid these wars! Protect us from attack! Up with NATO! Down with NATO! Up with Putin! Down with Putin! Up with Avenatti! Down with Avenatti! The clamor was both confusing and impossible to ignore. It also made it impossible to separate important signals from the noise.

When the US increased its security alert to DEFCON 3 with preparations for DEFCON 2 in place, Russian counterparts did the same.

And then ISIS made its move, exploding a small nuclear device in Syria but made no claim of responsibility, and all hell broke loose in the Middle East. Blame was placed on the US, Russia, terrorists ... The violent accusations and denials were everywhere. There was confusion on the whereabouts of the all the approximately 50 US nuclear weapons at Incirlik Air Base. No one could convincingly deny guilt. And now the US early warning system is showing incoming missiles.

Flanked by his closest advisors in Washington, President Avenatti is focused on trying to make sense of conflicting information coming from American nuclear early warning systems. Nothing in his training as a litigator or political operative provided what he needs now: a way out of a dark and extremely perilous box. Putin has not picked up the legacy analog phone of the US-Russian hotline throughout the escalating crisis.

Avenatti is a true Twitter aficionado. He feels his Twitter finger twitch and reaches for his phone.

WILDCARD (sidebar)

Thanks to increasingly cheap technology, every device in the EU's medical system has been networked and put in the cloud for efficiency. The EU system has been hailed as a modern medical marvel and other developed countries are following suit. A rogue hacker group in Russia has announced that they discovered a vulnerability in the network and are threatening shut down every single operating device in every medical facility across Europe. If they succeed, tens of millions will die instantly. Critics say they are bluffing, but there is no way to tell. Families all over Europe are clamoring outside of medical facilities to get to their loved ones.

CIRCUIT BREAKER

The most immediate challenge in this scenario is figuring out what's really going on while dealing with the increasing panic. Are there ICBMs in the air and if so from where?

The team assigned to this scenario thought the best way to do both the misinformation and the panic would be to have *already* established an international third-party organization, comprising foreign military leaders from United States and Russia and perhaps other nuclear

power states. The UN Security Council is seen more as a tarnished organization so a new organization likely is required. The organization would build on something the United States and Russian have already discussed. It would need to be a respected by the international community. Also, you want to make sure that the organization's mission is limited. It would not be a peace keeper but just provide information, purely information, so that the organization is less corruptible.

The organization would have three responsibilities:

1. **Sensing.** The organization would be in charge of technology systems designed to sense international incidents that could be related to nuclear war. This would include simple sensors that can detect nuclear weapon or missile launches as well as human observers—the already several million people who have downloaded an app and can both proactively send messages as well as respond to a query (e.g., Have you seen evidence of a missile launch at your location?). They would also need the ability to detect social media storms as they are brewing.
2. **Validation.** The organization would be able to see clusters or reports or activity from certain areas at certain times, then launch a deeper level of evaluation. If there is seismic detection as well as three reports of contrail in the sky, that's worth investigating. Or there might be a social media storm but no other information to validate what's being "reported."
3. **Dissemination.** The organization would be responsible for communicating with the leaders of the different nuclear powers, sending the same information to all parties: "our system is saying X and this is what we think is going on." This creates lines of communications at the leadership level. Tying the key dissemination notes to social media, the bottom of the megaphone, would also help mitigate panic.

IV.6. SHORT CIRCUIT 4: EMBRACE TIGER, RETREAT TO MAINLAND—CHINA, TAIWAN, AND THE UNITED STATES

December 1, 2021 – Taipei, Taiwan

The newly elected Taiwanese government's declaration of independence from China sent political and military shockwaves across East Asia. Since the upgrade of its status and forces in the first decade of the 21st century, China's Second Artillery Force viewed itself as "the arrow on the bow and poised to strike," able to exert tremendous pressure on the "Taiwan separatists." Thus, Taiwan's declaration was a direct challenge to the PLA's core identity. It hankered for orders from the Central Military Commission to show Taiwan it could not ignore the military power of the mainland.

It took less than a day for Chinese President Xi Jinping to direct the Chinese Navy to impose a naval and air blockade around Taiwan. In return, Taiwanese forces fired a barrage of anti-ship missiles at Chinese forces, simultaneously launching long-range missiles at offshore islands where Chinese ground and amphibious forces are massing to invade Taiwan. The United States

Navy sent destroyers into the Taiwan Straits, and US strategic submarines departed Guam and the US West Coast for open ocean.

Early reports suggest that US strategic bombers are also flying to and from Taiwan on a rotation out of Guam supported by long-range fuel tanker flying in daisy chains, while US special forces and marines have been introduced into Taiwan, along with two mobile missile defense units (Patriot+).

December 14

A Chinese submarine is sunk by a US torpedo drone after the submarine penetrated a US aircraft carrier group's underwater security perimeter. The Chinese People's Liberation Army (PLA) responds by volleying missiles at Taiwanese airfields and military bases, disabling the American destroyers with salvos of anti-ship missiles and drone attack fleets. China is now threatening to use nuclear weapons in response to any US or Taiwanese attack on Chinese mainland-based forces. In a startling announcement, the PLA also threatens nuclear weapons use against US allies—namely Japan, South Korea, and Australia—should they support or assist US forces in any way.

December 15

The march to world war accelerates with North Korea's dramatic entrance to the fray—a salvo of missiles fired at US naval forces in the West Sea and Kim Jong Un's threat to attack Guam with nuclear weapons. North Korea takes these actions in response to the United States Strategic Command (STRATCOM) claim that a low-orbit US satellite that disappeared earlier in the week was shot down by a Chinese missile.

China now threatens to close sea-lanes from the Indian Ocean to the Pacific, and from the South Pacific to Northeast Asia. A Chinese warship collides with an American destroyer in the South China Sea and the two vessels open fire. Both are disabled and wallow in the ocean waves within sight of the other.

December 16

As world powers gear up for the unthinkable—all-out nuclear war—experts scramble to make sense of the role social media is playing in the escalation of conflict.

Informants on the ground in China report that covert, state-run trolls have been encouraged to use social media to express nationalistic fervor and fury against the Taiwanese leadership and population at large. Their posts accuse Taiwanese citizens of being traitors—or worse, of being "non-Chinese" and other insults such as tái bāzi (台巴子) that belittle Taiwanese, casting them as unsophisticated peasants.

This social media onslaught relies on "rumor refutation" groups, such as *weibo piyao*, first established in 2010 as self-appointed or state-cultivated "self-purification" networks. These groups are dedicated to exposing false information propagated via social media; in the present case, the explosion of anti-Taiwanese sentiment is designed to put maximum pressure on soft-

line elements in the Chinese Community Party and the military, and rumored to set online traps for Taiwanese sympathizers.

A sudden social media wave starts in Hong Kong and spreads instantly across Shenzhen and beyond to all of China, claiming that the United States has reinstalled nuclear weapons in Taiwan alongside mobile missile defense units, and that Chinese missile forces are preparing to attack these sites and American aircraft carriers with nuclear weapons. Thousands of Chinese start to leave the coastal cities for inland and police forces deploy to block them.

An hour later (there's a 12-hour time zone difference from Shenzhen-Taipei to Washington, DC), US open source intelligence analysts alert STRATCOM that these rumors may have some basis in reality. Fire orders have been monitored in communications and electronic tracking of China's strategic communication systems, and missile launchers have been observed dispersing into tunnels. But they cannot confirm if these missiles are nuclear or conventionally armed—in part because the same communications system is used for the sending orders to the missile units whether they are carry conventional or nuclear warheads. Based on artificial intelligence analysis and close monitoring of Chinese missile units, American strategic bombers are already flying along the Chinese ADIZ within range of nuclear tipped air launched cruise missiles, in case they are ordered to fire and close enough to ensure that mobile units cannot move far before nuclear weapons detonate in their vicinity. Chinese military commanders are particularly alarmed at the possibility that AI has enabled a realistic American fire strike and press for an early first strike off their own against US forces in the western Pacific to limit their damage.

US allies are pressing for a powerful American response. An advisor to President Pence in the Oval Office can be heard describing him as "preternaturally calm" as he alternately consults scripture and considers his options.

The phone on his desk rings and interrupts his meditation. President Pence puts his Bible down on the Resolute Desk and answers the call. It's from US Strategic Command, reporting that a US satellite has observed two intermediate-range missiles launching from China, one apparently heading toward Okinawa and the other streaming toward Guam. They do not yet have radar readings on the incoming warheads in space or plunging back to Earth, but that is only a matter of minutes away, they say.

In China, state operatives report to President Xi that ham radio observers have posted on social media that US stealth bombers have taken off and headed west. They also report that Chinese satellites have located two but not all three of the American aircraft carrier groups sailing in the vicinity of Taiwan. They summarize the last three Pence tweets but admit that they have no idea what he might mean by his latest: *"Proclaim liberty throughout all the land, and unto [all] the inhabitants thereof." "Freedom will prevail, for as the Bible tells us, "where the spirit of the Lord is, there is liberty." "So freedom always wins when Faith in Him is held high."*

Xi's Chinese advisors do not recognize this text from Leviticus, let alone that it is inscribed on the Liberty Bell. It does not translate well. Perplexed, President Xi quotes back to his advisors and generals from the war classic *Romance of the Three Kingdoms*, "Borrow the east wind! A

general has only one chance to storm a fortress and all his ships had to depend on the east wind to make the surprise attack successful." The generals leave to issue orders to their respective commanders, assuming that Xi means to occupy Taiwan.

President Pence takes two minutes to kneel and pray. Then he stands up, ignores the Secretary of Defense, and requests an aide to assist with selection of nuclear strike options from the football—one for North Korea and one for China. He is advised that the only way to deliver nuclear warheads on these two countries is to fire them from the US Mid-West using land-based missiles, and that these will have to fly over Russia on their way to their targets, albeit in space, not Russian airspace. He shrugs.

Pence receives and notes a report from the CIA that there are no social media reports of these missiles taking off over the heads of millions of Chinese, all armed with smartphones, yet the contagious propaganda attacks continue. The CIA is unable to determine if crowd reporting of the launches is missing due to censorship controls, or its absence is confirmation that in fact no missiles were launched and that the satellite early warning sensors have mistaken missiles for some other infrared signature. They recount a blitz of social media in China issued by the government and individual celebrities that China will pay any price to stop the United States from separating Taiwan province from the mainland and calling on the Chinese diaspora to rise up and strike against the United States everywhere in the world.

He asks how long before the missiles strike and whether it is certain that they will hit land or the ocean? He picks up his phone when...

WILDCARD(sidebar)

After SnapChat, Kik, and Weibo folded, an open source, blockchain-enabled encrypted messaging platform called _U is the new social media channel that young people have flocked to. There is no visibility into their communications for large corporations, governments, or even their parents. Disenfranchised young people across the world are using _U to coordinate massive protests against oppressive governments across the developed world; Chinese youth are particularly active on _U because the government hasn't found a way to censor it yet. To the surprise of global leaders, they've recently begun showing up en masse to live stream and protest on battle fronts giving new meaning to Social Justice Warriors. They arrive prepared with helmets, combat boots, and gas masks and lay their bodies on the line, daring governments to kill them on live stream. Among the _U Warriors on the front lines are social media influencers and other famous young people.

CIRCUIT BREAKER

The key question in this scenario is: What's the nature of the missiles that are set to hit near or on Guam and Okinawa?

Relatedly, many questions needed to be answered to determine a de-escalatory strategy. What is President Pence hearing from the Chinese? How does President Pence manage his domestic political constituencies that have been active on Taiwan? How might a new information circuit

to Xi be established or restored? The United States has no diplomatic avenues or channels open to China in this scenario, so the group focused on how they might create one.

They posit that the Chinese decide they need to climb down from, rather jump out of the tree they have climbed, realizing the immense risk that President Pence might escalate to nuclear before the missiles land. In their first move, they shut down social media and anti-Taiwanese, nationalist and murderous rhetoric in China. This shift is intended to show that they are looking for a way to back down without losing face; they are sure it will be noticed instantly.

At the same time, they flip on a concerted, all-out *diplomatic* circuit breaker by calling on trusted persons to act as intermediaries. Their message is: a) the missiles are conventional; b) they won't hit land, only open ocean; c) they want the United States reverse the Taiwan independence declaration in return for China standing down their invasion force and missile attacks on Taiwan. They propose to setup a joint committee to examine how things got out control in the Straits and naval shootouts, and to establish maritime "rules of the road."

The group came up with several personal backchannels that the Chinese might deploy including Jack Ma, Alibaba, who has the personal cell phone of Jeff Bezos of Amazon, 97-year-old Senator Richard Lugar, a prominent Indiana University graduate who has an existing close relationship with China, knowing that Pence is a graduate of Indiana University; Ban Ki Moon in South Korea simply because he is well known to President Xi although his ability to influence Pence is unknown; and the Pope because the Chinese think that Pence might cut a deal with the Pope that might help him with his domestic evangelicals problem, as the latter have been key force promoting Taiwanese independence from the god-less communist mainland. (Although only a small fraction of Taiwanese are Christian, the evangelicals have setup churches all over Taiwan, and some of the apocalyptic evangelicals have pushed hard for all-out war with North Korea and China).

This circuit breaker scenario has many circuits, fuses, and breaker switches working with and against each other. In the final phase of Chinese outreach, part of the motivation is that a power struggle may be going on inside Chinese communist party; and social media wars are part of this power struggle, with one pro or anti-war faction using the Taiwan issue, and escalating conflict with the United States, to try to flush out their Chinese adversary to force them to show their "true color" as "patriot" or "traitor" to Chinese national cause. These factions have their own trolls, as do the evangelicals in US-South Korea and Taiwan who are mobilizing online to reach to Pence.

At the same time as calls are going out to the Americans, Xi has an intermediary from Shanghai Forum contact President of Taiwan to tell him mainland is not going to invade Taiwan; and that he is relying on Taiwan to work out a deal with the Americans to reverse Taiwan's independence declaration as the price of peace.

V. COMMONALITIES ACROSS SCENARIOS AND IMPLICATIONS

The following themes were noted across all the scenarios, suggesting that elements of some or all of these may be needed in a multi-dimensional, multi-pronged, multi-channel de-escalatory strategy. It is noteworthy that whatever the role of social media in short circuiting early warning systems or in amplifying conflict and degrading decision making, traditional high-level communications between persons trusted by the political and military leaders cropped up in all the circuit breakers envisioned in these scenarios. Other elements included high and low-level hostage exchange; doing whatever it took to slow escalatory spirals; and anticipating the loss of control induced by social media and other drivers by establishing hot lines, market and civil society-based communication channels, and trusted, third party, and impartial sources of authoritative information on the status of forces.

Also common across the scenarios was a sense that the unfolding of the scenarios may have followed inexplicable pathways that were the result of complexity beyond human comprehension and emerging at a speed beyond human recognition and decision-making. In general, the scenarios portend more uncertainty, not less; that there will be more noise, less signal to contend with in early warning systems; and that almost all "facts" bearing on critical decisions in a crisis will contested, both in the policy and decision-process, and public discourse. As one participant put it: "Everyone knows everything, but no one was sure of anything."

Despite these negative trends, it was also evident in the scenarios that history leads to the future via open, not closed doors, and which doors humans choose to enter remains for them to decide. Nothing is pre-ordained, and there are many steps humans can take today that will alleviate the situation in a crisis—reducing the number of nuclear weapons, reducing alert levels, separating warheads from delivery systems, adopting a no first use declaratory policy, shifting to a deterrent-only force posture that reduces the number of deployed elements in which something can go wrong.

Along the way, the participants were convinced that social media platforms and social media users can shift the center of gravity away from the current, celebrity-driven and conflict-amplifying social media dynamic that degrades the quality of much information and towards more reliable, authenticated information while preserving the ability of users to free speech and near-instantaneous networking. In this regard, cities and civil society emerged as a set of actors and networks that may be positioned to create new forms of governance and public information goods that restrains the aggressive use of social media that may contribute to false alarms and poor decision-making at the national level, while contributing to independent, impartial and validated information that is useful to nuclear early warning systems and nuclear commanders who may be relatively poorly served by traditional sensors, early warning systems, and conflict resolution mechanisms at the level of inter-state conflict.