NAVAL SCIENCE AND TECHNOLOGY

FUTURE FORCE™

VOL. 6, NO. 2, 2020

DEVELOPING A NEW APPROACH TO CYBERDIPLOMACY

COMBATING MISINFORMATION: AN ECOLOGICAL APPROACH

EXPLORING THE POWER OF CUTE

OPERATING IN THE INFORMATION ENVIRONMENT

VOLUME 6
#02
2020

FUTURE FORCE

NAVAL SCIENCE AND TECHNOLOGY MAGAZINE

IN THIS ISSUE

Investing in Social Cybersecurity 16

BEND: A Framework for Social Cybersecurity 22

Anticipating Russian Shenanigans: Moving toward Predictive Analysis 28

Technology for Communicators: The BITE Dashboard 30

Combating Misinformation: An Ecological Approach 38

42 Identifying Misinformation Campaigns

45 Exploring the Power of Cute

48 Compile to Combat Focuses on Delivery to the Fleet

50 Navy and University of California Collaborate on Data Science

COLUMNS

Speaking of S&T
Operating in the Information Environment 04

06 How We Got Here
Dezinformatsiya and the Cold War

Editor in Chief

- **Capt. John A. Gearhart, USN**
  Assistant Chief of Naval Research, ONR

Editorial Staff (Contractors)

- **Colin E. Babb, Managing Editor**
- **Warren Duffie, Jr., Assistant Editor**
- **Jeff Wright, Art Director**
- **John F. Williams, Photo Editor**
- **Moraima Johnston, Web Editor**

Guest Editor

- **Lt. Cmdr. Jennifer Franco**

Editorial Board

- **Dr. James B. Sheehy**
  Chief Technology Officer, NAVAIR
- **Dr. David Sanders**
  Director, Corporate Communications, NUWC
- **Gregg Brinegar**
  Assistant Vice Chief of Naval Research, ONR
- **Victor Chen**
  Director, Strategic Communications, NRL
- **Christopher Lawson**
  Director, Corporate Communications, NSWC
- **Dr. Joan Cleveland**
  Deputy Chief Scientist, ONR
- **Dr. Dimitris Tsintikidis**
  Discovery and Invention Business Portfolio Manager, NIWC PAC

To submit an article or subscribe to *Future Force*, please visit our website or contact the managing editor.

Future Force Magazine
Office of Naval Research
875 N. Randolph Street, Suite 1425
Arlington, VA 22203-1995

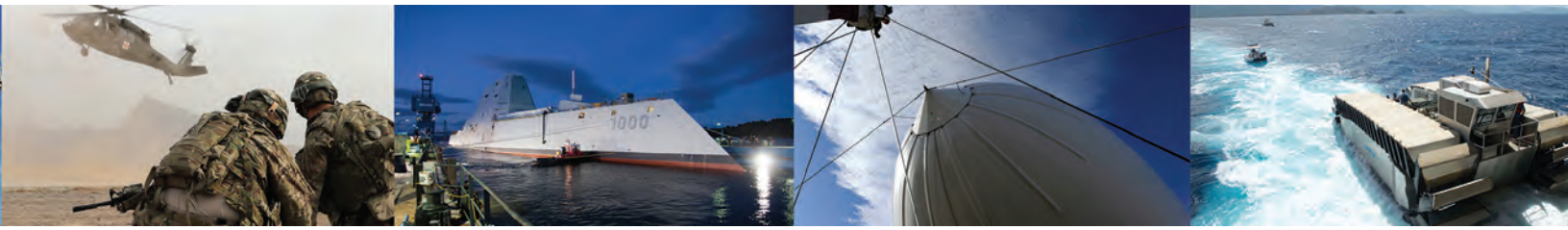
Email: futureforce@navy.mil
Phone: (703) 696-5031
Web: http://futureforce.navylive.dodlive.mil
Facebook: http://www.facebook.com/navalfutureforce

BEYOND FUTURE FORCE

Interactive features are enabled with the digital copy of Future Force:

futureforce.navylive.dodlive.mil

Mobile Download



08 Developing a New Approach to Cyber Diplomacy

It's difficult to remember now, but the internet was once a place that inspired utopian visions. Those visions may not be completely dead, but lately they have taken a considerable beating. Beginning to calm the crowd and restore civility to online discourse will take concerted research efforts.

32 NATO Integrates New Media, and So Do Adversaries

The NATO exercise Trident Juncture 2018 provided an excellent testing ground for a new tool that looks at the effects of video posts—still one of the trickier products of social media to analyze.

Future Force is a professional magazine of the naval science and technology community. Published quarterly by the Office of Naval Research, its purpose is to inform readers about basic and applied research and advanced technology development efforts funded by the Department of the Navy. The mission of this publication is to enhance awareness of the decisive naval capabilities that are being discovered, developed, and demonstrated by scientists and engineers for the Navy, Marine Corps, and nation.

This magazine is an authorized publication for members of the Department of Defense and the public. The use of a name of any specific manufacturer, commercial product, commodity, or service in this publication does not imply endorsement by the Department of the Navy. Any opinions herein are those of the authors and do not necessarily represent the views of the US government, the Department of the Navy, or the Department of Defense.

Future Force is an unclassified, online publication. All submissions must be precleared through your command's public release process before being sent to our staff. To subscribe to *Future Force*, contact the managing editor at futureforce@navy.mil, (703) 696-5031, or *Future Force* Magazine, Office of Naval Research, 875 N. Randolph Street, Ste. 1425, Arlington, VA 22203.

All photos are credited to the US Navy unless otherwise noted.

Front Cover: Illustration by Jeff Wright.

SPEAKING OF S&T

Rear Adm. Charles W. Brown, USN

Photo by MC3 Drew Verbis

The surest way to lose a competition is failing to recognize it has begun.

Today, we are in a fight in the cognitive domain. It is a fight with other Great Powers, a fight unlike any we've been in since the Cold War. Our adversaries are attacking America with soft censorship, propaganda, and disinformation. They are attempting to manipulate the perceptions of our citizens—including members of our armed forces—and doing so in order to achieve their own national goals at the expense of America and our Allies and partners.

Disinformation, propaganda, and censorship are not new challenges. What is new is the scope and speed of the delivery systems today.

Social-cyber is not the only attack vector in today's cognitive domain, but it is certainly the main difference in the public information sphere since the last period of competition between Great Powers. The mass media of the late 20th century—film, radio, and television—remain very relevant today, but they have yielded increasing influence to social media. While newspaper circulation has declined to pre-World War II levels, more than half of Americans now say they get their news from social media. That percentage continues to increase even as Americans recognize that social media is less trustworthy.

Whether there is any meaningful distinction between social and traditional media is itself debatable. As newsmakers and news reporters increasingly rely on social media to both make and break news, it becomes impossible to isolate social-cyber media from the rest of the public communication ecosystem.

The ubiquity of social media and the seemingly inescapable connectivity of our information systems have increased the potential damage of cognitive attacks by our adversaries. But technological changes have just as surely increased our own opportunity to fight effectively.

To win, we must understand the fight we are in.

In this issue, Dr. Rebecca Goolsby from the Office of Naval Research outlines the history of social-cyberattacks, trolling, bots, and other forms of malign information maneuvers and what we can do about it. Dr. Kathleen Carley from Carnegie Melon University goes further to propose how we can effectively fight in the social-cyber domain using the "BEND Framework," and other authors give us multidisciplinary perspectives to help us map and navigate the complexity of the digital information landscape.

Media literacy, especially social media literacy, has become part of our national security. When it comes to our own service members, media literacy is truly force protection in the cognitive domain. But we can't just focus on defense if we want to win. We have to put points on the board.

Winning this fight will require all hands.

Just like the sayings, "Every Sailor is a firefighter," and "Every Marine is a rifleman," every member of the Department of the Navy, uniformed and civilian, is a public communicator. Even if only as consumers of media, all of our people are cognitive combatants. As leaders, we must empower them not only to defend themselves in this domain, but to contribute to the public discourse in a way that helps us win critical battles of the narrative.

We all hope that this fight in the cognitive domain is not simply a prelude to direct conflict and open warfare. The military must be prepared to fight a kinetic battle if necessary, and to win. But there is also a chance that, like in the last competition between Great Powers, a winner is determined without actual direct kinetic

Chief of Information Rear Adm. Charles W. Brown, right, addresses members of the Navy public affairs community at Naval Support Activity Naples in November 2019.

OPERATING in the INFORMATION ENVIRONMENT

warfare. The Cold War stayed cold for more than 40 years, yet there is no doubt that when it ended there was a clear winner, and a clear loser. Ultimately, nonkinetic means decided the Cold War.

We may not be exchanging fire with our adversaries, but we are in a fight right now. We must approach this current fight in the cognitive domain as if it might be the deciding arena that determines the winner and the loser of the competition between Great Powers today.

Rear Adm. Brown is the Navy's chief of information.

►► By Colin E. Babb

DEZINFORMATSIYA AND THE COLD WAR

This political cartoon suggesting the collusion of scientists with the US military appeared in the 31 October 1986 issue of *Pravda*.

LONG BEFORE THERE WERE INTERNET TROLLS, BOTS, AND "FAKE NEWS" STORIES, THE SOVIET UNION WAS NOTORIOUS FOR NOT ONLY PROPAGANDA BUT ALSO DISINFORMATION—THE DELIBERATE SPREAD OF INACCURATE INFORMATION.

In 1787, as part of what today might be called a disinformation campaign, Grigory Potemkin, commander of all of Russia's armies and former lover of Catherine II, hosted his empress during her grand tour of southern Ukraine and the Crimea—recently wrestled from the Ottoman Turks—just as a new war with Turkey was brewing. Eager to reassure the *tsaritsa* that the new lands were filling with Russian settlers, Potemkin had mobile villages built that could be set up quickly as the royal entourage passed through, and then taken down at night and set up farther south as Catherine continued her journey. It is perhaps fitting that this most personal of all attempts at spreading disinformation has itself been questioned by modern historians as likely either an exaggeration (the empress probably knew the villages were fake) or a smear campaign by Potemkin's enemies.[1] The issue of what information is real—and what is not—is a modern problem with a long pedigree.

In the two and a half centuries since, through wars, revolutions, and ideological roller-coaster rides, it would be difficult to disagree with the observation that subsequent Russian regimes have had a certain proclivity for the deliberate spread of false information. (One recent history of disinformation even claims that the Potemkin villages story was the direct inspiration for this seemingly ubiquitous part of Russian statecraft.[2]) Soviet Russia's state security organ,

the KGB, would officially call this particular type of propaganda "active measures." In the West, however, it acquired the more forthright label of "disinformation," a word that came into English as a translation of the Russian word *dezinformatsiya* (the origin of which is somewhat murky, but it has been in use at least since the 1950s).

In their book *Dezinformatsia* [sic], authors Richard Shultz and Roy Godson define active measures broadly. "Active measures may be conducted overtly through officially-sponsored foreign propaganda channels, diplomatic relations, and cultural diplomacy," they write. "Cover political techniques include the use of covert propaganda, oral and written disinformation, agents of influence, clandestine radios, and international front organizations." They also could include military and paramilitary operations. In Soviet terms, active measures were a continuation of the revolution by other means, in times of "peace" as well as war. The Soviets had a formidable arsenal of communication conduits available to them, from various international front organizations such as the World Peace Council and the World Federation of Trade Unions, to Radio Moscow's world broadcast in English and dozens of other languages, to traditional print outlets such as *Pravda* and numerous foreign language journals and newspapers.[3] Throughout the Cold War, Soviet disinformation consistently sought to portray US military and political policies as the major cause of world conflict and to isolate the United States from its allies.[4]

The most notorious Soviet disinformation campaign—codenamed Operation Denver—was the attempt to portray the AIDS epidemic in the 1980s as the work of the Pentagon. Initially appearing in the Soviet weekly publication *Literaturnaya Gazeta* on 30 October 1985, the story claimed that scientists from the American Centers for Disease Control and the Army's Fort Detrick in Maryland had created the HIV virus from two known viruses found in Africa and Latin America in an attempt to make a biological weapon. This article sourced a supposed previous letter to the editor in the Indian newspaper *Patriot*, published in July 1983. (The editor of the *Patriot* subsequently claimed that no such letter to the editor ever appeared in the paper.) Over the next several years, Soviet media printed numerous stories reiterating and then embellishing their claims (US military personnel, for instance, were supposedly widely infected, and hence vectors for the spread of HIV overseas), many of which were picked up in outlets especially in the Third World. A commonly quoted "expert" was an East German scientist by the name of Jacob Segal, who claimed the virus was man made and originated in a lab in 1977. It turned out, however, there were limits to coordinating messages even in an authoritarian regime such as the Soviet Union: the leading Russian medical expert on AIDS openly and publicly condemned Segal's claims.[5]

Operation Denver was clearly focused on changing the perception of the United States (and the West more broadly) within the developing world—where the hottest battlefields of the Cold War took place, and where the information environment was most vulnerable to manipulation. Another purpose may have been to distract from the Soviet Union's own actual biological weapons program, conducted in violation of a 1972 treaty against such weapons. Other attempts at disinformation—such as the fabrication of the "bomber gap" (as well as the subsequent "missile gap") in the 1950s—began out of a need to hide or obfuscate military and strategic weaknesses.

Much of the literature on Soviet disinformation tactics is itself tainted by questions of veracity. Three decades after the fall of the Berlin Wall, the most notable works on the subject remain books written by former defectors who left the Eastern Bloc in the 1960s and 1970s. Some of them contain claims so bold that critics have wondered if the books are themselves attempts at deception.[6] More broadly, even now it remains difficult to measure the overall effects of Soviet active measures.

The analog disinformation campaigns of the Soviet regime were shackled to analog vulnerabilities. Traditional media—newspapers, books, radio, and television—were (and for the most part still are) driven by editorial hierarchies that tied information to clear authorship and provenance, making the spread of deliberately false information challenging when the outlet was not controlled by the state. Effective Soviet propaganda and disinformation was possible when there were active allies on the ground in the West (or Third World), be it sympathetic authors and publishers or actual Soviet agents. These issues are now fairly irrelevant in the internet age, as information has largely been uncoupled from the requirement to provide evidence of authorship and sources. *Who* is saying something, and *why*, and from *where*, are questions rarely asked of memes, viral videos, or Twitter posts. It is no accident that the current success of internet "fake news" is a result not only of the rise of technology that makes information instantly and globally available, but also the long-term erosion of faith in the processes of traditional media.

In this issue of *Future Force*, a host of authors looks at the ways in which the dissemination and manipulation of deliberately false information remains a particularly pernicious problem in the Internet Age.

References

[1]Both of Potemkin's modern biographers, Simon Sebag-Montefiore and Aleksandr Panchenko, find the story problematic.
[2]Ion Mahai Pacepa and Ronald J. Rychlak, *Disinformation: Former Spy Chief Reveals Secret Strategies for Undermining Freedom, Attacking Religion, and Promoting Terrorism* (Chicago: WND Books, 2013).
[3]Richard H. Shultz and Roy Godson, *Dezinformatsia: Active Measures in Soviet Strategy* (Washington, DC: Pergamon-Brassey's, 1984), 1-49.
[4]Shultz and Godson, 40.
[5]US Department of State, "Soviet Influence Activities: A Report on Active Measures and Propaganda, 1986-87," August 1987.
[6]Pacepa's work (cited above) as well as that of Anatoliy Golitsyn (*New Lies for Old: The Communist Strategy of Deception and Disinformation* [New York: Dodd, Mead & Company, 1984]) have been subjected to notable scrutiny in this regard. Their claims include that Lee Harvey Oswald was a KGB operative (Pacepa) and that Gorbachev's policy of *perestroika* was a KGB plot (Golitsyn).

About the author:

Colin Babb is a contractor serving as command historian of the Office of Naval Research and the managing editor of *Future Force*.

Photo courtesy of State Department

DEVELOPING A NEW APPROACH TO CYBER DIPLOMACY

By Dr. Rebecca Goolsby

IT'S DIFFICULT TO REMEMBER NOW, BUT THE INTERNET WAS ONCE A PLACE THAT INSPIRED UTOPIAN VISIONS OF COMMUNITY BUILDING FOR THE GREATER GOOD. THOSE VISIONS MAY NOT BE COMPLETELY DEAD, BUT THEY HAVE TAKEN A CONSIDERABLE BEATING IN THE PAST FEW YEARS. BEGINNING TO CALM THE CROWD AND RESTORE CIVILITY TO ONLINE DISCOURSE WILL TAKE CONCERTED RESEARCH EFFORTS.

The world of information changed sharply and rapidly in 2016. The majority of internet users, especially in the West, had been taking the changes in their information environments in stride. The 24-hour-a-day, seven-days-a-week news cycle was interesting and exciting. Facebook, Twitter, and YouTube were fun and entertaining pastimes. Chaos then descended from all sides. "Fake news" and disinformation—issues and concerns that had seemed far away from the day-to-day lives of many Western audiences—brought about confusion and uncertainty.

Information conflict has changed the world and is a cause of concern globally. The internet and its companion technology, the mobile phone, brought new possibilities for effective collaboration and knowledge sharing. It also created new vulnerabilities for social manipulation. Information conflict has been an evolving process, beginning in the early Usenet groups in the 1980s with the emergence of trolling, a form of group polarization and social hysteria creation that would change over time, with new tactics and new incentives. The potency of these social cyberattacks has increased substantially in the past four years, as states have begun to weaponize social media.

This paper discusses the emergence of new techniques of group polarization and crowd manipulation, and explains how these techniques began as part of a state-sponsored campaign to take advantage of already existing practices and problems. It considers what needs to be done in research and in civil society to create the capability for calming crowds and restoring civility to difficult discourses. These are key issues in developing a new approach to messaging and engagement with online audiences.

The Study of Bad Behavior on the Internet

In 2011, Belarusian-American writer Evgeny Morozov published *The Net Delusion*, a groundbreaking study about the use of internet capabilities by authoritarian governments to oppress and control populations. At a time when many regimes were looking to the internet as a new wellspring for peace and greater liberty, Morozov's work pointed out the dangers of what he called "cyber utopianism"—the idea that increasing access to the internet and technology could only be good for democracy and individual freedom. This attitude, he argued, ignored the mounting evidence that technology itself was neither good nor bad—it could be put to all kinds of uses, including repression and manipulation.[1]

Bad behavior on the internet has a remarkably long history. Distributed denial of service attacks, trolling, and the instigation of flame wars—the early seeds of group polarization—were all part of the internet scene as early as 1989. These affected very few people, however, compared to the billions today. In the 1980s, most regular users of the internet were university, military, and corporate researchers, some of whom were also users of the early Usenet groups, the first open forums available on the internet. Groups such as /reddit and other similar web forums got their starts in the late 1980s and early 1990s. Anonymity brought with it positive possibilities for communication and collaboration in relative safety, but also negative, potentially antisocial, and truly atrocious possibilities for crime and abuse.

By the time of the Arab Spring in 2011, cyberutopianists were lauding the internet, just as Morozov described, as a great and good thing, with seeming little concern for its downside—unaware, perhaps, of the long history of uncivil online behavior and its problematic aspects. Organizations and technologies founded on the principles of social justice, transparency, and collaboration for social good began to emerge all over the world, which seemed to justify the cyberutopianist view. Many substantive public-private partnerships emerged to create significant technological and socio-technical solutions to hard problems.

Technological solutions of note included groundbreaking projects in crowdsourcing such as Ora Okolloyah's project, Ushahidi, a platform for organizing responses to disaster and crisis that spawned a new wave sociotechnical innovation and activism based on social media for good. Groups such as Crisis Mappers emerged, technological projects such as Open Street Map opened shop, and nonprofit organizations such as Humanity Road and StandBy Task Force were founded. Disaster- and crisis-response use of the internet was developing into a new kind of expertise with many enthusiasts. Remarkably few understood the potential dangers.

When I first brought up the possibility of a downside to internet communication with disaster response experts at a conference in 2012, it became apparent that most of them had never considered there was potential for danger in their reliance on open and transparent platforms. Good information was repeated over and over; in their experience at the time, it outcompeted lies. Those of us who had studied online behavior for many years had to disagree. Swarming activities in Usenet groups, including group polarization, rumormongering, and social hysteria propagation were among the strengths of online mischief makers who were always trying to make a splash in online communication events. It was just a matter of time, organization, and opportunity for them to strike.

To help them understand the potential problems of social hysteria propagation and rumor, I collected multiple disturbing examples that showed the potential for Twitter, blogs, Facebook, and YouTube to cause harm to disaster-response and humanitarian-assistance operations. I focused on problems of ethnic strife in Assam, where a hoax had led to the train stations of India having to be shut down for three days, and an incident in Bangladesh where a Facebook post had led to attacks on police stations in rural districts. After the presentation of these facts, a two-hour discussion on its implications for disaster response and humanitarian assistance ensued. The organizer of the event, Dr. Lea Shanley, requested that I expand the two-page synopsis I had written into

Shutterstock photo by Karen Roach

a larger paper for publication on the concept of "social cyber-attack."[2] Shortly after this article was published, the "Syrian Electronic Army" hacked the Associated Press's Twitter account, sending out a false tweet about an explosion at the White House that sent the stock market tumbling.[3] Loosely organized groups such as Anonymous grew in strength and capability, looking for ways to mix online hacks, rumor, fake videos, and other novel means for getting attention, causing mischief, or advancing agendas.

Today, those initial groups have metastasized into loosely organized troll armies or troll factories. These are made up of mercenaries, ideologues, as well as those who are "in it for the lulz"—the pleasure of making mischief and causing trouble. The Russian Federation monetized these onslaughts, creating troll factories first reported by the *St. Petersburg Times* in 2013, primarily (it was reported) to focus on politics internal to Russia.[4] By 2014, we discovered the first Russian botnets spreading disinformation, social hysteria, and rumors to generate and promote civil strife in Ukraine.[5] Buoyed by their success, these Russian-backed troll armies began to attack the North Atlantic Treaty Organization's reputation and manipulate Western audiences. From there, they took on the world.

Influence and Emotion in Cyberspace Messaging Campaigns

Successful influence campaigns depend largely on the willingness of audiences to give the weaponized narrative attention and credence. Campaigns require an entry point—an existing narrative that can be turned, exploited, and adjusted to carry a "parasite payload" of information that pushes the audience to the desired point of view. This parasite information latches on to an existing narrative, such as "you cannot trust the government" or "the mass media are biased." It then expands it in desired directions, amplifying the narrative through social engineering—finding the right crowds who are likely to repeat the information because of confirmation bias and through technical means such as the use of botnets. In some ways, these new techniques are a form of "forced perspective"—a kind of narrative optical illusion that causes the audience to mistake a parasite narrative for something authentic and consonant with their attitudes, interests, and beliefs. Logical fallacies are a consistent feature of many of these narratives.

Researcher Ben Nimmo outlined the four primary tactics that Russian influence operations typically use to construct social cyberattacks in the information environment: distort, distract, dismiss, and dismay.[6] Distortion of facts includes adding attacks on persons or institutions (ad hominem fallacies), cherry-picking facts, and omitting contextual information. This was typical of the tweets, blog posts, and stories in the Russian media circulating in 2014-2016. Distraction tactics include fairly straightforward disinformation campaigns and hoaxes. Dismissal tactics are outright denial of well-documented facts, often by highly placed Russian officials.

Dismay tactics require special consideration. Russian efforts to fabricate frightening, disgusting, and horrific tales were designed to produce social hysteria by hitting its audiences' emotional "hot buttons." Human beings often are susceptible to manipulation when topics of deep commitment, intrinsic to their worldviews, are in play. Abortion is an example where one's values, emotions, and beliefs typically converge, making it difficult for an individual to be dispassionate and logical. When propagandists target these emotional topical constructs, they can create a psychological response called "amygdala hijack."

Amygdala hijack is a phenomenon recognized in psychology and psychiatry, where the area of the brain responsible for processing emotional response gets overloaded.[7] This causes the brain to lose connection to the cerebral cortex, where logical processing and the evaluation of facts occur. An individual in a state of amygdala hijack has difficulty reasoning.

The brain stem, deeply connected to the amygdala, brings up a "fight or flight" response, causing the individual to fight off information it cannot process or evaluate and thus avoids even trying to reason. An individual who experiences amygdala hijack is in a position to receive disinformation that confirms his or her bias. The influenced person may even proceed to behave more urgently in the desired direction of the cyberattacker by hitting the "retweet" button, and spreading the inflammatory messaging into his own network, which is liable to have similar hot-button vulnerabilities with respect to the message. Such messaging can spread virally in a more or less natural fashion.

Many psychologists and psychiatrists believe anger to be physiologically addictive. Anger delivers a dopamine response; many people appear to search for anger-producing experiences to help them cope with emotional issues, address fears and anxieties, and deal with even darker emotional issues that may have nothing to do with the topic space.[8] By targeting emotional hot buttons with angry-making content, disinformation campaigns can achieve rapid and pervasive amplification of a preferred stance. At the same time, this makes it much more difficult for anyone to correct the disinformation. Existing communities of habitually angry individuals are common, clustering around topic spaces known for heated disagreements such as politics and religion. Inflammatory, underlying feelings of racism are a hidden wellspring of anger that these social engineering techniques have been able to tap to move information virally. Vile remarks and discourse within the opposing community are rationalized as "only natural" or at least not as bad as what "those (other) people" are saying, thinking, feeling, and doing. The resulting vilification of

Shutterstock photo by Sundry Photography

In the first decade of the 21st century, there were many predictions that online collaboration would lead to exciting new projects. Massive open source software programs, such as Mozilla's Firefox web browser, and similar programs such as the OpenOffice software suite and the website Wikipedia, offered concrete evidence of what was possible when large numbers of coders, programmers, and writers donate their time without monetary compensation.

the opposing crowd keeps the fires of anger, hate, and disgust alive indefinitely.

A Brief History of Trolls

In cyberspace, this kind of baiting has a history that predates today's "fake news" fights. In web forums and Usenet groups of old, those who did this baiting were called trolls. Trolling began on the internet in alt.rec. usenet forums where people would instigate horrific fights known as flame wars. An instigator, for example, would make several accounts on a forum such as a Palestinian-Israeli friendship group. Then the instigator would begin to post questions that were bound to trigger someone's hot buttons. In that situation, questions about whether the Holocaust really happened set off both sides, with those on the Palestinian side readily believing the Holocaust was a hoax.

Noticing how these flame wars began with a "baited" question or post designed to promote angry responses and polarizing topics, Usenet pundits likened this practice to a fisherman "trolling" a fishing line—pulling a fishing line back and forth across the water, looking for some fish to bite. Many people took this metaphor and mistook it for the story about the troll under the bridge from fairy tales. This became an easier metaphor to work with to get to the solution: do not feed the trolls, do not take the bait, do not engage them, ignore them, and, if they continue, ban them from the forum. These maxims became common memes in the mid to late 1980s.

Trolls at that time were inciting incivility and conflict purely for the dark pleasure of causing mischief, anger, and pain; their capabilities were limited by the technology of the day. Audiences were small and the spread of the cure for trolling spread through the communities. These early audiences were culturally very similar to one another and spoke the same language.

Just as the practice of trolling was fairly easy to spot, the cure was also easy to accomplish—and spread readily after it was shown to be effective. This led to the belief (among trolls, primarily) that there was nothing wrong with trolling because it did no harm. This kind of psychological distancing from all things done in cyberspace is one of the most significant factors in the spread of malignant behavior.[9]

Today, audiences are culturally diverse and often technologically unsophisticated, and many lack experience with trolls, crowd manipulation, and rumor on the internet. Incidents of violence and ethnic conflict today often have some wellspring of trolling and its promotion of highly uncivil, polarizing discourses on the internet. The effects have been fatal. Individual acts of violence and large, polarized mobs have hit in both Western and non-Western societies. A partial list, for 2018 alone, includes the incel attacks in Toronto, alternative right group attacks in Virginia, as well as ethnic violence in India and Sri Lanka.[10]

The Profitability of Disinformation and Crowd Manipulation

Making botnets and trolling profitable was like putting a match to a flame. It set the stage for the campaigns of social hysteria, group polarization, and crowd manipulation in 2016 and beyond. Social media marketing concerns arose with the development of streams of passive income where advertising space was sold around the periphery of blogs and websites. Social influencer accounts received payments for tweets in Twitter, reportedly as high as three to five thousand dollars. Significant incomes can result from these passive streams, which rely on retaining very high readership (up to thousands of followers) and other demonstrable measures of validation, generally provided through Google Analytics. The influx of cash to conspiracy-oriented and extremist blogs and websites incentivized them to improve their performance dramatically. Additional services, from graphic design to editorial assistance, could be purchased cheaply. Thus, during this period there were many sophisticated changes to make these blogs look more like established news outlets.

This proved to be an advantage to the Russian campaign: it allowed Russian media content to be attached to narratives and viewpoints that already had some limited foothold in the target nation's narrative. Many different kinds of extremist narratives with some appeal to a small target audience could be paired with Russian content that could extend their narrative, and even validate it. It was simply a win-win relationship, with added resources for the host blogs and botnet managers. The extremists that Russian media co-opted had ready-made host narratives that were highly compatible with slogans such as RT.com's "Question more." The narratives "the government is lying," "mainstream media cannot be trusted," and all the various forms of saying "only we have the real truth" are common features of conspiracy theories of all kinds, from UFO cultists to doomsday preppers. Many people who follow conspiracy-oriented blogs and social media are almost hobbyists who do so for entertainment and escapist purposes.

Attempts by Russian propagandists to invade Western social-cyber communities were somewhat problematic. Grammatical errors aside, it is often difficult for outsiders to understand the nuances of communication, the stories and narratives, and appropriate gists that make an influence attempt compelling to audiences. The new paradigm, emerging in 2014 to 2015, was to develop narratives that seemed to be consonant with the worldview of their target audiences. The developers of these influence campaigns looked first to the hacker communities, already earning from legitimate and illegitimate activities on the internet: small advertisers, pornography sites, extremist propagandists, and scam operators using these same technologies to reach audiences. Extremist propagandists, on the fringe of political discourse, were by far the most effective in reaching audiences who were already disgruntled, frustrated, and aggrieved. Further, they had a degree of proficiency in understanding how to push divisive narratives into the fringes of mainstream audiences in their home countries.

Influence campaigns were developed that would amplify certain sentiments and beliefs and incorporate these parasitic subnarratives into host narratives of fringe political information actors. Artificial amplification of divisive Twitter rhetoric such as the #blacklivesmatter/#bluelivesmatter controversy helped to distort these positions so that they would fall into a false dilemma—the logical fallacy of "you are either with us or against us." Strategies of group polarization and social hysteria propagation became the tool of choice for disrupting civil discourses. Amplification of both far-left and far-right narratives were the result of apparent early experiments targeting these communities primarily through Twitter, blogsites, YouTube, and Facebook. Today we see experimentation in sending social hysteria narratives through WhatsApp.[11]

The Problems of Engagement with Influenced Audiences

It is evident that Russian media continue to subsidize the amplification of divisive narratives, rumors, and social hysteria through what appears to be a franchise operation, likely to include attacks on journalists and other human targets. The German Marshall Fund's Hamilton68 platform logs the activities of collections of both US-facing and German-facing botnets.[12] Their efforts show that bloggers and botnet operators target audiences in their home countries with divisive

messages, also carrying Russian media content, presumably on a pay-for-play basis. Examination of these outputs shows that these botnets primarily promote far-right political content on an ad hoc basis. Although highly divisive, content of all kinds is taken advantage of by these botnets as these events occur. Hamilton68's botnets also appear to target all sorts of popular hashtags occurring on a regular basis to get their content into larger audiences.

Recognition of this manipulation and group polarization, with the internet as a vector of infection, has been a positive step in the social processes necessary to help audiences recover from these influence attacks. The US Atlantic Council's series on bots and bot-based influence provided the public with a better understanding of botnet activities and influence. This awareness has brought increased vigilance among audiences to inspect their feeds for bot followers (and bots they might have followed). More needs to be done to help communities identify bot "infections" and fight off their own reactions to attempts to propagate social hysteria and rumor. This is a good first step—but will it be sufficient in the coming information conflicts, which will be amped up by new technologies of deceit and social engineering enhanced by artificial intelligence?

Festinger, Riecken, and Schachter's classic work on social hysteria, *When Prophecy Fails*, is as salient today as it was when it was first published in 1956. They observed that people who have a public, visible commitment to a stance are often rigidified into staying with that stance. When the prophecy fails and the Messiah does not come, the author's noted that adherents tend to "double down" on their commitment.[13] Influenced audiences, even when they know the facts, work hard to push those

Shutterstock photo by pathdoc

Many people find their identity through their engagement with online communities. When those communities lead down dark paths, it can be difficult to reach people and change the conversation. Empathy and understanding are important tools in moving people back towards the light.

facts away and hang on to their influenced decision. They often hold to those stances even more fervently when disconfirming facts are brought to everyone's attention and are difficult to deny. People in these situations rationalize away facts, push those realities to the wayside, and hold tight to their social ties within the influenced group to support their beliefs.

When Prophecy Fails explained the problem of this rigidified stance as a reaction to public ridicule, but there may be simpler and more subtle reasons than this. The psychological phenomenon of consistency bias—sticking with a stance because it has been expressed publicly—prevents many people from backing away from a previously held, publicly stated position. This is not the result of a completely emotional reaction. The need for consistency is a well-known phenomenon in influence studies, discussed by Robert Cialdini: people want to appear consistent in their actions, and often pursue that consistency in the face of disconfirming evidence. This effect is known to lead to hung juries when they begin a session with a public show of hands about the guilt or innocence of the accused. Juries that begin with a secret ballot are more open to disconfirming information. Because their initial votes are secret, they are freer to back away from positions.[14]

This is a worrying situation for the world of discourse in cyberspace. Groups who become polarized may be highly resistant to change. Campaigns to correct disinformation often appear to have had no discernible impact, at least not in the heat of the moment. Current research efforts to "find deceit" and counter it quickly, take down hate speech, and develop more punitive efforts may not have their intended effect. Instead, in at least some cases, these kinds of actions could lead to the growth of polarizing movements and a rigidification of the stances of influenced audiences.

Consistency bias may be just as important to understanding social-cyber crowds and the forces at work to influence them. In the past, influenced crowds on Twitter have done things to their profiles, changing their names and adding hashtags, they are reluctant to undo because it would signify a change or a capitulation. This phenomenon leads to the ability of bots to find their targets more accurately, identifying the hot button issues more clearly. A Twitter account is easier to walk away from, however, especially if it does not contain your real name or connections to your real life. The role of anonymity both in promoting "bad behavior" (such as taking on socially abhorrent stances or engaging in antisocial behaviors) also may help people to walk away from, and silently renounce, those stances and attitudes.

So how do we calm crowds and reconcile those who have been polarized into hate-filled, angry groups? Psychiatrist Mark Goulston outlines a plan for dealing with people in amygdala hijack, in one-on-one, face-to-face settings.[15] Goulston, a well-known expert on the topic of active listening, writes for those who are dealing with people who are "not in their right minds," such as teenagers in rebellion, parents with Alzheimer's, and angry, upset, and frightened people who have developed fixed beliefs because of a temporary misconstruction of reality set up by confirmation bias. He likens his approach to that of hostage negotiators who "talk people down."

In Goulston's solution, designed solely for face-to-face, one-on-one contexts, he presents several steps that could help people better negotiate discourses in cyberspace. First, the negotiator must recognize and set aside her own "hot buttons" and biases. It is important to know what those are because the target audience typically knows the negotiator's hot buttons very well. Many can set up a counterattack, distract the negotiator, and change their argument from direct action into a defensive position. Trying to have a civil discourse when everyone is in a position of emotional defensiveness leads to polarization of the discourse and sets up a no-win situation.

Goulston argues that the negotiator should "lean into the crazy" in a face-to-face situation with an influenced individual, creating a safe space to discuss the influenced beliefs and attitudes. Once trust has been established (however tentatively), the negotiator might begin to engage in discourse about why the individual feels and believes as he does and what might be done to address the fears that often underlie the anger and aggression. Posing gentle (not baited) questions to help the influenced individual "walk themselves back" from polarized, irrational stances can assist the individual, once calmed and out of amygdala hijack, to engage in more rational assessment. The negotiator will probably never get the influenced individual to give up their viewpoint or stance entirely. The point is to move the discourse, consideration of what is real, and how to improve the situation without people losing something important to them in the process.

Goulston's solution is a face-to-face encounter that relies on a preexisting relationship of trust and empathy to be effective. In addition, he specifically warns against trying his techniques directly on those who exhibit sociopathic behavior. Sociopaths require serious professional attention—or, in the words of George Bernard Shaw, "I learned long ago, never to wrestle with a pig. You get dirty, and besides, the pig likes it."[16]

Research Needs for Developing Cyber Diplomacy

Today's trolls and other malignant information actors have a much larger, more naïve, and more diverse audience on which to hone their craft. Trolls also are well financed, more technically capable, and more dangerous.

A great deal of energy and research capacity is currently being directed at "fake news." From a social-theoretical standpoint, identifying fake and deceptive content is probably beside the point. Logically, the discovery of who is being paid to carry deceptive content, especially state-sponsored content, may be better than finding deceptive content itself.

Cutting off the financial flow of state-sponsored largesse will naturally cause its own reactions. In internet forums, paid trolls have expressed concern about losing those outside incomes. Researchers have not explored or substantively discussed this potential means of addressing the problem of influence. This is, however, probably not the first solution to pursue. Such attacks on income, besides being difficult and legally problematic, would likely validate the narratives of malign adversaries. Income sources already are hidden and difficult to trace.

A promising first line of defense requires good methods for depolarizing influenced audiences. We urgently need more research into the susceptibility of audiences, polarizing narratives, and techniques of influence in cyberspace. Likewise, we need to explore and develop, test, and validate social-cyber and psychological-cyber techniques to defuse polarized audiences. The study of echo chambers, including their maintenance and function, as well as research on the social-psychological dynamics of polarizing discourse are required to develop new theoretical and practical foundations to create more effective interventions. What new techniques and capabilities need to be developed to support new cures for trolling, disinformation, and other social-cyber attacks? Who should use such techniques once they are developed?

No one has yet attempted to adapt Goulston's methods to the information environment. Such research should be pursued internationally, with appropriate scientific rigor and ethical oversight. Goulston's techniques are just one plausible direction of research for developing capabilities to pull groups away from incivility and bring them into problem-solving discourses, healthy exchanges of divergent viewpoints, and get them on the difficult path of finding sufficient common ground to address significant problems that affect society today. Social psychology, diplomacy studies, psychological anthropology, and communications research have additional insights to offer.

This discussion needs to include legal and ethical experts to determine its effects on free speech and freedom of association, as well as to balance these against other concerns. Trolling and botnets often work to suppress the freedom of speech of others, muting the impact of voices and opinions that oppose them, and directly, even illegally attacking individuals. There are serious arguments to be made on many sides of this question. Raising public awareness of these issues would be a good start.

References

[1]E. Morozov, *The Net Delusion*: *The Dark Side of Internet Freedom* (New York: Public Affairs, 2011).
[2]R. Goolsby, "On cybersecurity, crowdsourcing and social cyber-attack," *Woodrow Wilson Lab Policy Memo* 1 (March 3, 2013), https://www.wilsoncenter.org/publication/cybersecurity-crowdsourcing-and-social-cyber-attack.
[3]L. Smith-Spark, "What is the Syrian Electronic Army?" CNN, August 28, 2013, https://www.cnn.com/2013/08/28/tech/syrian-electronic-army/index.html.
[4]O. Khazan, "Russia's online comment propaganda Army," *Atlantic Monthly*, October 9, 2013, https://www.theatlantic.com/international/archive/2013/10/russias-online-comment-propaganda-army/280432/.
[5]N. Agarwal, S. Al-khateeb, R. Galeano, and R. Goolsby, "Examining the use of botnets and their evolution in propaganda dissemination," *Defense Strategic Communications* 2 (Spring 2017).
[6]B. Nimmo, "Anatomy of an info-war: How Russia's propaganda machine works, and how to counter it," Stopfake, May 19, 2015, https://www.stopfake.org/en/anatomy-of-an-info-war-how-russia-s-propaganda-machine-works-and-how-to-counter-it/.
[7]M. Goulston, *Talking to Crazy: How to Deal with Impossible People in Your Life* (New York: Amacom Press, 2016).
[8]Ibid.
[9]J. Bartlett, *The Dark Net: Inside the Digital Underworld* (London: William Heineman, 2014).
[10]A. Taub, M. Fisher, "Where countries are tinderboxes and Facebook is a match," New York Times, April 21, 2018, https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html; D. Bilefsky, I. Austin, "Toronto van attack suspect expressed anger at women," New York Times, April 24, 2018, https://www.nytimes.com/2018/04/24/world/canada/toronto-van-rampage.html; S.G. Stolberg, B.M. Rosenthal, "Man charged after white nationalist rally in Charlottesville ends in deadly violence," New York Times, August 12, 2017, https://www.nytimes.com/2017/08/12/us/charlottesville-protest-white-nationalist.html; Associated Press, "Inflamed by social media, mob kills at least three in India," New York Times, May 25, 2018, https://www.nytimes.com/aponline/2018/05/25/world/asia/ap-as-india-mob-killings.html.
[11]C. Bell, "No, the BBC is not reporting the end of the world," BBC News, April 19, 2018, http://www.bbc.com/news/blogs-trending-43822718.
[12]German Marshall Fund, "Hamilton 68: Tracking Russian Influence Operations on Twitter," 2018, https://dashboard.securingdemocracy.org/.
[13]L. Festinger, H. Riecken, S. Schachter, *When Prophecy Fails: A Social and Psychological Study of a Modern Group that Predicted the Destruction of the World* (Minneapolis, MN: University of Minnesota, 1956).
[14]R. Cialdini, *Influence: The Psychology of Persuasion* (New York: Harper Business, 2006 [1984]).
[15]M. Goulston, *Talking to Crazy: How to Deal with Impossible People in Your Life* (New York: Amacom Press, 2016).
[16]G.B. Shaw, *The Doctor's Dilemma* (New York: Brentano's, 1911): lxxxv-lxxxvi.

Note: This is article was originally printed in *Senior Leadership Roundtable on Military and Defence Aspects of Border Security in South East Europe* 105 (2019). It has been edited for length, and is reprinted with permission from IOS Press.

About the author:

Dr. Goolsby is a program officer at the Office of Naval Research. She is the co-lead of the NATO HFM-248 research technology group on crisis, disaster, and information technology.

INVESTING IN SOCIAL CYBERSECURITY

By Lt. Col. David M. Beskow, USA, and Dr. Kathleen M. Carley

THE SOCIAL MEDIA ENVIRONMENT CAN BE EVERYTHING FROM SILLY TO VICIOUS—AND OFTEN IT CAN BE HARD TO TELL THE DIFFERENCE. MACHINE LEARNING TOOLS ARE NOW HELPING TO FIND THE BOTS AND MEMES THAT HAVE MALICIOUS INTENT, AS WELL AS WHO IS USING THEM.

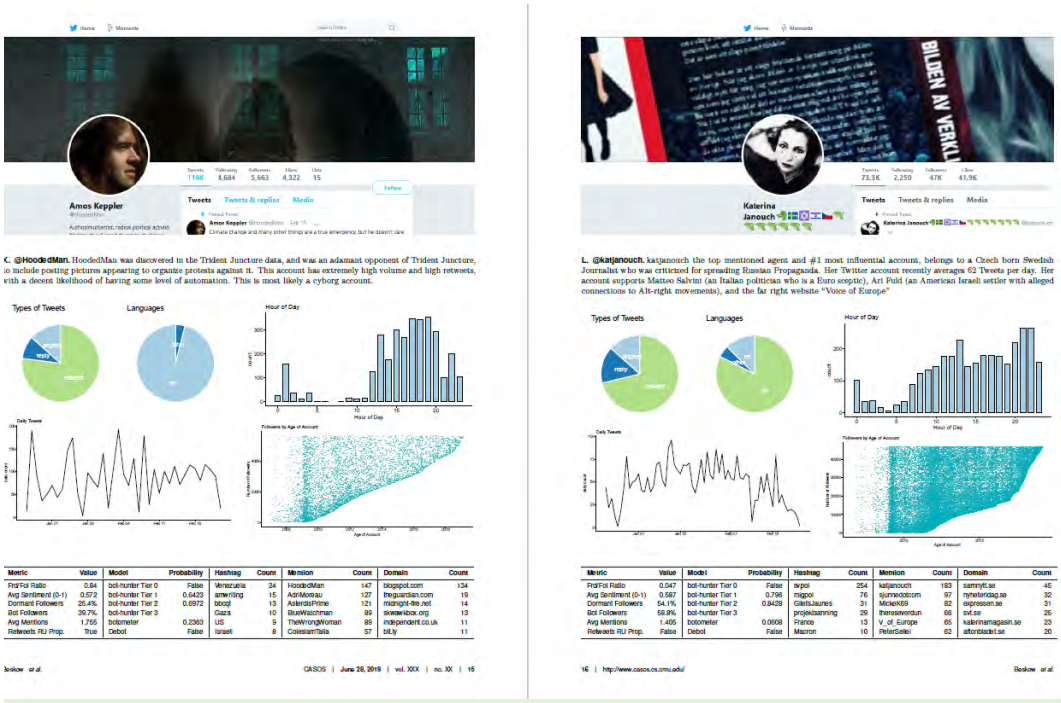
The modern information environment has created an entirely new warfare domain: cybersecurity. Much has been written about traditional cybersecurity, which focuses on humans using information systems to hack other information systems—but much less has been made about the capabilities required for social cybersecurity, which focuses on humans who use the same information systems to hack other humans. While "information operations" have existed since antiquity, the modern age has allowed them at a scale, complexity, distance, and impact unheard of even 50 years ago. As a response to this emerging threat, social cybersecurity allows a democratic society to continue to exist while retaining its core values. The National Research Council consequently has recognized it as a key computational social science area of relevance to the intelligence community.[1] To accomplish this, social cybersecurity professionals need multidisciplinary science and appropriate technology to quickly identify and neutralize modern disinformation threats that are taking aim at the core tenets of society.

Social media are the main weapons of disinformation operations. State and non-state actors execute disinformation operations across multiple social media platforms hoping to overflow into traditional media and grassroots movements. Within social media, actors attempt to manipulate the narrative as well as the network. Together, this manipulation forms an information campaign, where sophisticated actors develop multiple lines of effort that combine to support strategic goals. These campaigns are deployed in social media through curated actors (bots, memes, cyborgs, sock-puppets, etc.) and creative content (memes, videos, written propaganda, etc.). Research and acquisition efforts that support social cybersecurity must aid in identifying threat actors and content at the lowest level, and then aggregate this into a common operating picture of the threat campaign lines of effort and their strategic intent. We will discuss some of our teams' efforts to chip away at this important national security science and technology requirement.

Building a Framework

National security leaders require a framework within which to understand information warfare forms of "maneuver." We have developed such a framework, known as BEND. BEND creates information forms of maneuver similar to the forms of maneuver often used to classify offensive ground combat operations. This framework is discussed in detail in a March-April 2019 *Military Review* article as well as a separate article in this issue of *Future Force* (see page 22). The forms of maneuver encapsulated in BEND are an essential contribution to the science of social cybersecurity and are a starting place for all national security leaders trying to understand this emerging threat. In addition to building the framework, we are developing metrics that assist in detecting these forms of maneuver in social media streams. These metrics are available in ORA-PRO, a network analysis and visualization tool available from Netanomics, and in a future web version of ORA-PRO.

The national security establishment also must have the technology to outline a threat operation, associated narratives, targeted networks, and measures of impact. This tool must digest social media streams connected to emerging events and extract the threat situational template. This includes identifying the actors and content, target audience and networks, and likely desired end states. We are pioneering novel social cybersecurity techniques that would extract information campaign elements



These are example pages from the Bot Guide Guide, which gives detailed information about a variety of well-known bots.

Graphic courtesy of authors

and measures of impact from curated social media streams. These techniques blend artificial intelligence and dynamic network analysis to address social cybersecurity concerns ranging from bot detection to the spread of disinformation.

Types of Actors

A bot is any social media account that allows a computer to execute basic social media activities (such as tweet, retweet, friend, follow, like, reply, etc.). A savvy computer programmer can automate most of these activities with only a few lines of code. Researchers often try to classify accounts as either bots or humans, but many accounts are hybrids combining the activities of both. These cyborg accounts often have a human conducting nuanced two-way dialogue while the computer conducts activities at scale in the background. Bots can be positive, neutral, or malicious. Positive bots include personal assistants and accounts that warn people of impending natural disaster. Neutral bots generally focus on spam, proliferating content that ranges from commercial advertising to adult content. Malicious bots are involved in intimidation, propaganda, slander, etc. Troll accounts have human operators that specialize in aggravation as an end in itself, where divisive actions are initiated for the sole purpose of building or widening fissures in a society in an attempt to make it less cohesive. Sock-puppets are the false identities attached to troll, bot, and cyborg accounts to make them fit in with their target audience/network. The artificially intelligent assistants discussed below can assist analysts in differentiating types of actors.

BotHunter

Researchers have developed sophisticated machine learning algorithms to detect bots. This has resulted in a cat-and-mouse cycle in which bot puppet masters develop increasingly sophisticated bots to stay ahead of increasingly sophisticated anti-bot algorithms. We have developed a machine learning tool known as BotHunter to assist in finding malicious bots.[2] This is a supervised machine learning tool that has been trained on multiple bot training data and can detect bots at various data granularities. BotHunter is different from other detection algorithms in that it is designed to scale while conducting prediction on existing data. The unique focus of this algorithm is its ability to render a quality prediction on researchers' own data at a scale that is not feasible in other bot-detection approaches. In the past, social cybersecurity researchers were required to sample their data for bot detection because existing models did not scale. With BotHunter, these researchers can conduct bot detection on all their data without sampling. For example, we were able to run BotHunter using a single computer processor on 60 million tweets associated with a large world election event and received results within 24 hours (it can process approximately 4.5 million tweets per hour per thread). This same prediction would have taken months with other algorithms. In addition, BotHunter can run on existing data.

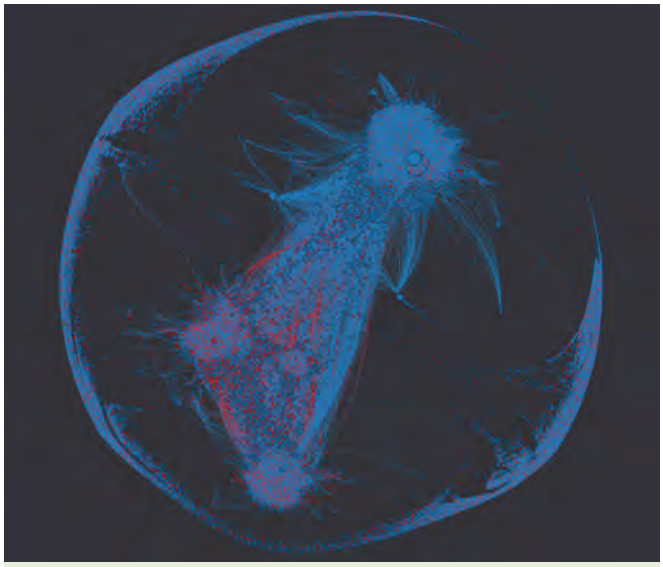
Researchers often collect data associated with a world event, and then think about executing bot detection only later. Current bot-detection algorithms often "re-scrape" the data, which is time consuming, possibly out of date, and unable to get detection data on accounts that are suspended or otherwise shut down (which are often the most interesting accounts). By running on existing data, BotHunter overcomes these limitations. The primary production BotHunter algorithm is trained on approximately 20,000 accounts that attacked NATO and the Atlantic Council's Digital Forensic Research Lab, ensuring that these predictions are relevant to national security analysts. BotHunter, in conjunction with MemeHunter (discussed below), provides state-of-the-art machine learning algorithms to analysts to assist in sifting through large social media streams.

MemeHunter

Internet memes, often thought of as humorous and harmless artifacts of the digital age, are increasingly used in information warfare. Since almost all are anonymous, increasingly political, and require sophisticated multimedia and multimodal machine learning to dissect, memes are becoming a mainstay of propaganda and disinformation operations. Memes offer the combination of an image and witty text to connect a propaganda message with a target audience, often appealing to existing biases. In addition, memes propagate in a different manner than normal viral content. Memes, as originally envisioned by evolutionary biologist Richard Dawkins in his book the *Selfish Gene* in 1976, propagate through mutation and evolution. This means they can be introduced in anonymous platforms such as 4chan and Reddit, hop into mainstream social media outlets such as Facebook and Twitter, and then move quickly to other places on the internet.

We have developed a multimodal meme detection algorithm that takes into consideration the image, text, and faces in an image to determine if it is a meme. To make this possible we also have developed a meme-specific optical character recognition (OCR) process. Traditional

Graphic courtesy of authors



This visualization shows how 56,000 probable bots (red dots) were identified out of a dataset of more 330,000 unique Twitter accounts.

OCR tools often fail when used with memes. Our meme-specific OCR preprocesses meme images so that traditional OCR algorithms can be used to extract text. We also have developed graph learning techniques to take meme embeddings and cluster them to discover the evolutionary tree that maps the mutation of memes. MemeHunter is a deep learning algorithm that can classify roughly 7,000 images per hour per thread. It automatically detects and uses available cores, and on a medium-sized server (38 cores) can process approximately 250,000 images per hour.

Describing Bots—A Field Guide

Even though bots are extremely prolific, most humans struggle to identify them. To make identification easier, we've developed a bot field guide that—like an animal field guide—provides many examples and descriptions of various malicious bots that we've found. This field guide provides a brief description and screen capture of the accounts, provides some descriptive visualizations to understand the accounts' behaviors, and offers metrics to understand how we can identify the accounts as well as what type of messages the accounts are sending or amplifying. The draft field guide has 11 sections:

1. Normal users (personal, commercial, and government accounts)
2. Amplifier bots
3. Cyborg bots
4. Chaos bots
5. Coordinated Bots
6. Social influence bots
7. News bots
8. Overt bots
9. Intimidation bots
10. Russian and Iranian bots
11. Random string bots

By walking through the field guide, analysts, journalists, and others can learn how to recognize these accounts. They find out how to look for high volumes, high retweet counts, odd friend/follower ratios, anonymity, and other distinguishing characteristics. They also will begin to understand how these accounts are used, who the targets or benefactors are, and in what conversations they are participating. They also can identify telltale signs of a bot puppet master trying to leverage a single account in multiple conversations (for example, using an account for the US election cycle and then pivoting to an anti-EU campaign in Italy).

Use in Intelligence and Public Affairs

The science and technology capabilities discussed above can provide advantages to specialists across the Department of Defense, including intelligence and public affairs professionals. As information operations are increasingly used either as an end in themselves or as shaping operations, intelligence analysts will be increasingly required to detect, monitor, map, and analyze these campaigns. Without science and technology investments in social cybersecurity, these analysts will spend most of their time looking for needles in haystacks. By using machine learning, these analysts can spend more of their time making sense of the patterns and preparing their analyses.

Public affairs offices also need to have some basic social cybersecurity techniques to see how many of their followers and retweeters are bots, cyborgs, or just dormant accounts. They need to understand what an intimidation campaign looks like and when actions should be undertaken to counter these subversive attacks. These social cybersecurity tools will help public affairs personnel monitor the threat narrative and strategic aims to make sure their message creates an appropriate counter narrative and is not being manipulated in social media.

For analysts, public affairs officers, and many others to be successful, defense leaders must set appropriate policy to enable access to the right data by the right people. Application

Graphics courtesy of authors

These are samples of typical memes directed against the United States, many of them focused on an audience of young military service members.

programming interfaces are the access point for both offensive and defensive social cybersecurity. Some specially trained individuals in the intelligence and public affairs disciplines must have the required authorities to access data and conduct analysis. Intelligence and PAO analysts arguably require "pull" authorities, whereas information operations analysts arguably require both "pull" and "push" authorities. What is meant by this push/pull information relationship is how information is curated. For the intelligence community, pull and push information is from and to a desired audience. The PAO community's primary mandate, however, is to push information to stakeholders.

Currently, much of the access to the data streams is provided by commercial tools and served to analysts. While these tools undoubtedly provide value, they do not provide all necessary data and analysis. In addition, the government is left with no actual data—which is owned and maintained by commercial entities—to incorporate into workflows and tools.

What Can These Tools Tell Us?

We examined influence campaigns in Twitter by looking at 1.6 million tweets from 330,000 unique accounts, each of which either mentions, replies, or retweets overt Russian propaganda outlets such as Russia Today and Sputnik. We ran BotHunter on the entire dataset and found that 56,000 accounts had a bot probability greater than 65 percent. These bots are visualized in the Russian propaganda conversational network below.

We then ran MemeHunter and extracted 1,616 unique memes. Visual analysis of these memes illuminated a worldwide campaign to discredit Western powers, with a focus on the United States, France, and the United Kingdom. Below we have sampled some of the memes taking aim at the United States in particular. Notice that some of these are trying to sow doubt particularly in the minds of young military service members.

Putting the Tools into Action

Our team has tested BotHunter and MemeHunter in multiple case studies and research initiatives. This includes monitoring and identifying external manipulation in multiple election events, including the 2018 elections in Sweden and the 2019 elections in the Philippines and Canada. Our team also has used these techniques to monitor anti-NATO actors and actions surrounding the 2017 and 2018 Trident Juncture exercises in Europe. We continue to monitor multiple actors manipulating information in the Middle East, often pitting pro-Saudi Arabian versus pro-Iranian information operations. We have monitored several intimidation attacks such as a 2017 attack against NATO and the Atlantic Council's Digital and Forensic Research Lab, as well as a 2017 intimidation attack against journalists in Yemen. We have monitored ongoing manipulation in Ukraine as well as global efforts by Russian and pro-Russian proxies. Finally, we have used these tools to assist defense and joint public affairs officers to understand their audience and followers better, highlighting the presence of bot, cyborg, and dormant accounts. In all cases, the tools discussed here allowed rapid triage of large and messy information streams in order to identify malicious actors and content.

National security in the 21st century will require investments in social cybersecurity. This will involve basic research into the interaction between technology and social behavior and beliefs. It will necessitate increasing investments into appropriate tools for identifying and neutralizing external manipulation of open and free societies. We also need accompanying policy changes that reflect the technical complexity of the modern information environment while remaining true to our national values. In the end, the appropriate research investments coupled with wise policy with a whole-of-government approach will ensure our nation and society continue unchanged in their essential forms with democratic institutions.

References

[1]National Academies of Sciences, Engineering, and Medicine. *A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis* (Washington, DC: The National Academies Press, 2019).
[2]David M. Beskow and Kathleen M. Carley, "Introducing Bothunter: A Tiered Approach to Detection and Characterizing Automated Activity on Twitter," in H. Bisgin, A. Hyder, C. Dancy, and R. Thomson (eds.), *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation* (2018); David M. Beskow and Kathleen M. Carley, "Bot Conversations are Different: Leveraging Network Metrics for Bot Detection in Twitter," *International Conference on Advances in Social Networks Analysis and Mining* (2018): 176-183; David M. Beskow and Kathleen M. Carley, "It's all in a name: detecting and labeling bots by their name," *Computational and Mathematical Organization Theory* 25 (2019): 24-35.

About the authors:

Lt. Col. Beskow is a doctoral candidate in the School of Computer Science at Carnegie Mellon University.

Dr. Carley is a professor of societal computing in the School of Computer Science at Carnegie Mellon University.

BEND:

A Framework for Social Cybersecurity

By Dr. Kathleen M. Carley

TODAY'S SOCIAL MEDIA LANDSCAPE DEFIES EASY CATEGORIZATION—ESPECIALLY WHEN IT COMES TO INTENT. IS THAT MEME WITH PUPPIES IN IT REALLY JUST ABOUT PUPPIES, OR SOMETHING ELSE? A NEW FRAMEWORK SEEKS TO MAKE SENSE OUT OF WHAT'S REALLY GOING ON IN SOCIAL MEDIA.

In today's high-tech battlefield, hearts and minds are won and lost in social media. The collateral effects of this fight show up on traditional media outlets. State and non-state actors who are actively engaged in influence operations on the internet end up challenging credible news sources. Lone wolves, as well as large propaganda machines, disrupt civil discourse, sew discord, and spread disinformation. Bots, cyborgs, trolls, sock-puppets, deep fakes, and memes are just a few of the technologies used in social engineering aimed at undermining the status quo and supporting adversarial agendas. Maintaining information dominance in this volatile news environment is an enormous challenge for the US Navy.

Social Cybersecurity

In response to these cyber-mediated threats to democracy, a new scientific discipline has emerged: social cybersecurity. As defined by a 2019 National Academies report,[1] social cybersecurity is an applied computational social science with two objectives:

- Characterize, understand, and forecast cyber-mediated changes in human behavior and in social, cultural, and political outcomes
- Build a social cyber infrastructure that will allow the essential character of a society to persist in a cyber-mediated information environment that is characterized by changing conditions, actual or imminent social cyberthreats, and cyber-mediated threats.

Social cybersecurity's methods are of relevance to both public affairs as well as intelligence operations. This field provides the tactics, techniques, and procedures to support a wide range of missions from providing humanitarian assistance to deterring foreign aggression. Social cybersecurity uses computational social science techniques to identify, counter, and measure the effects of communication objectives. The methods and findings in this area are critical, and advance industry-accepted practices for public affairs research. These methods also provide evidence about who is communicating what about the Navy, what methods are being used, and how it can be countered.

Some of the research in social cybersecurity is concerned with disinformation. The spread of disinformation via social media is one specialty area for communication practitioners and researchers. The term disinformation is a translation of the Russian term *dezinformatsiya*, whose origin has deep roots going back to Soviet times and even beyond (see page 6). Today, these operations have been adapted to the internet and used against the United States and numerous other countries. Four basic operations typically discussed within national security source documents are the "four Ds"—distract, distort, dismay, and disrupt. Russia and China are ahead of the United States in their ability to conduct information operations. As noted by retired US Air Force Gen. Philip Breedlove, "Russia is waging the most amazing information warfare blitzkrieg we have ever seen in the history of information warfare."

In the 21st century, the pace of warfare has accelerated. According to retired Marine Corps Gen. Joe Dunford, this accelerations means is it critical that the military seamlessly integrates "information operations, space and cyber into battle plans." Artificial intelligence is a critical capability, both to deal with the ever-increasing pace of social media engagements as well as to handle the vast quantiles of data. Some of the research in social cybersecurity is concerned with the way artificial intelligence-based tools, such as bots, can be used to manipulate groups in social media. The role of bots, cyborgs, trolls, sock-puppets, and deepfakes is another specialty area for influence operations practitioners and researchers.

Within social cybersecurity, artificial intelligence is coupled with social network analysis to provide new tools and metrics to support decision makers. Recent research in social cybersecurity has enabled new tools to support research methodologies and metrics-based decision making for communicators. The following case studies highlight research findings made possible with these new tools.

Case Study 1: Information Warfare in Social Media

In Ukraine, there was a group of young men sending out provocative images of women. The men did not know each other—they were just posting images they liked. Bots were used in an influence campaign to send out tweets mentioning each other and many of these young men at once. This led the men to learn of others who, like them, were sending out these images. They formed an online group—a topic group. Once formed, the bots now and then tweeted information about where to get guns, ammunition, and how to get involved in the Russian separationist effort in Crimea during the Euromaidan protests in 2013. Why did this work?

The cyber landscape is populated by topic groups—assemblages of actors all communicating with each other about a topic of interest. Each actor can be in many topic groups. Actors can be people, bots, cyborgs (a human with bot assistance), trolls (a person seeking to disrupt), a corporation or government account, and so forth. Members of a topic group are connected by the fact they interact with each other. Some actors will be opinion leaders, some will have a disproportionate ability to get messages to the

Shutterstock photo by Sergey Nivens

community (superspreaders), some will be highly involved in the mutual give and take of an ongoing discussion, and some will be just lurking on the sidelines. The members of topic groups are also loosely connected because they are sending or receiving messages about the same topics. Topic groups range in size and how they are organized, and some actors will be more actively engaged and send more messages. With new tools and research methodologies that measure communication impacts through social media, it is now possible to measure and visualize data to demonstrate how topic groups that become overly connected become echo chambers.

For Ukraine, bots were used to send communications that introduced these young men to each other through the use of social media mentions (e.g., the use of @mention and then the name of a person in a tweet). These bots also sent provocative images. The young men then began to follow each other, forming a topic group. In Ukraine, social media influencers created/controlled the bots that conducted a "build" campaign to misinform (i.e., engendering social connections between the young men by mentioning them together). These influencers are actors in social media who have a disproportionate ability to influence a group based on their network position. These actors appeared to be working in Russia's interests and against Ukraine. At the same time, these influencers conducted an "enhance" campaign by rebroadcasting some images and pointing to others, and an "excite" campaign with new positive language. Once the group was established, a "distort" campaign appeared bringing in information relative to the 2014 revolution.

Case Study 2: Increasing Communicative Reach in Social Media

Syrian expats and sympathizers with ISIS were engaged in social media conversations. This included listening to the prayers and spiritual guidance of a prominent Imam. A group of actors infiltrated this group and redirected attention to a site collecting money for the children of Syria. How was this done?

In social media, your followers may not receive your messages, or your messages may not be prioritized so they appear prominently to those interested in your messages. Social media platforms use your social network position (how you are connected to others), and the content of your message, to decide to whom to recommend your message, when, and in what order to receive it. Who you mention in posts, which hashtags you use, whether you use memes or link to YouTube videos, the frequency with which you post, the number of others who follow you or like your posts—all of these affect whether your message is prioritized.

In this Syrian ex-pat community, influencers created/controlled a social influence bot, the Firibi gnome bot, which was used to conduct a sophisticated influence campaign. Multiple copies of this bot were released that proceeded to send messages mentioning each other—thus engaging in a "build" campaign to misinform. The result was a topic group of bots—which meant that messages from any one bot would be recommended to others interested in similar topics. These bots then started following retweeting messages from an imam, who may not have been aware of this activity. This boosted the social influence of the imam, and engaged the bot with the community. Since the imam was a superspreader, this also meant that messages from the Firibi gnome would be prioritized to the imam's followers. The Firibi gnome bot then engaged in an "enhance" campaign and started sending messages recommending the charity website. This message was then prioritized.

An Alternative Research Approach

We have developed a methodology to measure more adequately the impact of social media communication research, planning, and objectives, called the BEND framework. This framework is a set of methods and tools for looking at who engaged in what communications, directed at whom, and with what impact. It is referred to as BEND as it characterizes communication objectives into 16 objectives, such that eight are aimed at shaping the social networks of who is communicating with whom and eight are aimed at shaping the narrative. For the social network, there are four positive objectives (the four Bs) and four negative objectives (the four Ns). Similarly, for shaping the narrative there are four positive objectives (the four Es) and the four traditional negative objectives (the four Ds). These are described in Table 1.

We developed the BEND framework to describe the fundamentals of "play" in online discourses. The goal of this playbook is to provide a standardized methodology for measuring effective communications in social media. The framework is the product of five years of research on disinformation and other forms of communication-based influence campaigns, as well as on the communication objectives of various adversarial communities. BEND addresses the well-documented Russian communication tactics, and is based on research from the commercial

Table 1: BEND Communication Objectives

	Manipulating the Narrative		Manipulating the Social Network	
POSITIVE	**Engage**	Messages that bring up a related but relevant topic.	**Back**	Actions that increase the importance of the opinion leader or create a new opinion leader.
POSITIVE	**Explain**	Messages that provide details on or elaborate the topic.	**Build**	Actions that create a group or the appearance of a group.
POSITIVE	**Excite**	Messages that elicit a positive emotion such as joy or excitement.	**Bridge**	Actions that build a connection between two or more groups.
POSITIVE	**Enhance**	Messages that encourage the topic group to continue with the topic.	**Boost**	Actions that grow the size of the group or make it appear that it has grown.
NEGATIVE	**Dismiss**	Messages about why the topic is not important.	**Neutralize**	Actions decrease the importance of the opinion leader.
NEGATIVE	**Distort**	Messages that alter the main message of the topic.	**Nuke**	Actions that lead to a group being dismantled or broken up.
NEGATIVE	**Dismay**	Messages that elicit a negative emotion such as sadness or anger.	**Narrow**	Actions that lead to a group becoming sequestered from other groups or marginalized.
NEGATIVE	**Distract**	Discussion about a totally different topic and irrelevant.	**Neglect**	Actions that reduce the size of the group or make it appear that the group has grown smaller.

sector, studies of algorithms, and dozens of case studies of communication campaigns across many platforms, in many countries, since 2014. Early evidence suggests that excite, enhance, dismay, and distort are common communication objectives used to spread disinformation.

Associated with the BEND framework is a series of measures and indicators for each of the objectives; these have been operationalized and made part of the ORA-PRO social media tools. Theae BEND measures and indicators, as built into ORA-PRO, also have been tested on Twitter data, and were used in assessing data during the Baltic Operations and Trident Juncture exercises. We find that in many cases complex influence campaigns involve using multiple BEND objectives as was described in the two case studies.

Using Social Network Analysis and Artificial Intelligence

One of the key tools in social cybersecurity is high-dimensional dynamic social network analysis, which is the analysis of who interacts with whom. Network techniques have long been used in intelligence for identifying groups and tracking adversarial actors and by marketers for identifying key informants and opinion leaders. With social media such techniques have been expanded to enable scalable solutions for massive data that take into account multiple types of relations among actors as well as relations among resources, ideas, and so forth. Today, such high-dimensional dynamic network techniques underlie social media analysis.

This analysis makes use of two interaction networks—such as who likes or retweets whom and who shares what message content with whom. The techniques to identify these interaction networks are embedded in ORA-PRO and are used for identifying topic groups and the influential actors within these groups; the depth of this data is not possible with other off-the-shelf analysis tools. Running social network techniques on social media provides indicators that then can be used in machine learning tools to identify actors and messages of interest such as bots, cyborgs, and trolls.

Artificial intelligence (AI) techniques, particularly machine learning (ML) and natural language processing techniques, are important tools in social cybersecurity. Many point to AI and ML as force multipliers in dealing with the vast quantity of digital data available today. Such technologies are of value, but they are not the panacea envisioned. The problems faced by the military in social cyberwar are continually changing and often occur only once, so new techniques for responding are needed continuously. In addition, current AI and ML techniques often are focused on easily measured data rather than the more volatile sociopolitical context.

Language technologies are used for translation, sentiment, and stance detection. Most sentiment tools simply inform readers if a message containing a word of interest is positive or negative, which often has no relation to the sentiment about the word of interest. We find that as much as 50 percent of the time the sentiment toward the word of interest is the opposite of the sentiment of the message as a whole. In contrast, the NetMapper system used with the BEND framework identifies the sentiment about the word of interest, and measures a set of subconscious cues in the message to assess the sender's emotional state.

Machine learning techniques are used to identify bots, false statements, and message on particular topics. An example is BotHunter, which can identify the likelihood that potential actors are bots. This indeed can support analysis and help communicators understand adversaries' communication objectives. The tools that are based on "supervised" learning, however, have a limited shelf life. They require large training sets, which need to be created by humans tediously coding messages and their senders into categories required for the AI tool.

Today, bots are evolving faster than the tools to find them in large part because it takes too long to create training sets. Training sets also are often biased—sentiment training sets, for example, tend to be biased toward lower-middle-class ways of expressing sentiment in English. The AI tools themselves give probability scores and no explanation on why they reached the conclusion they did. Bot detection tools often disagree because the tools were "trained" differently—leaving the ultimate decision in the hands of the analyst. These factors reduce how long these technologies will be useful and in what contexts. Today's technology advances are being made in developing AI techniques that do not require massive training sets and that provide explanations—BotRecommender is such a tool.

There are many types of disinformation (shown in Table 2). Fact-checking tools using humans or human-AI teams are providing valuable guidance, but so far these tools take too long to determine if a story contains an inaccuracy. Assessing intent is difficult: were senders intentionally trying to deceive (disinformation), or were they just mistaken (misinformation)? Many disinformation campaigns are not based on inaccurate facts, but on innuendo, flights of illogic, reasoning from data taken out of context, and so on. Many times, stories labeled as disinformation are simply alternative interpretations of facts, so AI only helps for some situations. It is less useful the more distinctive the storyline, and the faster the story spreads.

AI techniques are only useful as part of the toolkit. AI can support classifying messages by using BEND objectives. The BEND framework and associated tools, some of which employ AI, can be used to assess how communications are spreading and measure the effect. For example, MemeHunter was used to identify an influence campaign from Russia using a dismay objective that implied that compared to Russia, NATO was weak because the heads of many countries defense establishments were women, rather than supposedly strong male military leaders. This meme was spread by bots and humans alike.

Disinformation Types	Example	Potential for AI Techniques to Detect
Fake news (story made to look like news)	Navy destroyer crash in Hurricane Harvey.	AI could be used to identify sites, and do fact checking.
Fabrication with visual	Parkland student ripping up Constitution.	AI could be used to create and identify fake images.
Fabrication without visual	Opposition peso scam in Philippines.	AI might be of some assistance in finding all instances of story.
Propaganda	Duterte's helicopter scaring off the Chinese.	AI could help classify underlying BEND objectives.
Conspiracy	Pizzagate.	AI could be used to do fact checking.
Misleading—due to misquoting	Captain Marvel-Brie Larson is a racist/sexist.	AI could be used to do fact checking and stance checking.
Misleading—due to being out of context	Voting makes you lose your hunting license.	AI might provide support tools.
Innuendo and illogic	Antivax campaign.	AI might provide some support but won't solve.

The Way Forward

New technologies are needed, as well as new frameworks and procedures that can assist with analyzing the dynamic environment of social media analytics. These technologies need to be light, scalable, and interoperable, and they have to move beyond monitoring activity on the internet to engaging and countering attacks. Universities and small companies are advancing new usable solutions that need to get into the hands of warfighters, and they need improved and more streamlined acquisition procedures to do this more effectively.

In the social cyberspace, adversaries already are manipulating narratives, networks, and using media tools such as bots and memes. ORA-PRO provides a set of indicators for assessing social media and identifying which of these communication objectives are being used in accordance with the BEND framework. Companion tools that are interoperable with ORA-PRO, such as BotHunter and MemeHunter, support assessment of how the maneuver is being conducted and by whom. Such social cybersecurity technologies support the information dominance goals of the US Navy.

References

[1]National Academies of Sciences, Engineering, and Medicine. *A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis* (Washington, DC: The National Academies Press, 2019).

About the author:

Dr. Carley is a professor of societal computing in the School of Computer Science at Carnegie Mellon University.

Shutterstock photo by photo.ua

ANTICIPATING RUSSIAN SHENANIGANS:

Moving Toward Predictive Analysis

Dr. Steven R. Corman and Dr. Scott W. Ruston

EVEN AS RUSSIAN MILITARY FORCES WERE ENTERING CRIMEA ON THE GROUND DURING THEIR INVASION IN 2014, ANOTHER CAMPAIGN WAS TAKING PLACE ONLINE. NEW TOOLS MAY HELP PREDICT WHEN FUTURE INFORMATION OPERATIONS LEAD TO REAL-WORLD CONSEQUENCES.

This article is about "framing": using words, phrases, images, and other rhetorical devices to promote one interpretation of a fact (or set of facts) and discourage other interpretations. For example, here is a fact: If you die with more than $5 million in assets ($10 million for married couples) you must pay a tax on the portion of your assets over that amount. In federal law, this is called an "estate tax" because an estate is legally defined as assets left behind at death. But "estate" has other connotations too, defined by synonyms such as castle, manor, mansion, and palace. Thus, using the word estate frames the tax (to people other than lawyers and accountants) as something that applies only to a small number of very wealthy people who tend to live in mansions.

Another side of the American political spectrum does not like this tax. In 1993, the 60 Plus Association, a political advocacy group, began calling it a "death tax." The phrase caught on, was included in the Republican Party's "Contract with America" in 1994, and is still used today. This frames the tax in a different way—as an example of government overreach by taxing some people even after they are in their graves. Use of the phrase by people to accomplish a purpose (i.e., repeal of the tax) makes it an example of strategic framing, the goal of which is to frame a topic in a way that gets a desired response from a target audience.

Framing the Message

Strategic framing is routinely used in some communication planning efforts, especially in a military context when communicating the "commander's intent" of an operation to the general public and the media. In 2016, as the European Phased Adaptive Approach (a program to provide missile defense systems in Europe) proceeded with land-based ballistic missile defense installations established in Romania and Poland, US officials consistently described the capability as part of a commitment to collective defense, particularly against potential rogue threats such as Iran. NATO officials echoed this narrative. Conversely, the Russian perspective, as evidenced by Russian-owned media coverage, held that NATO ballistic missile defense capabilities upend the balance of power in the region.

Russian propaganda outlets such as Sputnik News and RIA Novosti frequently use strategic messaging to foment wedge issues in public discourse. In Latvia, for example, there is an annual event commemorating the Latvian Legion, a formation of the German Waffen-SS, which fought against the Soviet Army at the end of World War II. Russian news outlets focus coverage on this commemoration and frame it as proof of the rising tide of fascism in Europe.

For example, one media outlet recently reported that, "the glorification of the Nazi criminals in Latvia is not only being met with no opposition among the authorities but is very often carried out with their assistance." This type of framing is used to split the Latvian citizenry down ethnic lines (there are many ethnic Russians living in Latvia) and generates sympathy for Russia when the reality of public opinion within Latvia is neutral.

Framing an Invasion

The Russians also used messaging techniques to great effect in their 2014 invasion of Crimea. Analysts and observers, including then-Joint Chiefs Chairman Gen. Martin Dempsey, noted that the Russians have a "playbook" that involves stoking outrage among ethnic Russians in a targeted area and encouraging them to protest. Next, they criticize treatment of the protestors by the adversary government, enhancing unrest and furthering maltreatment, and then invade on the pretext of protecting the abused Russians.

In 2017, the Office of Naval Research (ONR) funded a study to examine whether this playbook was implemented during the Crimea invasion. The study proved that Russian propaganda created outrage, and used both social media and traditional media to spread discord. Researchers partnering with ONR collected about 30,000 news stories from known Russian propaganda and pro-Russian news outlets. The effort also trained a machine classifier to identify instances of five different narratives associated with the Russian playbook. Each of these categorized narratives were based on certain claims and events:

- **Fascist vs. antifascist struggle:** A government and/or society is fascist and seeks to oppress the antifascist Russian minority.
- **Discrimination against Russian minorities:** Russian minorities are marginalized, abused, and/or denied their human rights.
- **Assault on Soviet history:** The targeted government/society subverts, suppresses, or revises Soviet history and accomplishments.
- **Criticism of government and politics:** The targeted government and political system are corrupt and ineffective.
- **Invasion of Crimea:** An invasion of the Crimean Peninsula is just/necessary.

In terms of process, the classifier measured the density of each of these narratives daily going back to 2010, to establish the baseline of "normal" narratives, and to 2017, to establish the baseline of the new normal following the 2014 invasion. The study found there was considerable day-to-day variation in the distribution of messaging across the five categories.

To uncover evidence that this variation was strategic we looked at shifts in messaging using a measure of divergence, which is the measurement of the degree of change across the variables (in this case the five frames) over time. For example, if on one day the framing was distributed equally across the five categories, and the next day the framing was very lop-sided (80-5-5-5-5 percent), that would be a very large divergence. If the next-day's distribution showed only small changes (say 22-18-20-22-18 percent) that would be a very small divergence. Applied in this way, the divergence measure represents a messaging signal in the information environment. Assuming an information operation is under way, large divergences are likely because of changes in messaging by the Russians to accomplish their strategic goal of behavior change (e.g., support for their military operation to recapture Crimea).

According to the data, there are relatively low levels of divergence, except from September 2013 to September 2014. The largest spike begins and grows around the time of the pro-EU Euromaidan protests and peaks in late January 2014, just before the ouster of Ukrainian president Viktor Yanukovych on 23 February 2014 and the incursion of "little green men" and the invasion of Crimea in the following week. The second spike began during late March when Crimea declared independence, and peaked during mid-May, when there were votes for unification with Russia. This example shows that tracking messaging divergence in Russian propaganda sources aligns with plausible changes in messaging in advance of important periods in the conflict. Divergence starts to accelerate months before the secession of Crimea and its annexation by Russia.

This begs the question could this technique be used to detect future cases when the Russians are "softening up" a country in advance of a possible invasion? ONR's academic partners believe the answer to this question is yes, which is why ONR is funding further research to improve detection techniques and test them in other scenarios. Given the growing importance of aligned communication efforts with US partners to support kinetic action, we may be able to predict communication tactics of other militaries and shape the information battlespace.

About the authors:

Dr. Corman is the director of the Center for Strategic Communication at Arizona State University.

Dr. Ruston is a research scientist with the Global Security Initiative at Arizona State University

NATO photo by Chief Petty Officer Brian Djurslev

Technology for Communicators:

THE BITE DASHBOARD

By Chris Kurcz

BALTIC OPERATIONS 2019—AN ANNUAL NATO EXERCISE INVOLVING FORCES FROM 18 NATIONS—WAS A PERFECT OPPORTUNITY TO TEST THE BOT IDENTIFICATION AND THREAT EVALUATION (BITE) DASHBOARD, A NEW TOOL FOR EXAMINING THE INFORMATION BATTLEFIELD IN REAL TIME.

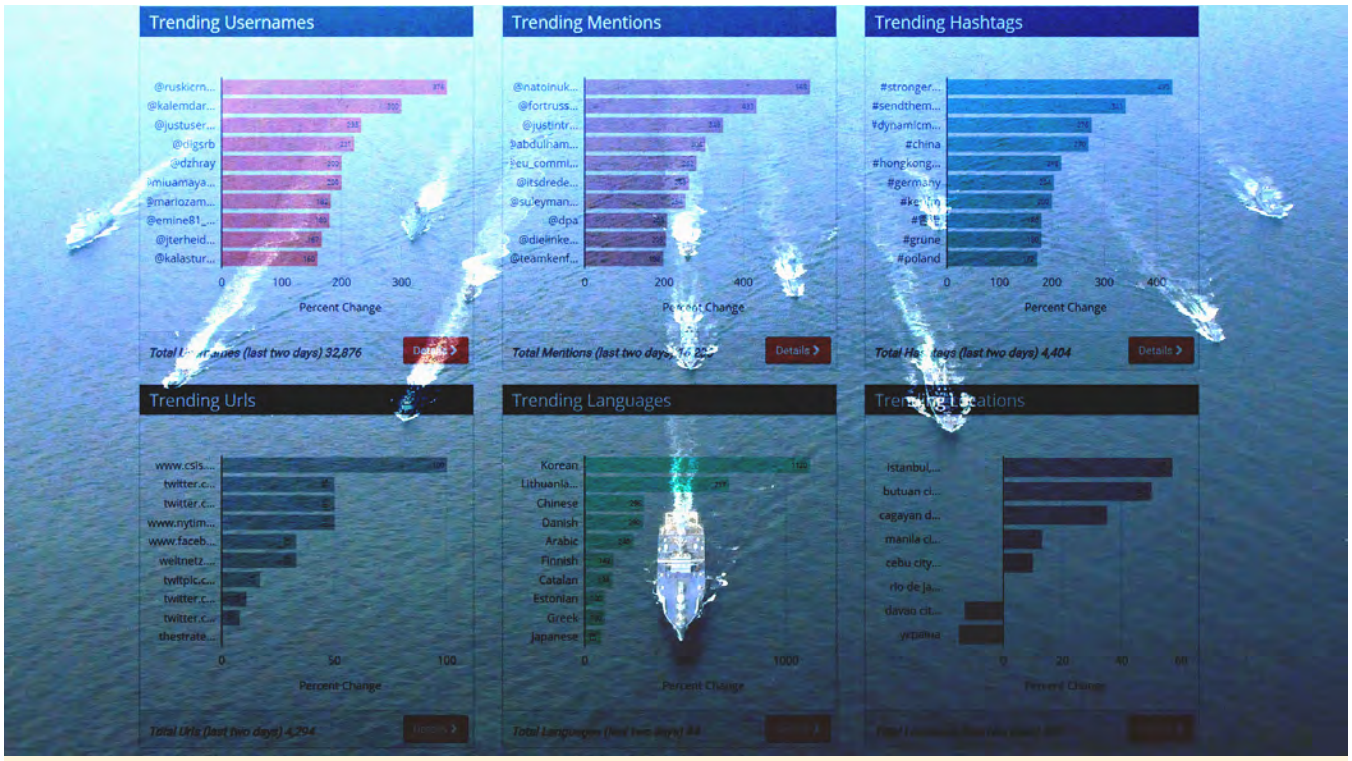
Analysis of social media as well as publicly available information more broadly has become relevant to the Department of Defense DoD, as it provides a means to gauge messaging from adversaries who use it to influence views and behaviors of targeted groups. In addition, publicly available information provides valuable situational awareness that is vital for many types of operations, including humanitarian or disaster relief. Publicly available information also plays an important role in engaging the extended DoD community and provides a conduit to spread the values of the organization.

One of DoD's challenges in the information environment is identifying, tracking, and performing analysis of manipulated discourses on social media. The Office of Naval Research (ONR) has funded research and development of capabilities that identify bot-fueled artificial amplification, echo chambers, group polarization, social hysteria propagation, rumor dispersion, and propaganda. These research efforts face many hurdles, including the sheer volume and velocity of the data streams, the ubiquity of noise, and the sophisticated information strategies employed by adversaries. Successful approaches must provide state-of-the-art analysis capabilities in network science and natural language/multimedia processing as well as the means to navigate and explore the analytic results and raw data intuitively.

This article describes the Bot Identification and Threat Evaluation (BITE) dashboard, one of the ongoing research efforts funded by ONR to address these challenges. The BITE dashboard monitors social media by collecting data, in real time, based on user-specified search queries. The data is collected, aggregated, and displayed to show entities (such as hashtags) that are changing most rapidly (i.e., trending). Views of this data are highly dynamic and are simple to adjust by applying filters specified by a robust query language and are designed to facilitate exploration. Similar to traditional sensor systems that use tipping and queueing principles for data collection and processing, analysts and operators can use trending entities as a tip to focus subsequent analysis or discover other entities of interest. Toward this end, the BITE dashboard provides export capabilities to other tools that perform complementary analysis.

The annual NATO Baltic Operations (BALTOPS) is an international maritime-centric exercise that takes place in the Baltic Sea. BALTOPS 2019 was led by the US 2nd Fleet, and included forces from 18 nations. In past NATO exercises such as Trident Juncture 2018, social media monitoring has shown that regimes opposed to NATO have used these exercises to disseminate anti-NATO narratives through artificial dissemination networks.

Analysts and operators addressed social media during BALTOPS from two perspectives: information operations and public affairs. During the exercise, analysts used the BITE dashboard to find abnormalities or spikes in trends and then applied other tools for deeper analysis or collected data from additional sources. In particular, analysts and operators could monitor social media



Graphic courtesy of author

This chart shows some of the ways in which real-time social media data from the BITE Dashboard can be displayed.

to determine the use of coordinated distort, dismiss, dismay, and/or distract tactics. To support analysis of the information environment during BALTOPS, several DoD operators and analysts were trained on the use of the BITE dashboard, which added to the arsenal of tools available.

Several social media monitors were established to collect data representing different regimes of the information environment. One such monitor collected social media posts for particular Russian keywords. Analysts used the exploratory capabilities of the BITE dashboard to discover suspicious accounts that were spreading propaganda.

Relevant data was exported to other tools capable of conducting a more detailed and in-depth analysis using advanced network and content-based methods. This analysis uncovered a targeted campaign to disseminate anti-NATO narratives posted in other venues using a sophisticated network. For example, content was disseminated that claimed the NATO exercise was a guise for the United States to invade Kaliningrad (the small Russian enclave between Poland and Lithuania). Discovering this connection led to further analysis and data collection to understand these information operations more fully.

To support public affairs efforts, analysts and operators analyzed NATO, US Navy, and other friendly accounts. The intent was to understand the effectiveness of the blue force's use of the information environment. For example, Navy Live, the official blog site of the US Navy, produced several articles during BALTOPS that emphasized the cooperation between NATO allies and described World War II-era mine-clearing exercises and other information about the US Navy's involvement during the exercise. Public affairs officers used social media to spread this content throughout the naval community and beyond. The BITE dashboard monitored this activity and enabled analysis that focused on understanding how the content flowed through naval communities, whose accounts are most influential in enabling the spread of information and detecting adversarial tactics employed in response to this content.

Analysis of publicly available information associated with NATO exercises serves many valuable purposes. One benefit is that NATO and its participating nations receive a richer understanding of the geopolitical landscape and climate in real time. Perhaps more significant, however, is the opportunity to improve understanding of, and to maturate skills in, the information space. As the world becomes more connected, critical intelligence for successful operations can be found more consistently through digital channels. Analysis of the publicly available information around exercises such as BALTOPS allows NATO forces to gain a better understanding of adversarial actors in the digital information environment.

About the author:

Chris Kurcz works for the MITRE Corporation and is a published authority on computing methodologies.

Photo by MC3 Joe J. Cardona Gonzalez

NATO INTEGRATES NEW MEDIA, AND SO DO ADVERSARIES

By Dr. Katrin Galeano, Lt. Col. Rick Galeano, and Dr. Nitin Agarwal

THE NATO EXERCISE TRIDENT JUNCTURE 2018 PROVIDED AN EXCELLENT TESTING GROUND FOR A NEW TOOL THAT LOOKS AT THE EFFECTS OF VIDEO POSTS—STILL ONE OF THE TRICKIER PRODUCTS OF SOCIAL MEDIA TO ANALYZE.

In December 2019, Stanford University released their annual artificial intelligence report. The report indicated the rapid pace of technology since 2012.[1] With this comes the need for advanced software to monitor and analyze big data with a fiscally affordable platform. This case study demonstrates the effective use of such platforms in the information domain.

Research by the RAND Corporation concluded that in order to review large amounts of data in enormous networks, "big data tools" are needed to focus on investigation efforts.[2] Our effort undertaken here consequently used a number of applications to support research of this magnitude for data collection, such as YouTubeTracker (a new application developed by a member of our team) and ORA for data analysis.[3]

This article will articulate how the platforms identified three key actors that commented heavily on anti-NATO videos, raising their rankings. These comments falsely portrayed NATO as an aggressor, undermined the credibility of the alliance, and questioned its abilities.

The Information Environment

As preparations for Trident Juncture 2018, the largest NATO exercise held in Norway since the 1980s, were in full swing on land, in the air, and on sea, another domain gained increasing importance: the information environment. Whether the intent is to manipulate narratives or insert new narratives, the actors (also called nodes for this research) add content to the information environment at an overwhelming rate. They are not necessarily individuals—they can be a group, a corporation, a news outlet, a government organization, or bots and cyborgs.

Actors use a wide variety of platforms to spread their content. These platforms are not limited to the most actively used global social media platforms such as Facebook, YouTube, WhatsApp, Instagram, and Twitter, but also include niche user-content-generated outlets (e.g., the Russian networking site VKontake and Reddit) as well as online news media and blog sites. While the latter two are used to provide an audience with detailed information, a teaser in the form of a catchy, attention-grabbing headline is spread throughout social media featuring a link to the website reaching a broader audience. Both credible news media and fake media outlets alike use this tactic, as they share the common goal of reaching the widest audience possible.

Content is always busy and overwhelming in these media and often crosses over from one platform to another; it is diluted and blended to form the stories that are told on social platforms. Interpreting that data is an ever-evolving process. Twitter has become the gold standard for many data scientists to mine and explore digital messaging. Information maneuvers on Twitter have evolved rapidly. Content engagement through the use of tags and hashtags supports sorting of content easily; simultaneously, software development has followed suit. Platforms such as YouTube are new to data scientists because video is not analyzed effortlessly with software, and it can be manipulated easily. High view counts on videos result in revenue for click farms (view-selling sites) and public opinion being misled on the actual information. For example, the view counter on YouTube videos shows how many times a video has been seen. The more views and the longer retention rate a video has, the higher the engagement. Higher engagement scores push a particular story up in results during a normal Google search because of the inherent biases of search engine optimization algorithms. This is where the software application YouTubeTracker would benefit from commercialization alongside industry partners by parsing the big data into decipherable information or even actionable intelligence from YouTube data.

As expected, information confrontation was notable in several areas of the information environment during Trident Juncture, ranging from alleged GPS electromagnetic interference to social media manipulation. NATO digital natives posted on multiple social media outlets, while the digital immigrants watched with curiosity. Disinformation was corrected by official NATO channels on select social media platforms, such as Twitter, but adversaries encroached on one global platform: YouTube.

Information Actors on YouTube

YouTube is the largest storytelling platform that incorporates videos from across the globe, allowing for freedom of expression, information, and opportunity, as well as the "freedom to belong," according to the website itself.[4] With YouTube being a key player in the overall online realm of social communications, it is also the most relevant video sharing platform globally. Each day, more than one billion users watch

NATO spokesperson Oana Lungescu replying to a tweet published by the Embassy of the Russian Federation in London.

more than a billion hours of content. The top countries for YouTube usage are the United States, India, and Russia.[5]

The Collaboratorium for Social Media and Online Behavioral Studies at the University of Arkansas at Little Rock collected the YouTube data with the help of the YouTube Application Programming Interface (API) in order to evaluate how NATO's communications efforts through official channels performed compared to adversarial actors. To assess Trident Juncture's information environment, content was divided into three groups by the strategic communications cell at NATO headquarters:

- **Owned communication**: NATO official accounts and channels from NATO delegations (7 percent)
- **Earned communication**: What everyone is saying, how audiences are reacting (64 percent)
- **Hostile communication**: The activities and communications of anti-NATO information actors (29 percent).

Comparing NATO-owned, -earned, and -hostile content related to Trident Juncture, we found that hostile content outperformed NATO-owned and -earned content. Hostile videos received higher user engagement (views, comments, etc.) on average than NATO-owned or -earned videos. NATO-owned and -earned videos had entirely organic engagement, but hostile videos exhibited strong indications of inorganic, or robotic, activities.

Seven percent of all Trident Juncture-related videos were published by NATO, its operational headquarters, and participating militaries. The majority, however, consisted of coverage provided by news outlets, military enthusiasts, locals, and hostile actors. Channels targeting the military enthusiast community used footage from the Digital Video Information Distribution System (DVIDS) for their videos. The use of catchy, attention-grabbing titles such as "This is How U.S. Marines Will Take the Fight to Russia in the Arctic" sparked the curiosity of the audience. Many viewers engaged with the content by liking, disliking, sharing, and/or commenting. To achieve high engagement, little effort and expense was required for these information actors to have an effect.

Information Actors and Their Tactics

The goal of many channels publishing Trident Juncture content was to generate revenue from advertisements. Actors frequently used deception and smoke screening tactics. Numerous channels used the words "news," "military," and "defense." This, in combination with channel verification, caused confusion among viewers. The recipe followed is a simple one: First, download videos and their description from DVIDS. Next, add your own branding, and then upload it to YouTube to appeal to the military enthusiast community. The videos are then monetized, which means the uploader earns money for the advertisement that is played before or during the videos.

Throughout the exercise, exaggeration and hyperbole were evident. Videos focused on enhancing friction with Russia, undermining NATO's reputation and credibility, and questioning its capabilities while painting a picture for the public that World War III was imminent. Distraction tactics were also used. The world was informed, for instance, that the US Marines had caused a beer shortage in Iceland. The most watched video, however, was the collision of the Norwegian frigate KNM *Helge Ingstad* with an oil tanker toward the end of the exercise.

"BREAKING! Norwegian navy frigate-collides with oil tanker in fjord" was published by Weapons of the World on 8 November 2018. It received more than 330,000 views, which amounts to more than three times their subscribers, and quickly attracted more than 1,000 comments.

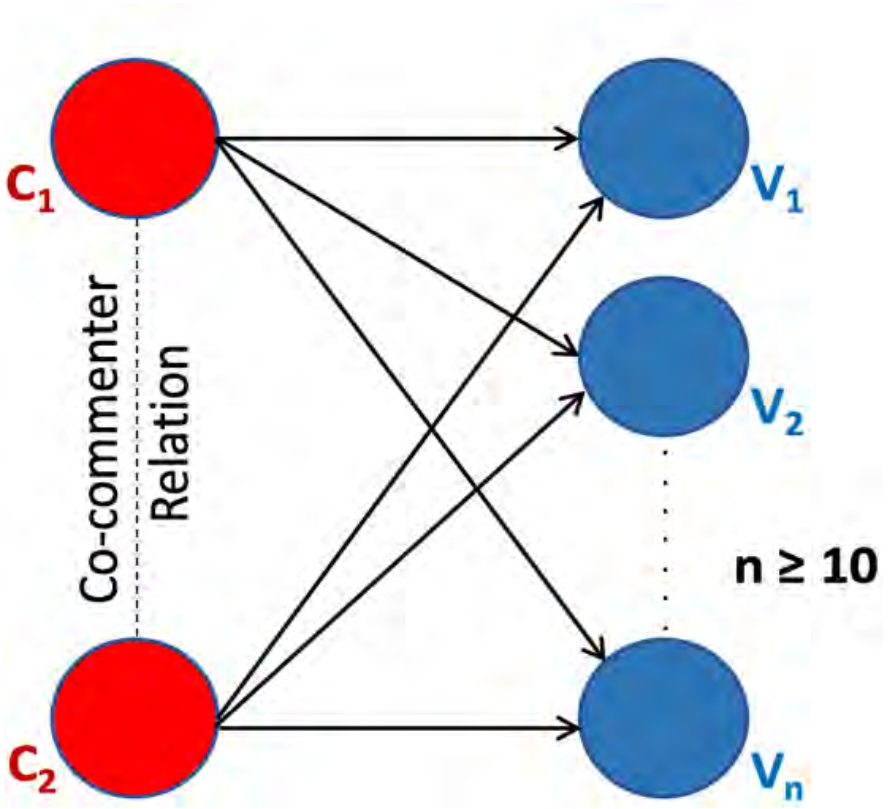
Within hours of the incident, the video "Spoofing Attack—Vlad Putin jamming the GPS of NATO ships, HNoMS Helge Ingstad (F313)" published by Servitutt was uploaded. Even though the video did not gain much friction initially, the story idea did, and hostile information actors used this event to deploy their tactic of defamation.

The Narratives

The combined efforts of Russia Today (RT), Sputnik, and Ruptly in Russian, English, Arabic, French, and German reported on the exercise and local protests. The target audiences in Russia and beyond were fed the false narrative that NATO was an aggressor and that Russia's technology was superior, as demonstrated by Tupolev 160 bomber flights and a missile launch. The technology demonstrations shifted the attention away from the exercise and offered anti-NATO information actors the opportunity to refocus on Russian activities while undermining NATO's reputation. Orchestrated events in the guise of talk shows, similar to authentic and neutral television shows, were produced with the intent of influencing viewers. We observed several talk shows in English, Russian, and German using video footage from RT and other Russian information actors to create tension and, potentially, to divide the public within NATO nations. Others used hate speech and ridicule to question NATO's capabilities.

Influencing public perception

Hostile content was uploaded quickly, but NATO's own content was added much more slowly, and with less attention-grabbing headlines, resulting in low viewership and engagement. The hostile comments often mirrored the anti-NATO content published by these channels. A commenter may want to amplify or distort the narrative, create a toxic/trolling environment for other viewers, or exploit YouTube's recommendation algorithms by flashflooding/flashmobbing the comment space—a behavior that misleads YouTube's algorithms into believing that the video is going viral and therefore elevating its rank. Many hostile channels interacted with their viewers by asking open-ended questions, replying to comments, and liking viewers' comments, which pushed up the ranking.



Graphic courtesy of authors

Commenters (c) would have to comment on the same videos (v) at least ten times in order to have a connection.

Comment Analysis

In order to analyze the complex comments section of YouTube, we created a simple algorithmic model to identify the relationships among commenters. For the purpose of this study, commenters and co-commenters would have to comment on the same videos at least ten times to have a connection, as depicted in the figure to the left.

Data was collected using YouTube API. We also reviewed YouTube content manually on a daily basis during our data collection period that supported the channels to analyze by parsing out irrelevant data. Using this data and analyzing the co-commenter network was imperative, as we identified 35,601 users who commented on 503 videos that were published by the NATO-owned, -earned, and -hostile channels during a 34-day timeframe.[6] This data was linked when both C1 and C2 commented on at least ten videos or more together. Results

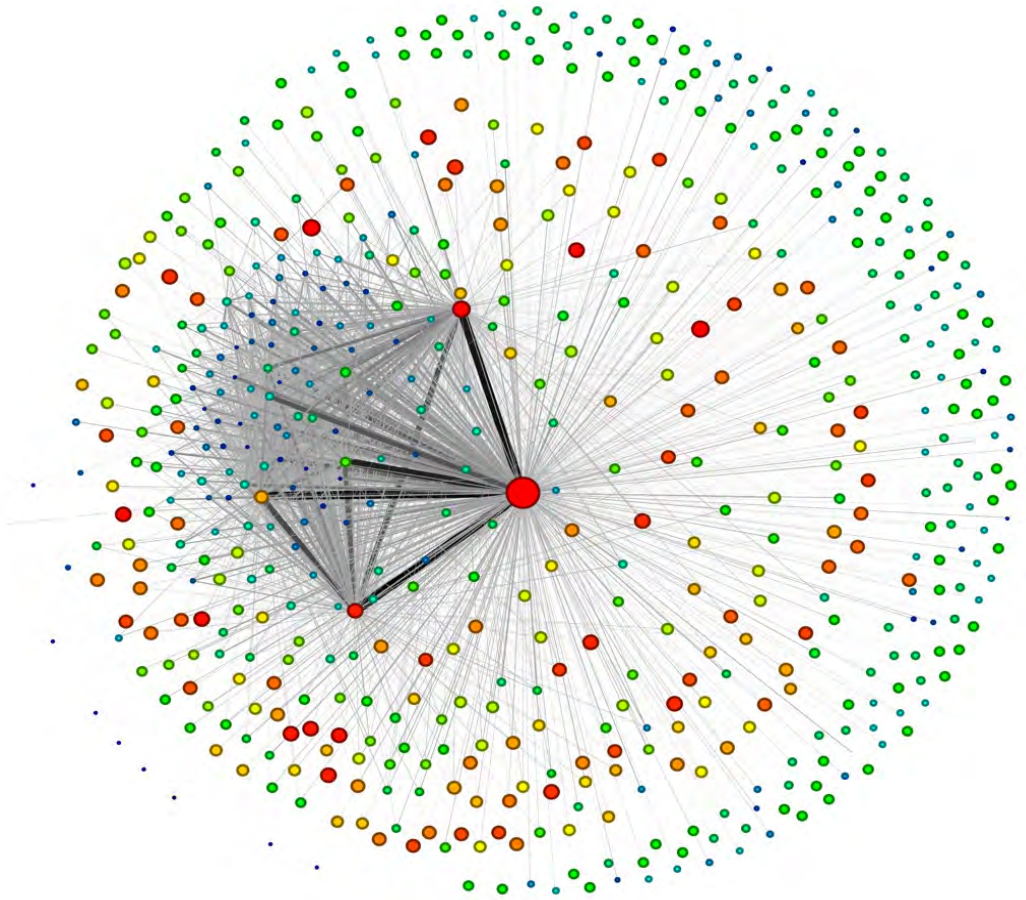
identified a mean of 14.82, which indicated that most commenters commented on 10-14 of the same videos.

These comments resulted in an extremely large data set. The commenters and co-commenters tied together more than 9,000 nodes that were connected by more than 4.4 million edges.[7] This required us to fold the network data using ORA. By narrowing it down to at least ten videos, the network was reduced to 583 nodes (which represents the commenters) connected by 5,844 edges (comments).

To gain further insight on the key information actors in the network, we performed centrality calculations. Three commenters consistently ranked in the same order for the four types of centrality measures used: total-degree, betweenness, closeness, and Eigenvector.

The co-commenter network created for this study displayed in the figure below visualizes the network's most central actors. The nodes are colored based on their closeness centrality on a hue color scale with red being the most central and blue being the least central. The node size also was adjusted based on centrality. The larger the node, the higher its centrality. The links are colored by value. The greater the number of shared videos two commenters commented on, the darker the link. In addition, we also increased the width of the links based on value. These results echoed the aforementioned betweenness centrality in which the agents remained in the same ranking order (agent 1, 2, 3).

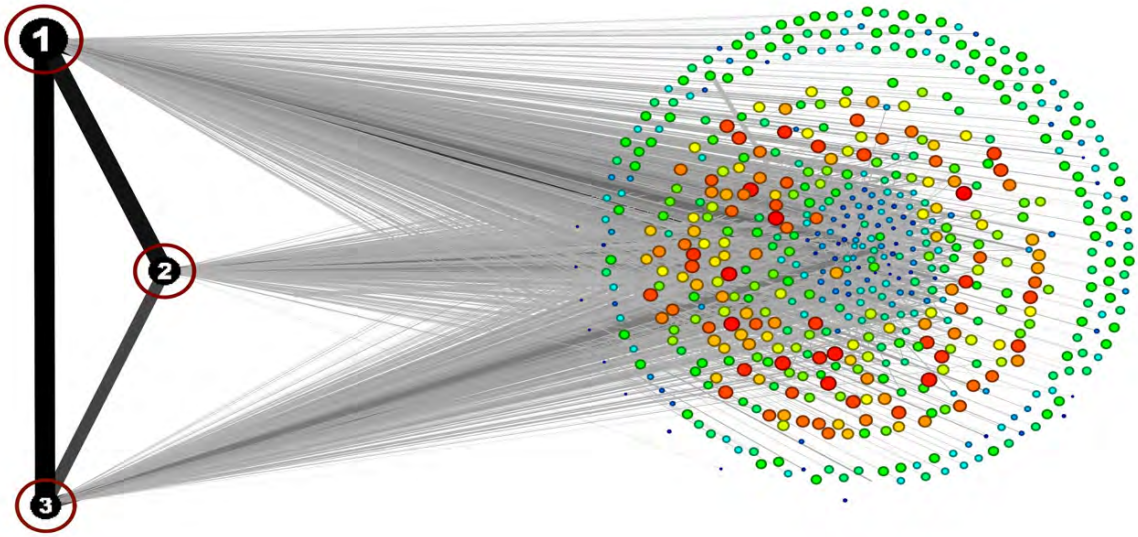
Additional study of this network allowed us to dive even deeper. We further analyzed these nodes that were most centralized and stood out and compared to the rest by looking at their sphere of influence. This network (also known as an ego network) was analyzed and showed that the top three nodes directly connected to 575 other nodes with a total of 5,834 links (see the figure above). In addition, these nodes were directly connected to 99.14 percent of all nodes within the network.



Co-Commenter network prominently displaying the node with the most links (564 edges).

Graphic courtesy of authors

Graphic courtesy of authors



An ego network of the top three commenters. The nodes are colored based on their closeness centrality on a hue color scale with red being the most central and blue being the least central. The node size was also adjusted based on centrality. The larger the node, the higher its centrality.

Results

A YouTube commenter network is the elementary network in relation to an overall feed in which comments are made under a post. By allowing for comments on posts the outreach increases and audiences can expand rapidly. Combining this tactic of commenting and co-commenting drives channels' visibility and its likelihood of being recommended by YouTube's algorithms through the roof. Identifying influential actors without the use of the YouTubeTracker for big data collection as well as the use of ORA for social network analysis presented in this article would have been like trying to find a needle in a haystack.

Conclusion

In the end, this analysis drives home dynamic questions for commanders and decision makers while operating in this information environment:

- What is the so what?
- Should I be concerned?
- Do I need to counter this?
- How can I counter this?

These questions should be answered by the strategic communications team, but recommendations are as follows:

1. Strategic communicators should incorporate social network analysis into the overall communications plan, aligning with operational effects through information activities
2. Use the targeting cycle and ensure analysis is processed early (for baseline metrics) and continues throughout any operation
3. Annotate influential actors and monitor/conduct counter operations as authorized.

Letting our guard down gives adversaries the opportunity to sow discord, weaponize narratives, or, worse yet, manipulate data by denying signals or changing the information received. Exercises in 2020 will continue to generate a plethora of data points (such as Trident Juncture 18). The big data tools that RAND Corporation referred to were presented (YouTubeTracker and ORA) as off-the-shelf solutions. These tools fall in line with the Stanford study with advances in software and are ready to be distinguished by outside organizations and/or shaped to support specific models as industry sees fit.

References

[1]R. Perrault et al, "The AI Index 2019 Annual Report," Stanford University (December 2019), https://hai.stanford.edu/ai-index/2019.
[2]E. Bodine-Baron et al, "Countering Russian Social Media Influence," Rand Corporation (2018), https://www.rand.org/pubs/research_reports/RR2740.html.
[3]YouTubeTracker, http://youtubetracker.host.ualr.edu/; ORA-LITE, http://www.casos.cs.cmu.edu/projects/ora/software.php.
[4]"About You Tube," YouTube, accessed on 4 February 2020, https://www.youtube.com/yt/about/.
[5]"You Tube for Press," You Tube, accessed on 4 February 2020, https://www.youtube.com/yt/about/press/.
[6]For a full report of the co-commenter analysis, please email Dr. Nitin Agarwal at nxagarwal@ualr.edu.
[7]An edge in SNA represents a relationship between two nodes represented by a solid line connecting them.

About the author:

Dr. Katrin, Lt. Col. Rick Galeano and Dr. Agarwal are researchers with the Collaboratorium for Social Media and Online Behavioral Studies at the University of Arkansas Little Rock.

Shutterstock photos by Eric Isselee and Christian Weber

COMBATING MISINFORMATION: AN ECOLOGICAL APPROACH

By Dr. Bryan Ek, Dr. Lucas A. Overbey, and Michael Grass

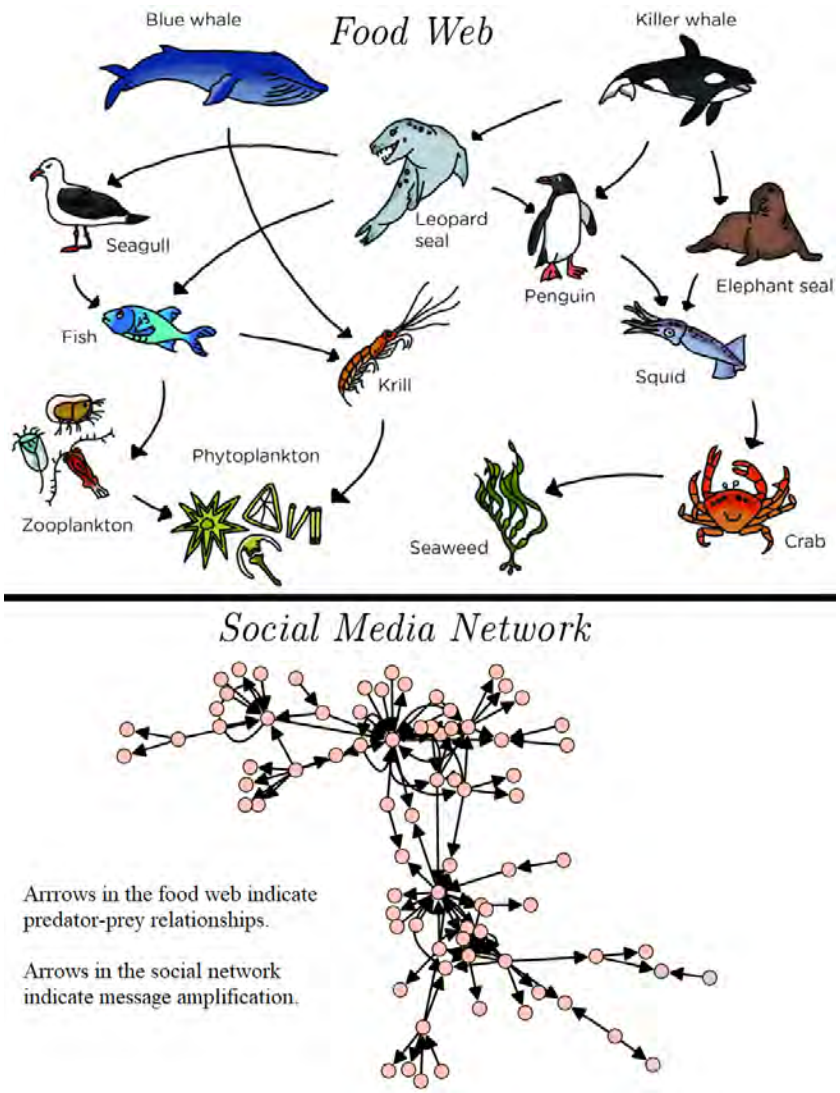
THINKING OF THE MALICIOUS MISHANDLING OF INFORMATION AS A KIND OF COMPETITION BETWEEN PREDATOR AND PREY IS ONE WAY TO VISUALIZE WHAT IS AT STAKE IN THE INFORMATION AGE.

The ubiquity of social media usage has led to an escalation of coordinated misinformation campaigns across platforms. Actors exploit the inherent capabilities of these platforms to deploy information warfare campaigns aimed at disrupting populations, propagating social hysteria, manipulating crowds, and increasing group polarization. Some adversaries are so proficient at navigating and developing new constructs within the open-source domain, they are able to achieve real-world results at a fraction of the cost of similar effects by conventional military forces. The goals of these new campaigns include the targeting of regimes, armed forces, and populations by means of influence to move them either to support aggressors or to act against their own self-interests.

On social media, these approaches often manifest themselves through coordinated activities magnified by the use of automation. One such example are bots: automated accounts programmed to act maliciously, thereby allowing for mass amplification of messages and increasing the perception of influence. The United States needs analytical methods to identify and combat internet and media operations relying on information and network maneuvers that make use of coordinated misinformation campaigns and bots.

Effective methods to identify fake, semiautomated, and automated accounts have only recently emerged, and frequently rely on common heuristics such as the ratio of tweets to retweets, posting regularities, and anonymized profile information to feed supervised machine learning techniques. These developments have led to automated bot detection algorithms that use these temporal and content-based features (and, more recently, network measures) as inputs. Supervised learning techniques are limited, however, by historical examples that have been documented. As efforts are made to diminish the effects of malicious accounts, bot characteristics quickly adapt to counter detection. Consequently, detection algorithms will need to evolve continuously. This leads to an evolutionary battle between bot creators masking their tactics and bot detectors uncovering them.

We are often interested in understanding how and why these accounts behave as they do and not just which accounts are automated or malicious. In most cases, bots attempt to spread influence or to facilitate and amplify the flow of information in a human network that would otherwise reject them. Observing bot behavior allows us to understand that for them to have this effect they must alter network connections.



Showing information-sharing as a food chain involving predators and prey helps as an effective way to demonstrate how social media networks function.

Graphic courtesy of authors

We are currently investigating novel methods that apply unsupervised network science approaches to social media to identify unusual coordination between groups of accounts rather than of individuals. In addition, obtaining a larger field of view from a coordinated group is a step toward answering the how and why. To achieve this, we are applying a set of mathematical graph theoretic concepts built around a construct from ecology called "competition graphs." Originally developed by mathematical biologist Joel E. Cohen, competition graphs are a way to analyze indirect relationships between organisms, predators and prey alike, within an ecological system. In an ecosystem, a competition graph is a mapping of a food web, represented by a directed graph, onto an undirected graph such that nodes represent predators and edges represent resource competition between predators. Thus, the indirect relationship between predators can be discovered through their similar behaviors toward like prey. Common enemy graphs are the result of considering interactions in the opposite direction; prey with common predators are grouped together.

In the social media application, online environments can be considered like an ecosystem of interacting actors, where groups are coordinating with or competing for attention of other users. The "prey" in this scenario are now users with common sources of information—i.e., the "predators." Coordinated behavior in this construct is often aimed at amplification of the source account's influence: to promote them by overstating their strength of connections to human users. Within the BEND framework of social cybersecurity introduced by David M. Beskow and Kathleen M. Carley of Carnegie Mellon University, bots with similar goals and tactics (and potentially created *en masse*), act as force multipliers to conduct information transactions at scale. This serves to increase the presence of certain information, making it more likely to be seen and spread from a purely statistical standpoint.

We are studying a variety of mathematical variants to competition graphs and related concepts to develop an analytical workflow for identifying groups and tactics of social media accounts working in coordination to spread misinformation. Using a combination of common enemy graphs, a novel edge-weighting scheme, filtering, and graph clustering, we have been able to identify groups of accounts composed largely or wholly of botnets exhibiting such coordinated behavior. The tactics of these botnets, or hybrid human/botnets, are elucidated by the component size, neighborhood size, source, and edge weights within our common enemy graph

workflow. These properties can be used as a set of features, among others, incorporated into a supervised bot-detection model; they are also valuable by themselves for analysts to be able to understand the tactics, behaviors, and goals of malicious sets of coordinating actors.

We are also researching how these structures and behaviors change over time: what happens if botnets are dormant and suddenly become active, adapt to platform suspensions, or evolve with changes to mission tactics and goals? Our common enemy graph-analytic workflow, for example, discovered the dynamic adaptation of a botnet that adjusted to the suspension of its amplified account by "promoting" a worker bot to its source account.

Although we can identify and act on individual accounts to combat the spread of disinformation or other malicious behavior in social media, botnets or hybrid networks may be robust or adaptable to individual disruptions. If communities of coordinating behavior can be identified and characterized, analysts can develop counter tactics to nullify misinformation campaigns more effectively.

In addition, we are currently studying how a similar workflow involving competition graphs and a graph metric called "boxicity" can assist with the recognition of the structure of "amplified accounts" that operate in spheres of shared prominence or within echo chambers. These spheres of shared prominence are analogous to "ecological niches" in environmental systems. These amplified accounts are frequently (though not always) human accounts and sometimes act as coordinating trolls sowing discord. Identifying spheres of shared prominence of amplified accounts can yield insight into the flow of information and connections within and between communities, provide an assessment of in-group and out-group prominence, and offer information on existing echo chambers or filter bubbles.

Further network science techniques may be useful for providing analytical intuition into where connections can be added or removed from the network to combat malicious coordinating sets of actors. Our research entails studying variants and related concepts, including competition graphs to derive logical correlations. In conjunction with natural language processing, supervised machine learning, and traditional social network analysis techniques, unsupervised graph theoretic techniques such as these are key to assessing and combating current and near-future threats within the information environment.

About the authors:

Dr. Ek is an applied mathematician working in the science and technology division at Naval Information Warfare Center Atlantic.

Dr. Overbey is the science and technology lead in data science and analytics at Naval Information Warfare Center Atlantic.

Michael Grass is a science and technology program manager at Naval Information Warfare Center Atlantic.

IDENTIFYING MISINFORMATION CAMPAIGNS

By Capt. Iain Cruickshank, USA, and Dr. Kathleen M. Carley

SEPARATING "FAKE NEWS" FROM THE REAL DEAL HAS BECOME, UNFORTUNATELY, ONE OF THE GREAT TASKS OF OUR TIME. DETERMINING WHETHER SOMETHING IS PART OF A DELIBERATE CAMPAIGN OF MISINFORMATION CAN STILL BE HARD, BUT MACHINE-LEARNING ALGORITHMS ARE MAKING IT EASIER.

Online misinformation campaigns are a growing threat to society. The use of online mediums such as blogs and social networking sites to propagate inflammatory and hostile narratives already has had profound effects on various aspects of society. These campaigns, which are often a blend of many types of information of varying veracity, have been implicated in everything from affecting national elections to inspiring hate crimes and domestic terrorism. Given all of this, the ability to counter and combat misinformation campaigns is imperative for the safety of society.

Among the great difficulties in combating disinformation campaigns is identifying and characterizing them. While many campaigns rely on fake news and fabricated facts, they also use distortions and partial truths as well. Furthermore, with the advent of machine-learning technologies that can generate fake data, the issue with trying to characterize and

Shutterstock photo by Yeti studio

identify a misinformation campaign by false information will become increasingly difficult. Thus, one cannot definitively identify misinformation campaigns by the use of detecting fake information.

Recent research has found the social-network context to be a significant factor in misinformation. Studies have shown that fake news spreads differently than genuine news in online social networks. In addition, misinformation tends to start and have its greatest influence in some online communities, such as echo chambers, but not others. Furthermore, recent research in detecting fake news has found that the inclusion of a social network into supervised learning gives a significant boost in accuracy over just using the content of the fake news by itself. Misinformation does not spread in a vacuum—it also has a significant social dimension.

To identify and characterize misinformation campaigns better we are developing new methodologies that can incorporate both content and the social-network context into one coherent model. Since misinformation campaigns often do not have labels that identify them as such, we opt for an unsupervised technique that focuses on representing the data for a human analyst. In addition, since graphs, or networks, are an extremely flexible mathematical model of any data, we use them as the models for the various modes of the online data.

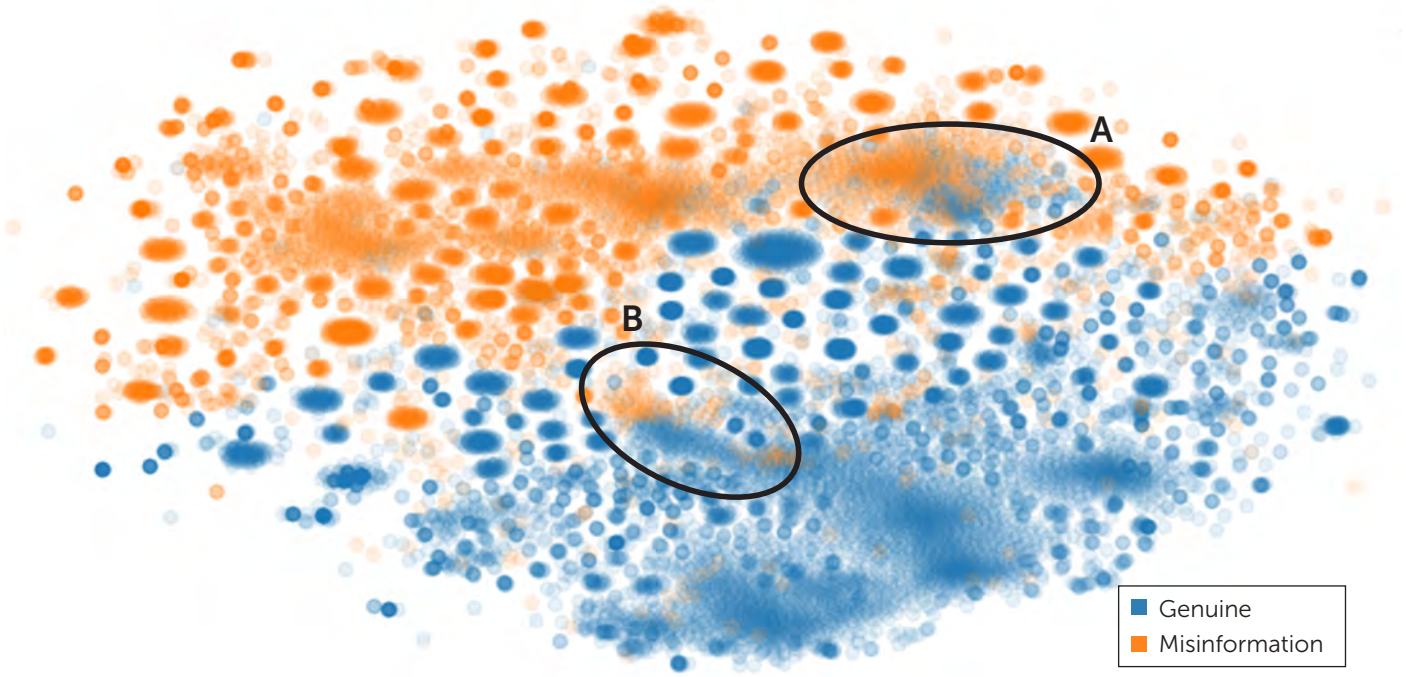
At a high level, our methodology works in three phases. In the first phase, we find the best-fit network for each mode (i.e., text, friends and followers, images, etc.) of the data. In the second, we use a scoring-and-sampling procedure based on common network-science metrics and Metropolis sampling (a random sampling method used when direct sampling is difficult) to produce a final fused graph from each of the modal graphs. Finally, we analyze the fused graph using common network science techniques and visualizations to identify and then characterize possible misinformation techniques. We call our current technique for unsupervised learning on multimodal data "graph annealing."

To illustrate how our methodology can be used to identify and characterize misinformation campaigns, we have employed it in a fake news identification data set. The set comes from a public data science competition website and consists of approximately 18,000 (mostly English) articles taken from various online websites in 2016. It features the author and text of the articles. The articles are about 40 percent fake and 60 percent genuine. They were hand labeled as being fake news or not, and have some error introduced by human bias into the labeling.

To leverage fully the richness of the data set, we extracted three modalities from the data. The first modality was the diction of the different articles, with the idea being that those articles of a dubious nature may use slightly different words, such as superlatives. The second modality was the parts of speech of the different articles based on recent research showing greater use of adjectives and adverbs in fake news. Finally, we incorporated the social context of the articles by including the republication network of the articles, as authors often republish articles that support their narratives. Using these three modalities of the articles, we annealed a graph to produce a cohesive, analyzable data model.

The fused model of the data produced some interesting and useful results for understanding the misinformation aspects of the data set. The annealed graph had sub groups that tended to form around particular topics, such as Middle East relations or President Obama's legacy, and around various veracities of the articles. In particular, more contentious, popular topical sub groups, such as the 2016 election and the state of the US economy, tended to have both genuine and fake news articles present in the subgroup, while other topic sub groups, such as personal health or war with Russia, tended to be entirely all genuine or fake. Thus, the

Latent Spatial Positions of Misinformation-labeled and Genuine-labeled Online News Articles



sub groups in the data not only tend to indicate whether an article is fake news or not, but also what particular topics are most inclined to having fake news.

Visualizing the fused data also showed interesting disinformation trends. To visualize the different sub groups of the data and their relationships, we used a t-Stochastic Neighbor Embedding (a machine learning algorithm for visualization) of the annealed graph. The following figure displays the visualization of all of the articles in the data set with their given hand labels.

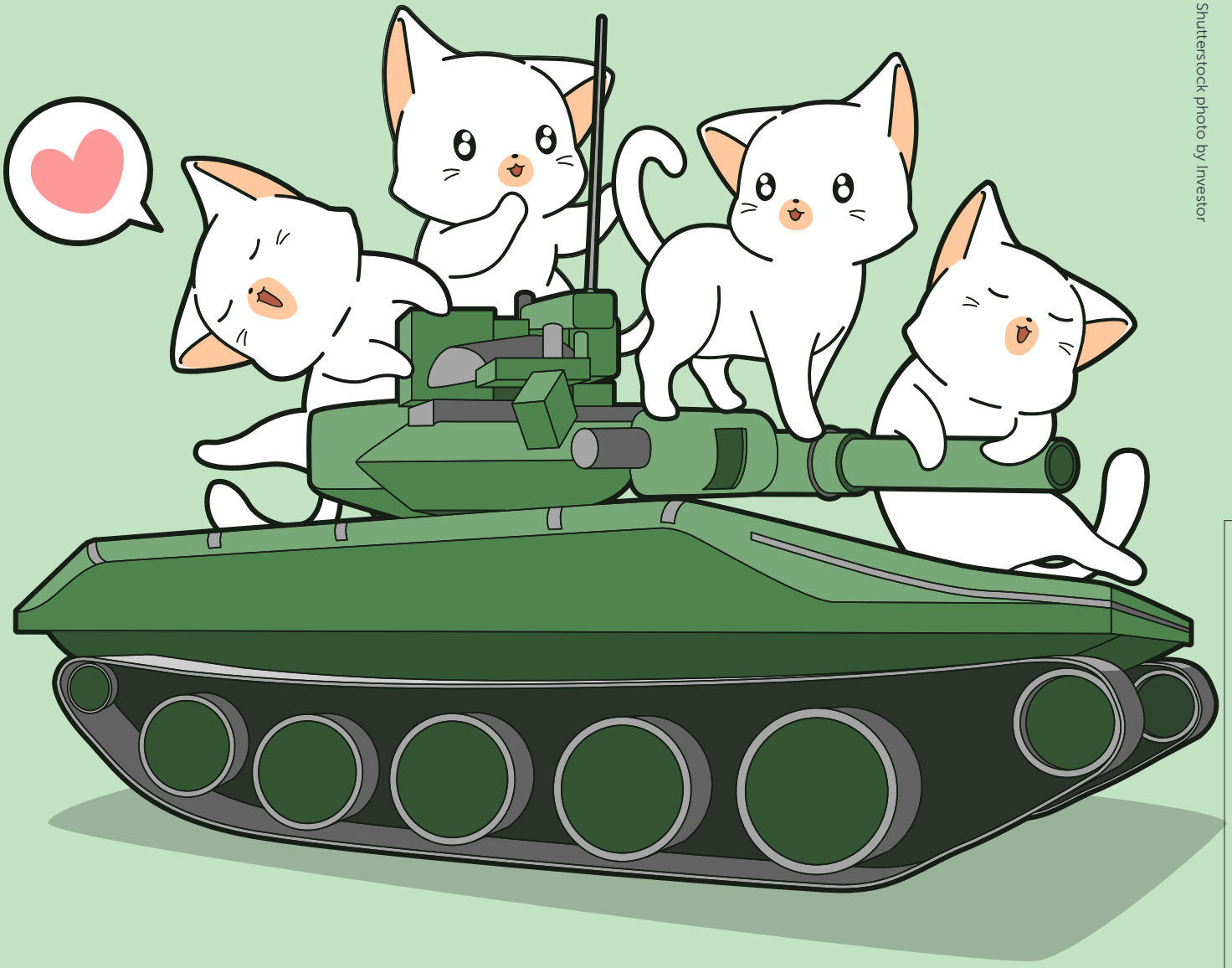
With the visualization we can observe that most of the fake news articles tend to lie in a separate area of the space than the genuine news articles. Some notable exceptions to this are the regions labeled A and B. Region A consists of an especially large sub group of articles that are both fake and real and focus around US economic news. What is interesting to note is that these articles are difficult to separate, as both the real and fake articles feature more alarmist language concerning the possibility of an economic downturn following the 2016 presidential election. It would seem many of these fake news articles were aimed at stoking these fears and possibly part of larger misinformation campaigns to sow unease in the United States with the coming elections. In region B, there also is a mix of fake and real news articles, which feature President Obama's healthcare and education initiatives as the main topic. This particular topic sub group featured a lot of emotionally charged language from both true and fake articles. Thus, it would seem the fake news articles in this particular topic group might have been part of larger misinformation campaigns to inflame emotional tensions during the election timeframe. Overall, visualizing the annealed graph allows one to quickly locate major subgroups present in the data and gain better understanding of possible relationships within the data.

Misinformation campaigns remain a major threat to society. One of the great difficulties in countering these misinformation campaigns is the ability to identify when one is occurring and characterize the nature of that campaign. To do so requires more than fake news classification or even the identification of faked data. As such, we are developing techniques that can incorporate many data types, including the social context, in an unsupervised manner to create a cohesive data model that enables an analyst to better identify, and characterize misinformation campaigns. Our preliminary research has demonstrated strong findings in its ability to enable identification and characterization of misinformation campaigns.

About the authors:

Capt. Cruickshank is a National Science Foundation Graduate Research PhD student in societal computing at Carnegie Mellon University.

Dr. Carley is a professor of societal computing in the School of Computer Science at Carnegie Mellon University.



Shutterstock photo by Investor

EXPLORING THE POWER OF CUTE

By Dr. Susannah Paletz and Dr. Ewa Golonka

CUTENESS HAS A WAY OF MAKING NEARLY ANYTHING—EVEN THE MOST SERIOUS OF SUBJECTS—INTO SOMETHING DISARMING, FUNNY, OR POSITIVE. PUTTING A CAT ON A TANK MIGHT SEEM RIDICULOUS, BUT SCIENCE SAYS THE POWER OF CUTE IS NOTHING TO LAUGH AT.

Program managers and researchers at the Office of Naval Research (ONR) have supported and partnered with academic experts to explore the polarization of news. Our research team drew on the latest in the psychology of emotions to create a new methodology to examine the role of emotions in social media sharing.

The internet is a dark place, filled with hate speech, deliberate misinformation, trolls, and bots. After the multipronged Russian influence campaign during the 2016 US election, information warfare has intensified. In 2018, researchers at the University of Washington visualized how trolls at the Russian Internet Research Agency played both sides of the Black Lives Matter debate against each other, increasing polarization.[1] With an increase in these types of subtle and not-so-subtle influence campaigns, it becomes even more pressing for researchers to understand the nature of malign influence.

The study of social influence and propaganda is not new. Recently, Lt. Col. David Beskow and Professor Kathleen Carley created the "BEND framework," to describe communication tactics that might be used on social media (see page 22). Communication tactics support achieving desired, overarching communication objectives. Some typical tactics include distorting the truth and attempting to dismiss a topic. Other tactics include exciting a group, or raising discussions that bring joy and cheer, and distracting attention from a topic, person, or brand, which involves sidelining a conversation to something irrelevant. In the course of this research, it became clear that although the internet is filled with negativity, it also could be a positive experience. Cute messages on social media might serve to excite a group as well as distract—by spreading cute images, social media manipulators, be they genuine humans or bots, can change the conversation to something positive and innocuous.

Cute things have played a role in propaganda before. ISIS used cats and kittens in some of their recruitment materials, including having a soldier posed with a kitten in the 15th issue of *Dabiq* (ISIS's official magazine).[2] In addition to potentially softening their image, photographs such as these were probably a reference to a companion of the Prophet, Abu Huraira, who was fond of cats.[3] Research on cuteness and its emotional reactions has a long history in the scientific literature: as early as 1943, Konrad Lorenz claimed that certain child-like features such as large eyes, a small nose, a round face, and/or playful behavior evoke caretaking behaviors by adults.[4] In more recent studies, these features have been linked to cuteness.[5] There is research on the characteristics of cuteness and how people are attracted to cute stimuli, as well as on the emotions cuteness elicits.[6] The current challenge, however, is to develop ways to measure cuteness and the reactions cuteness evokes. Being able to quantify cute content and reactions to it is especially important when the messenger intends to appeal to specific audiences, such as in marketing and product advertising or in information warfare.

Bad actors using cute things in their messaging is a concern, particularly because social media cross national borders. ONR funded our social media studies to examine specific emotions that might promote social media sharing.[7] Our research used a rating scale, which ranged from 0 (not present) to 100 (extremely intense), which included more than 20 emotions. Using a trained research assistant or scientist, each social media post was annotated. These annotators assessed each post independently, but then came together to debate each emotion assessment for each post to achieve consensus and improve validity. The annotators also recorded their reactions to the posts using the same list of emotions. Those reactions were averaged and assessed for breadth across the annotators, but not debated.

By using this process, researchers separated the effects of the content of a post from the reactions elicited by that post, such as when a post does not contain angry content but makes people who read it angry. This methodology was limited in terms of scale—annotation can be time-intensive and depends on human judgment—but it captured the full range of multimedia (e.g., photos, video, memes) on social media, not just text. It also covered a broader array of emotions than is typically examined in social media research. In addition to emotions such as anger, fear, disgust, surprise, and happiness, researchers also looked for gratitude, nostalgia, pride (including

ethnonationalism), hate, contempt, love, admiration—and kama muta, a Sanskrit term usually translated as the emotional reaction of feeling heart-warmed when seeing cute, infant-like things.

To continue this research team's advances in social media research, ONR awarded our team (which includes our colleagues Dr. Anton Rytting and Dr. Cody Buntain and researchers Egle Murauskaite and Devin Ellis) a grant through the Minerva Research Initiative to assess 1,000 Facebook posts and 300 YouTube videos from Polish and Lithuanian sociopolitical influencers. The goal of this research is to examine the role of emotion and narratives in social media sharing by political and social influencers in these two NATO member states.

By determining ways to measure cuteness and emotional reactions to cute stimuli, it is possible to then conduct statistical analyses on the impact of cute content and reactions on social media sharing. To prove this hypothesis, researchers conducted a pilot study in 2018 using Twitter to test the role of different emotions, including kama muta, on social media sharing. Our methodology was based on previous work on cuteness that focused on the specific characteristics of infants, or baby schema (such as appearance features such as small size, chubby cheeks) as well as child-like behavior (clumsy walking, playfulness, etc.). Researchers also adapted the concept of kama muta as a particular emotional reaction to cuteness.

As expected, researchers found that tweets containing cute images or cute behavior and tweets that evoked heartwarming responses were more likely to be shared.[8] However, after controlling for relevant confounds (extraneous factors that are potentially associated with both cuteness and sharing, such as the amount of time between when a tweet was posted and when researchers downloaded the retweet), a simultaneous positive effect for the heartwarming feeling and negative effect of the cute content on sharing was found. This complex finding suggested that the heartwarming feeling is essential to sharing on social media. Specifically, when a tweet depicting or discussing something cute did not evoke kama muta in the reader (not all people perceive cuteness the same way), that tweet was less likely to lead to sharing.

Although intuitive, research that directly measures and tests the effect of cute content, or, more importantly, the heartwarming emotion that arises from viewing cute content, is relatively new. Recent research proves that while online content tends to favor negative emotions, it is possible for positive emotions, such as kama muta, to stifle content designed to divide.

References

[1]A. Arif, L. G. Stewart, and K. Starbird, "Acting the part: examining information operations within #BlackLivesMatter discourse, "*Proceedings of the ACM on Human-Computer Interaction* 2 (2018): Article 20.
[2]Lizzie Dearden, "Isis using kittens and honey bees in bid to soften image in Dabiq propaganda magazine," *Independent*, August 2, 2016, https://www.independent.co.uk/news/world/middle-east/isis-kittens-honey-bees-dabiq-propaganda-recruits-photo-soften-image-terror-a7168586.html.
[3]J. P. Farwell, "The media strategy of ISIS," *Survival* 45 (2014): 49-55.
[4]K. Lorenz, K. (1943). "Die angeborenen Formen möglicher Erfahrung," *Z. Tierpsychol* 5 (1943): 233–519; cited in M. Borgi, I. Cogliati-Dezza,, V. Brelsford, K. Meints, and F. Cirulli, "Baby schema in human and animal faces induces cuteness perception and gaze allocation in children," *Frontiers in Psychology* 5 (2014): 411.
[5]M. L. Glocker, D. D. Langleben, K. Ruparel, J. W. Loughead, R. C. Gur, and N. Sachser, "Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults," *Ethology* 115 (2009): 257-63.
[6]D. Jones, "Sexual selection, physical attractiveness, and facial neoteny: Cross-cultural evidence and implications," *Current Anthropology* 36, no. 5 (1995): 723-48; K. K. Steinnes, J. K. Blomster, B. Seibt, J. H. Zickfeld, and A. P. Fiske, "Too cute for words: Cuteness evokes the heartwarming emotion of kama muta," *Frontiers in Psychology* (2019).
[7]S. B. F. Paletz, ed., *Measuring emotions in social media: Examining the relationship between emotional content and propagation* (College Park, MD: University of Maryland Center for Advanced Study of Language, 2018).
[8] "Kama Muta Lab," accessed November 5, 2019, http://kamamutalab.org/.

About the authors:

Dr. Paletz is a research professor at the University of Maryland's College of Information Studies and an affiliate at the University of Maryland's Applied Research Laboratory for Intelligence and Security (ARLIS).

Dr. Golonka is an associate research scientist at ARLIS.

COMPILE TO COMBAT FOCUSES ON DELIVERY TO THE FLEET

By Philip Baptiste and Patric Petrie

In 2019, Naval Information Warfare Center Pacific completed and important phase of Compile to Combat entitled "Digital Abe," which is a digital representation of a cloud-based information warfare system that will be installed on USS *Abraham Lincoln* (CVN 72) in 2020.

Photo by MC3 Zachary Pearson

In 2018, the chief of naval operations issued Navy-wide guidance to transform the enterprise information environment and rapidly deliver capability using the "Compile to Combat in 24 Hours" (C2C24) framework. This framework puts the Navy on the path to modernizing data and information technology architecture, including using commercial industry best practices for software modernization.

C2C24 provides a standardized structure to transform the Navy's information environment through the adoption of a service-oriented application architecture and common standards for data formats and interfaces. It is an afloat, end-to-end architecture designed to deliver capability in a matter of hours—compared to previous efforts that required a nine-month process or longer. By using commercial technology and open standards for maximum agility, C2C24 aligns with the latest maritime strategy document, "A Design for Maintaining Maritime Superiority Version 2.0," and its "lines of effort" to achieve high-velocity outcomes and deepen naval integration with joint/coalition partners.

As one of the principal contributors, Naval Information Warfare Center (NIWC) Pacific led a series of C2C24 pilots in April and May 2019 aboard several Navy units that demonstrated the ability to deliver a software capability "at the speed of relevance."

NIWC Pacific's Collaborative Software Armory is a crucial enabler for C2C24, and is an instrumental piece of the end-to-end development security operations pipeline. The armory uses the commercial cloud, automated development, and test tools to provide an environment where systems and software are developed with security incorporated early, fully accredited, and ready for installation within a few days instead of months. The armory also will enable program managers to reduce Risk Management Framework accreditation timelines significantly and will enable improved cybersecurity monitoring.

NIWC Pacific also is developing a digital distribution service, known as Application Arsenal (formerly PEO C4I Storefront), in alignment with the C2C24 framework. This service is an integral part of the Navy afloat platform and designed to deliver approved software to the fleet.

In April 2019, NIWC Pacific completed the next phase of C2C24, informally titled "Digital Abe." "Digital Abe is named after the USS *Abraham Lincoln* [CVN 72] because it provides a digital representation in the cloud of the Information Warfare Platform 2.0 to be installed on the ship in 2020," said Delores Washburn, chief engineer at NIWC Pacific. "More importantly, it provides a significant change to the way we do business because it allows us to do continuous test and integration in the cloud."

Digital Abe was a dry-run program of record implementation/integration test of C2C24. The objective was to conduct a cloud-based event and use an operationally representative ship, using methods demonstrated during the pilots and enabled by the Collaborative Software Armory. Participants included Maritime Tactical Command and Control (MTC2), Consolidated Afloat Network and Enterprise Services (CANES)/Agile Core Services (ACS), and Application Arsenal.

The Digital Abe live demonstration of MTC2 application containers, which traversed through the C2C24 development security operations pipeline and operated on a digital representation of an afloat platform, showcased the culmination of significant engineering efforts and the maturation of the C2C24 framework.

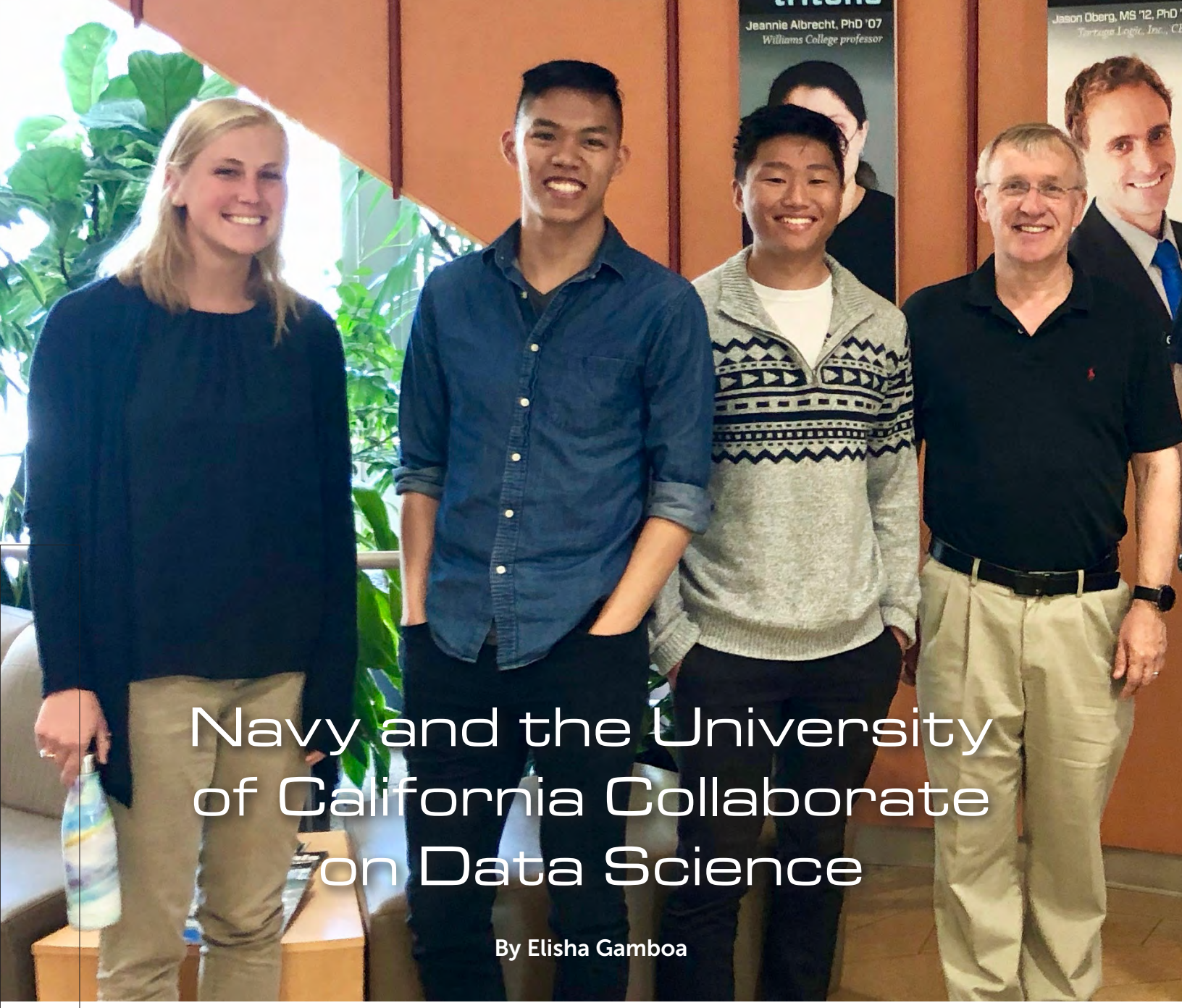
In May 2019, the NIWC Pacific C2C24 team conducted an experiment to containerize and re-host a US Air Force application—developed by the Air Force's new experimentation lab, Kessel Run—on the Navy's ACS platform. Adapting a Kessel Run application to run on ACS demonstrates the ability to move software code between two distinct platforms, similar to having the same app on an iPhone and Android. This experiment reveals the Navy's path on using a C2C24 development, security, and operations process for integrating and deploying products developed by the Air Force, resulting in a much more flexible pipeline in support of C2C24 and deepening naval integration of other joint apps in the future.

The C2C24 team received one of the Department of Navy's Information Management/Information Technology Excellence Awards winners during a ceremony at the Department of the Navy Information Technology Conference/AFCEA West Symposium in February 2019. The C2C24 team was recognized for demonstrating a culture of transformation with new modernized processes and use of the commercial cloud.

From pilot projects to program-of-record systems to joint applications, NIWC Pacific continues to develop the C2C24 framework and scale up the ability to deliver software capabilities to the fleet.

About the authors:

Philip Baptiste is the lead for C2C24 Digital Abe.
Patric Petrie is lead staff writer for Naval Information Warfare Center Pacific.



Navy and the University of California Collaborate on Data Science

By Elisha Gamboa

Naval Information Warfare Systems Command (NAVWAR) is collaborating with the University of California, San Diego (UCSD), to study the use of data science techniques and analysis to increase competitive intelligence and improve decision-making within the Navy.

Established in mid-2018, the NAVWAR data science team comprises the command's Logistics and Fleet Readiness Competency, UCSD's Halıcıoğlu Data Science Institute, and multiple Navy organizations focused on examining data-driven decision-making to produce calculated insights and solve complex problems for increased operational efficiency.

"Modern advancements in technology have drastically increased the amount of information that is being collected and stored," said NAVWAR executive director Pat Sullivan. "That information has the potential to enable a competitive edge for the fleet, but only if the Navy has the ability to interpret it and apply it to the current environment."

UCSD established the Halıcıoğlu Data Science Institute in 2018, an academic unit focused on training students, faculty, and industrial partners to use data science in new and novel ways that will allow them to understand some of the world's most pressing problems.

"Partnering with UCSD has greatly accelerated our efforts to become more innovative in our approach to answering difficult questions through data," said David Byres, NAVWAR information technology specialist and project manager for the data science team.

Photo by Elisha Gamboa

The NAVWAR data science team established a project group and developed basic processes for team collaboration, data extraction, exploration, and analysis.

Together they built on a hypothesis that one could combine data, in the form of problem descriptions, from the Regional Maintenance Center's support database with other incident management data to reveal trends and produce insights about possible solutions for new fleet support incidents.

"During their research, the team applied natural language processing techniques to relate short problem description narratives among different and unrelated fleet support IT systems," said Byres. "By relating these separate sources of information, potential solutions to fleet problems that are 'bottled up' in our IT systems could potentially be leveraged to more rapidly identify solutions to new problems."

The project team concluded this effort in June 2019.

In addition, the team has taken on a more involved data science endeavor. Expanding on a RAND Corporation study on Navy network dependability and applying it to current fleet challenges, the team is analyzing factors that influence the "user perceived dependability" of complex NAVWAR systems.

While the availability and reliability of IT systems determine network dependability, user-perceived dependability refers to how reliable a user understands or perceives a system to be.

Today, the Navy increasingly depends on networks and associated netcentric operations to conduct its missions. To increase their understanding of network dependability and user-perceived dependability with intraship and multiship networks, the NAVWAR data science team is investigating the relationship among existing sources of data including hardware, software, and human factors.

More specifically, the team is trying to understand which fleet IT staffing and training factors most affect a ship's ability to support their Consolidated Afloat Network and Enterprise Services (CANES), and why some ships are more successful in operating CANES than others.

"We're working closely with our many project partners to arrive at a much deeper understanding of what factors most impact a crew's ability to operate and maintain their complex shipboard IT networks," said Byres. "We've really just scratched the surface in terms of identifying and working with Navy data that might provide answers to our most basic questions."

The long-term project is currently ongoing, with an end date projected for fiscal year 2020. As the demand for data science in defense continues to grow, NAVWAR will continue to pursue the use of more advanced data science techniques and analytics to better capitalize on existing information to ensure the Navy can fight and win today and in the coming decades.

About the author:

Elisha Gamboa is a public affairs specialist with Naval Information Warfare Systems Command.

Jara Tripiano, Naval Information Warfare Systems Command (NAVWAR) division head for cybersecurity, engaged with current and prospective students during the launch of the University of San Diego's Cyber Bootcamp on 4 December 2019. Photo by Elisha Gamboa

FUTURE FORCE is a professional magazine of the naval science and techology community published quarterly by the Office of Naval Research.

Future Force
Office of Naval Research
875 N. Randolph Street, Suite 1425
Arlington, VA 22203-1995

Email: futureforce@navy.mil
Phone: (703) 696-5031
Web: http://futureforce.navylive.dodlive.mil
Facebook: http://www.facebook.com/navalfutureforce