Disrupting Deterrence Signaling: Examining the Fifth Wave of Technology’s Impact

Brendan Balestrieri
Korea University, brendan.j.balestrieri@gmail.com

Follow this and additional works at: https://digitalcommons.usf.edu/jss
pp. 1-10

Recommended Citation
DOI: https://doi.org/10.5038/1944-0472.16.2.2109
Available at: https://digitalcommons.usf.edu/jss/vol16/iss2/1

This Article is brought to you for free and open access by the Open Access Journals at Digital Commons @ University of South Florida. It has been accepted for inclusion in Journal of Strategic Security by an authorized editor of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.
Disrupting Deterrence Signaling: Examining the Fifth Wave of Technology’s Impact

Abstract
A critical component of deterrence is signaling, or how states communicate desired images of themselves and their intentions.[i] While scholars advocate for a re-examination of classical deterrence theories in light of the emergence of artificial intelligence (AI) and autonomous systems, some view the impact of AI and autonomy as a force that will diminish human agency as a component of deterrence and interactions between nations. This article argues the contrary and posits that the emergence of AI and autonomous systems will magnify the importance of human agency by enabling decision-makers to more accurately gauge credibility, will-to-fight factors, and the national resolve of their opponents based on aggregated sentiment analysis and psychological profiling of civilian populations. Furthermore, this article contends that traditional forms of signaling are undermined by the pervasiveness of AI, autonomous systems, social media, metadata, and the ability of nation-state actors to more accurately surmise the credibility and national resolve of an opponent.

Introduction

At its core, signaling is about humans attempting to understand what other humans are thinking. However, deterrence theory assumes that opposing actors receive and decode these signals even if they do not necessarily believe the messages conveyed, the result often being faulty communication and unexpected behavior. Faulty communication and unexpected behavior are only a few of the foundational weaknesses upon which traditional interpretations of signaling rest.

The inexactitude of signaling is inherently immense because success relies on the recipient not only correctly interpreting what the sender signals, but also finding it credible. Part of what makes signaling behavior so unpredictable is attributed to what Robert Jervis called “signaling reputation,” which Jervis defined as an actor’s reputation for following through on announced intentions. Numerous examples of failed deterrence throughout the 20th and 21st centuries exemplify the potential for miscommunication.

Actor motivation also plays a part. Paul Huth writes that there is a higher likelihood that defending states have an incentive to always act as if they are determined to resist an attack in the hope that an adversary will back down from conflict, the result being that potential aggressors may discount statements or military maneuvers from the defender as bluffs. Actor motivation is perhaps of greater significance in democratic countries where leaders may be uncertain about the level of domestic support for a given policy. Substitution factors for why and how deterrence may or may not be working also continue to plague the canon of deterrence research and undermine scholarly claims of successful or failed deterrence.

Much like deterrence writ large, establishing whether deterrence works, and whether the intended recipients correctly receive and interpret the signals proves difficult. Ultimately, it is dependent on circumstantial evidence, and even then, alternate forms of explanation for why an actor changed its behavior may exist. Moreover, the absence of an attack doesn’t always mean that deterrence has been achieved, since the potential attacker possibly never intended to attack in the first place, and if so, the attacker was never deterred. While numerous examples of deterrence
failure exist and can be more readily studied, examining the reasons why two nations did not go to war is more challenging.

Absent clear evidence of deterrence success or failure, policymakers must self-assess the effectiveness of their messaging, opening the possibility for several biases that have the potential to undermine signaling behavior. Case study analysis of past disasters shows that cognitive biases may distort decision-making, particularly the biases of groupthink, fallacy of control, framing, fallacy of plan, and confirmation bias. Perhaps foremost among the challenges of signaling is assessing the credibility of what is signaled.

Assessing Credibility

Crucial to successful deterrence is the ability of the deterrer to effectively signal threats, and that these threats are perceived as credible. However, the credibility of those threats is often undermined by misunderstandings, and in some cases, lack of faith in the truthfulness of what is being messaged since the act of deception in the pursuit of strategic interests at times may lead nations to engage in lying, spinning or concealment. Lying, spinning or concealment, and the potential for breakdowns in trust, are potential factors when it comes to alliances and the credibility of alliance commitments to defend a third party. Breakdowns in trust between allies could result in suboptimal deterrence outcomes, especially as the Cold War era debate over the credibility of the United States guarantee to “sacrifice New York for Paris” is re-introduced in the era of extended deterrence and North Korea’s growing nuclear and missile arsenal. The statements made by United States officials after leaving office, which sometimes contradicted promises made to alliance partners while in office, further undermine the credibility of policymakers and traditional signaling behavior.

Numerous factors shape how one side assesses the credibility of the other. Factors influencing how an adversary perceives the credibility of an opponent include assessments of capability (military-economic strength), the perceived resolve of a nation’s people to accept casualties, and also the personal dynamics that may exist between political-military leaders of nations, which may result in some measure of “micro-signaling.” During the 20th Century, assessments of credibility mostly relied on the first and
third factors, since assessments of national resolve were largely limited to anecdotal evidence or the pre-conceived experiences of decision-makers with the other country (for example, the German High Command had a “deep-seated respect for both French and British military prowess” and initially only advocated for a limited aims strategy for attacking France).  

A similar story played out in the East as well. Despite concerns from the German Staff over Hitler’s plans to invade the Soviet Union based on geographic-military factors, his claim that “we have only to kick in the front door and the whole rotten Russian edifice will come tumbling down” helped inform his decision-making. What Hitler may have been thinking is conjecture, but it may be safe to say that he perceived Soviet resolve to withstand an attack to be minimal based on observations of the Soviet Union’s performance during the Winter War between the Soviet Union and Finland several years prior. Ultimately, it is perceptions of the aggressor that matter most and not the prospects for success or the objective assessments of the consequences of attacking.

Closely tied to credibility is the notion of resolve. While previous research during the Cold War considered a nation’s reputation for toughness a key component of how an adversary may assess an opponent’s resolve, recent research has shown that it may be based more on specific situations and how leaders rely more on timing, risk preferences, self-control, and honor orientation. Deterrence theorists have also argued that potential attackers will rely on the specific intrinsic interests of the defender (such as losing territory or economic or military ties with an ally) as well as the defender’s bargaining behavior when assessing the defender’s likely resolve. However, in the wake of Russia’s failure to conquer Ukraine quickly, greater weight to the importance of national resolve beyond simply the application of intrinsic interests and bargaining behavior is needed.

In the case of Russia, the deluge of TikTok videos that show Russian soldiers surrendering or unwilling to fight is one example of how social media may be undermining traditional forms of signaling. Russia’s inability to deter Western support for Ukraine, despite the numerous forms of traditional signaling practiced by Russia’s political-military elite, could perhaps be attributed to assessments made in the West that rely more on what is being seen first-hand on the battlefield, and what is being
communicated directly from the people, and soldiers, of Russia. If so, the use of such information mediums for making assessments could present a significant shift in how deterrence theory defines signaling, the conduct of foreign policy decision-making, and most importantly, how nation-states assess the readiness, credibility, and resolve of their opponents.

True Readiness

Beyond assessing credibility and resolve, signaling has another foundational weakness, as it mostly involves conscious and deliberate actions that decision-makers on either side take while trying to shape their desired images and intentions. Based on the latest advancements in technology, decision-makers may now be able to form opinions and assess credibility based on the collective resolve of the opposing nations and are no longer as reliant on the deliberate decisions, and subsequent signals, conveyed by decision-makers on either side. The potential may now exist to aggregate assessments of national readiness from the collective consciousness of a nation’s people, compiling information from keywords, social media posts, search engine queries, text message conversations, phone calls, and other mediums through which thoughts about military service or support for national military strategies can be gauged.

The emergence of AI, and autonomous systems could potentially enable the collection of metadata, providing a more honest assessment of an opponent’s capability, and allowing them to see beyond what is formally communicated. Moreover, these technological innovations are no longer nascent and can leverage the pervasiveness of social media. According to Freedom House, “Social media surveillance refers to the collection and processing of personal data pulled from digital communication platforms, often through automated technology that allows for real-time aggregation, organization, and analysis of large amounts of metadata and content.”16 Many nations are quickly approaching the point where AI and automated technology may enable the close observation and aggregation of individual data.

The countries of North and South Korea present a good example of how AI, autonomous systems, and social media might impact signaling behavior, assessment of capability, and how North Korea perceives the true readiness of the South Korean military to wage war. Globally, South
Korea has the second-highest rate of active social media users, consisting of 91.2 percent of the total population. The high penetration rate of social media extends to members of the South Korean military, who are now allowed to have cell phones while participating in their mandatory military service. The potential for AI and autonomous systems to gauge individual sentiments, and then aggregate them, has the potential to provide the enemy with greater contextualization when assessing the messages conveyed by national leaders.

In the case of South Korea, a national reliant on its 2.3 million reservists to mobilize quickly in response to a North Korean attack, the potential for AI and autonomous systems to aggregate data on individual reservists could undermine defense strategies. For example, South Korea relies on reserve divisions to guard the nation’s critical infrastructure and logistics networks in the rear area, and the soldiers assigned to defend them are tasked to mobilize at reserve centers throughout the country. However, as South Korea’s youth pursue jobs in Korea’s megacities of Seoul and Busan, those individual reservists’ ability to respond quickly is potentially undermined due to their relocation away from those rural mobilization centers. Geotracking is but one of many examples of how AI and autonomous systems, weaponized by North Korea or a partner state to monitor and aggregate the locations and perceptions of individual South Koreans, especially South Korea’s 2.3 million citizen-soldiers, could potentially influence North Korean perceptions of South Korean readiness, and in the process, undermine both South Korean deterrence initiatives and the credibility of South Korean political-military elite signaling.

In the case of extended deterrence, particularly in the extended deterrence scenario of San Francisco for Seoul, AI and autonomous systems could potentially obtain a fuller picture of how willing the American people might be to support an alliance partner. Opponents of the United States could also potentially gain greater insight into the level of support that exists in countries like South Korea and Japan for alliance partnerships. The capacity for greater insight into what the people of a nation are thinking has the potential to shape leaders’ perceptions when considering decisions for aggression, especially in democratic countries where leaders face greater accountability to the opinions of the voting public.
As the fifth wave of technology continues to unfold, one of the most powerful mediums through which national resolve is potentially communicated is through social media, smart phones, and online media, which is a direct line to the mind of a nation’s people. But perhaps what makes the aggregated opinion and perspective of a nation’s people so important is that it has the potential to be honest, and lessens the reliance on the crafted, and biased, images and messages created by national leaders. Removed are the biggest unknowns and sources of unpredictability when nation-state leaders are engaging in overt or oftentimes nuisance deterrence signaling. AI and autonomous systems could potentially allow nation-state leaders to deduce the resolve, and by default capability and credibility, of what the other side is saying. Additionally, these technologies could also help national leaders assess the resolve and ability of their own nation to fight and win a war.

Monitoring the Fifth Wave

Recent literature on the implications of AI and autonomous systems for greater government surveillance also helps shed some light on the potential for those same systems to influence deterrence and perceptions of national resolve. The emergence of social media as a permanent fixture in modern society has coincided with governments seeking out the means to monitor their citizens’ online presence and social media discourse, the example of Cambridge Analytica demonstrating the reach and potential for organizations to amass personal information of private citizens. The technique employed during the Cambridge Analytica scandal is that of psychological targeting or the practice of extracting the psychological profiles of people based on their digital footprint to influence their behavior.19

Freedom House goes a step further in describing the potential for digital systems to collect data on individuals:

[AI has] opened up [sic]new possibilities for automated mass surveillance. Sophisticated monitoring systems can quickly map users’ relationships through link analysis; assign a meaning or attitude to their social media posts using natural-language processing and sentiment analysis; and infer their past, present, or future locations. Machine learning enables these systems to find
patterns that may be invisible to humans, while deep neural networks can identify and suggest whole new categories of patterns for further investigation. Whether accurate or inaccurate, the conclusions made about an individual can have serious repercussions, particularly in countries where one’s political views, social interactions, sexual orientation, or religious faith can lead to closer scrutiny and outright punishment.\(^\text{20}\)

The potential for AI and autonomous systems to assess an adversary’s national resolve, particularly in hyper-connected countries like South Korea and Japan, could result in unexpected deterrence outcomes, particularly if these technologies undermine national political-military elite signaling strategies. The pervasiveness of social media, and the degree to which individuals express their opinions via text messaging and other digital mediums, provide further means for nations with aggressive cyber capabilities to potentially tap and aggregate population sentiment. The phenomenon of aggregating population sentiment for the purpose of assessing national resolve could even potentially provide more repressive regimes with an advantage over their democratic adversaries. Nowhere is the contrast clearer than on the Korean Peninsula.

The contrast between the repressive authoritarian regime of North Korea and the open democratic society of South Korea is one example of how contemporary signaling strategies need to be re-examined. While North Korea maintains tight control over its population, and what outsiders can observe and learn from its population, South Korea’s population is wide open to observation by outsiders. Decision-makers should not discount the potential for North Korea’s aggressive cyber operations to use AI and metadata to gather information and public sentiment from South Korea’s hyperconnected population. For democratic countries like South Korea with elected leaders who are beholden to the voting public, the ability of an adversary to accurately gauge public opinion could potentially give them an edge when responding to traditional forms of signaling conveyed by that democratic country’s political-military elite.

Although research into the digital footprints of human behavior is still nascent, and primarily focuses on predicting election outcomes and marketplace preferences, there exists a likelihood that such research will both test social science theories in new ways and open up space for the
creation of new theories.\textsuperscript{21} Similar to how the nuclear deterrence theory of Thomas Schelling can partly be traced to deterrence in criminology and the writings of Cesare Beccaria and Jeremy Bentham, contemporary nuclear (and conventional) deterrence theory will undoubtedly have to pivot off of research taking place in other fields of study. The implications for how deterrence theorists view the component of capability, and signaling behavior among nations, could be significant.

Conclusion

Scholars consider signaling one of the most important elements of deterrence because it enables the deterrer to communicate and convey capability, resolve, or credibility. All these factors shape perception and how each side sees one another. However, deterrence research has not sufficiently explored the connection between these intrinsically linked aspects of deterrence, and most importantly, how the fifth wave of technology could potentially undermine legacy forms of signaling between nations.

The goal of this article is to help assess traditional understandings of signaling and deterrence theory in light of the fifth wave of technology and argue that more research is needed to fully understand its implications on foreign policy decision-making and how nations assess the credibility and national resolve of both their own nations and those of their opponents. Perhaps of greater importance, the advent of new technologies may allow nations to determine the true readiness state of their militaries more effectively. By aggregating sentiment analysis of individual citizens, AI, autonomous systems, and metadata can potentially amplify the role of human agency in national political-military deterrence decisions.

This article drives research further into the relationship between sentiment analysis, psychological profiling, and deterrence and thus, helps foreign policy decision-makers gain a more accurate understanding of opponent resolve and credibility behind signaling. Robert Jervis wrote that “in a better world, signaling and perception would be not only discussed between the same two covers but also intimately tied together.”\textsuperscript{22} The emergence of AI, autonomous systems, and the prevalence of social media as a defining element in society as part of the fifth wave of technology, may just allow for this to be possible.
Endnotes


