ARI 25/2020
17 March 2020

Bias and Misperception in Cyberspace

Miguel Alberto Gomez | Center for Security Studies, ETH Zurich | @mgomez85

Theme

With cyber operations serving as an instrument of foreign policy, it is fair to posit that cognitive factors that account for behavior in the physical domain are equally applicable to cyberspace.

Summary

A Psychological Turn. Our understanding of interstate behavior in cyberspace over the past decade rests firmly on systemic and technological attributes as determinants of strategic choices in this increasingly relevant domain. Scholars and policy specialists alike invoke established concepts such as the offense-defense balance, coercion, and signaling to account for state-associated cyber operations. Yet despite technological advancements, cyber operations continue to deliver limited strategic outcomes. This is paradoxical when accelerating investments in cyber capabilities are contrasted against lackluster performance thus far. Consequently, one may argue that attempts to frame strategic choices as a function of material and strategic realities hinders rather than enlightens attempts to comprehend state behavior in cyberspace. This, however, is not necessarily the case.

Recent cybersecurity scholarship acknowledges the importance of micro-level attributes. Whereas emphasis is commonly placed on the balance of power, dependence, and technological expertise; it is becoming apparent that cognition plays a crucial role in the decision-making processes that influence strategic choices. This psychological “turn” is not a novel occurrence as associated disciplines such as political science and international relations long recognized its importance. With cyber operations serving as an instrument of foreign policy, it is fair to posit that cognitive factors that account for behavior in the physical domain are equally applicable to cyberspace. Consequently, this ARI demonstrates this by discussing recent scholarship and how these affect the stability of cyberspace. In doing so, it surfaces the importance of taking a simultaneous top-down and bottom-up approach in evaluating state behavior in this man-made domain.

Analysis: uncertainty and cyberspace

One may argue that strategic choices made by states are an attempt to settle questions of uncertainty. Whether this pertains to the credibility of an emergent threat or adversarial resolve, uncertainty is a fundamental aspect of the international system. Yet uncertainty

is not a monolithic construct. As argued by Rathbun[1], it emerges either due to the lack of information or the ambiguity of the available information.

Concerning the former, the lack of information is resolved either through a test of arms that favors pre-emption as a means of settling questions of power balances or through information collected using espionage to gain an advantage over another. As Rovner[2] notes, this situation manifests itself in cyberspace as either military or intelligence contests. Whereas military contests suggest the need to dominate adversaries, intelligence contests are characterized by persistent competition and restrained action. Empirically, cyber operations over the past two decades are better represented by the latter given muted effects and restrain exercised by capable actors. This, however, does not account for the emergence of strategies such as persistent engagement that limits opportunities for further information collection due to its emphasis on pre-emption and disruption of adversarial operations. This calls into question whether uncertainty in cyberspace is indeed due to a lack of information available to decision-makers.

As Dunn-Cavelty[3] concisely argues, threat perception in cyberspace is due to the unknowability, vulnerability, and inevitability that characterizes the domain. Given its complexity, it is difficult to guarantee the security of cyberspace and its underlying components. Furthermore, recognizing the second to n-order effects of disruption is equally difficult. This complexity similarly gives rise to vulnerabilities within crucial systems that are not identified before deployment that, in turn, grants malicious actors the opportunity for exploitation. With cyberspace serving as an instrument of state power, adversarial motivations to resort to cyber operations becomes apparent and seems to shift the advantage in their favor. But this may not necessarily be the case.

Although low-hanging vulnerabilities are easily exploitable, these are likely to result in limited effects. Recent studies highlight the substantial material and organizational requirements necessary to develop and execute operations that adversely affect the strategic calculus of adversaries[4]. Relatedly, the apparent vulnerability of critical systems may also be used for deceptive purposes, luring adversaries into a sense of success when in fact defenders are in a position to observe and develop the necessary countermeasures against future operations[5]. Finally, the attribution problem often associated with the cyber operations is not as insurmountable when technical evidence

[1] Brian C. Rathbun (2007), "Uncertain about uncertainty: understanding the multiple meanings of a crucial concept in international relations theory", *International Studies Quarterly*, *51*(3), 533-557, https://www.jstor.org/stable/4621727?seq=1.

[2] Joshua Rovner (2019), "Cyber War as an Intelligence Contest", *War on the Rocks*, September 16, https://warontherocks.com/2019/09/cyber-war-as-an-intelligence-contest/.

[3] Myriam Dunn Cavelty (2013), "From cyber-bombs to political fallout: Threat representations with an impact in the cyber-security discourse", *International Studies Review, 15*(1), 105-122, https://academic.oup.com/isr/article-abstract/15/1/105/1791182.

[4] Allyson Pytlak and George E. Mitchell (2016), "Power, rivalry and cyber conflict: an empirical analysis", in Karsten Friis and Jens Ringsmore (Eds.), *Conflict in Cyber Space*, Routledge, 81-98. Rebecca Slayton (2017). "What is the cyber offense-defense balance? Conceptions, causes, and assessment", *International Security*, 41(3), 72-109.

[5] Erik Gartzke and Jon R. Lindsay (2015), "Weaving tangled webs: offense, defense, and deception in cyberspace", *Security Studies*, 24(2), 316-348, https://deterrence.ucsd.edu/_files/Weaving%20Tangled%20Webs_%20Offense%20Defense%20and%20Deception%20in%20Cyberspace.pdf.

is analyzed alongside the strategic environment in which these occur[6]. Consequently, the amount of information available to decision-makers is substantial and continues to grow with advances in technology and forensic processes. This is not to say, however, that the information is unambiguous.

Although states may be aware of adversarial operations on their systems, the underlying intent remains obscured. Buchanan[7], for instance, argues that the discovery of malicious code does not provide a clear indicator as to whether this is part of an on-going espionage operation or a precedent for future disruptive operations. The opacity of intent is further complicated by the criticality of the affected systems. Gartzke and Lindsay[8], for instance, argue that its appearance in sensitive systems such as Nuclear Command, Control, and Communication (N3C) infrastructure increases the likelihood of escalation. Aggravating matters further, forensic analysis is limited in its ability to identify the individual or organization directing the execution of specific cyber operations[9]. As was seen in past cases, the line between state and non-state actors in cyberspace is easily blurred, increasing the difficulty of assigning responsibility. Consequently, uncertainty in cyberspace is more a question of ambiguity rather than scarcity.

Heuristics driving strategic choices

As Rathbun argues, decision-makers overcome informational ambiguity through the use of cognitive heuristics. These serve to simplify complex environments, allowing decision-makers to achieve closure as efficiently and quickly as possible. From the perspective of operating in an ambiguous international system, these trigger two cognitive processes relevant to understanding strategic choices: first, setting and influencing expectations, and second, determining the plausibility of certain propositions[10]. The former is crucial in determining how state actors are expected to behave while the latter limits the range of plausible explanations for deviant behavior and appropriate responses to such.

Yet despite its efficiency, heuristics are constrained by the extent to which they match empirical reality. Otherwise, misperception and inappropriate strategic choices are likely to follow as observed in events such as the Yom Kippur war. In this case, Israeli assumptions regarding Egyptian capabilities and intent where made based on invalid assumptions. For cyber operations, similar cognitive mechanisms appear to be at work.

Thus far, our understanding of the extent to which elite decision-making is a function of heuristics is informed by a growing body of scholarship employing experimental designs and wargames. These studies reveal the extent to which individuals resort to analogical

[6] Thomas Rid and Ben Buchanan (2015), "Attributing cyber-attacks", *Journal of Strategic Studies*, 38(1-2), 4-37, https://ridt.co/d/rid-buchanan-attributing-cyber-attacks.pdf.
[7] Ben Buchanan (2016), *The cybersecurity dilemma: Hacking, trust, and fear between nations*, Oxford University Press.
[8] Erik Gartzke and Jon R. Lindsay (2017), "Thermonuclear cyberwar", *Journal of cybersecurity*, *3*(1), 37-48, https://academic.oup.com/cybersecurity/article/3/1/37/2996537.
[9] Herbert Lin (2016), "Attribution of malicious cyber incidents: from soup to nuts", *Journal of International Affairs*, *70*(1), 75-137, https://jia.sipa.columbia.edu/attribution-malicious-cyber-incidents.
[10] Robert Jervis (2009), "Understanding beliefs and threat inflation", in Trevor Thrall (Ed.), *American Foreign Policy and The Politics of Fear*, Routledge, 34-57.

reasoning, enemy images, and schemas when deciding on how best to utilize and respond to cyber operations.

In her longitudinal study of wargames conducted at the United States Naval War College, Schneider[11] illustrates how analogical reasoning influences and constraints the use of cyber operations during periods of conflict. Specifically, the escalatory potential of cyber operations was deemed comparable to that of nuclear weapons. In effect, participants were hesitant to authorize these before the onset of militarized conflict despite having the balance of power in their favor. Even in cases wherein conflict has already been militarized, hesitation persists with participants only willing to authorize operations that are easily recalled or those that are unlikely to be attributed.

Given the novelty of cyber operations and the lack of real-world cases of significant and extended physical effects, it is unsurprising that decision-makers are drawing parallels between cyber operations and nuclear weapons. The persistence of narratives that emphasize the destabilizing potential of malicious behavior in cyberspace in conjunction with a continued shortage of expertise in this domain is likely to motivate individuals to draw analogies between these and those with seemingly comparable features. For instance, on-going research at the University of Haifa[12] demonstrates the equivalency between cyber operations and terrorist activities based on comparable emotional responses generated by these events in a laboratory environment. Given our understanding that emotion precedes cognition, it is plausible that two distinct events that generate a similar emotional cue may lead to related, if not similar similar, strategic responses (i.e. retaliatory action employing a kinetic response) that may or may not be suitable for the given situation.

Besides analogies, enemy images are also found to be crucial in forming attributional judgments and evaluations of intent. These are cognitive constructs in which certain actors are believed to behave consistently in bad faith. This is built over time through repeated exposure to malicious behavior directed towards another such as repeated aggression or threats. Given the prevalence of rivalries among the cyber powers, enemy images likely play a crucial role in attributing and evaluating intent. Experimentally, this is demonstrated in a series of survey experiments where participants are informed of an incident involving a known adversary. Responsibility for the said incident, it seems, is consistently placed on the adversary irrespective of available information[13].

While this tendency to resort to beliefs is unsurprising, it signals a number of troubling possibilities. First, in attempting to avoid cognitive dissonance, elites may be susceptible to false flag operations as a means of either redirecting blame or further aggravating a situation that benefits a third party. As with analogical reasoning, this is further

[11] Jacquelin Schneider (2017), “Cyber and crisis escalation: insights from wargaming”, *USASOC Futures Forum,* March, https://pacs.einaudi.cornell.edu/sites/pacs/files/Schneider.Cyber%20and%20Crisis%20Escalation%20Insights%20from%20Wargaming%20Schneider%20for%20Cornell.10-12-17.pdf.

[12] Michael L. Gross, Daphna Canetti and Dana R. Vashdi (2016), “The psychological effects of cyber terrorism”, *Bulletin of the Atomic Scientists*, *72*(5), 284-291.

[13] Miguel Alberto Gomez (2019), “Sound the alarm! Updating beliefs and degradative cyber operations”, *European Journal of International Security*, *4*(2), 190-208.

aggravated by a lack of domain expertise. Second, adherence to enemy images increases escalatory risk. While escalation has yet to take place in response to a cyber operation, the absence of observable cases does not guarantee its impossibility. As noted by Gartzke and Lindsay, operations that affect crucial systems in the context of an established rivalry increases the likelihood of escalatory spirals. Third, while it is possible to alter beliefs, the dissonance required to achieve this is rare in the international environment. Consequently, advancements in forensic capabilities that generate more information are, on their own, unlikely to dislocate firmly embedded beliefs of adversarial malintent.

Finally, cognitive schemas offer elites a pre-established mechanism used to select an appropriate response to a cybersecurity incident. Schemas are cognitive constructs that *"represent knowledge about a concept or a type of stimulus, including its attributes and relations among those attributes"* [14]. Cybersecurity incidents, given its underlying geopolitical context, may contain specific stimuli that trigger schemas initially formed in response to non-cyber interactions.

Cross-national survey experiments conducted by Valeriano and Jensen[15] suggest that strategic preferences vary across groups. In their study involving participants from the United States, Russia, and Israel; responses to cybersecurity incidents appear to have unique national features. This appears to confirm earlier research that notes specific "national ways" of cyber conflict[16]. Relatedly, an on-going series of cross-national wargames conducted by Gomez and Whyte[17] illustrates the application of strategic solutions, previously employed in other domains, enacted in cyberspace. For instance, groups diverge on the appropriate approach in evaluating the culpability of malicious actors. Participants from the United States, for instance, drew parallels between a state hosting a terrorist group within its borders and its responsibility to address this security threat. In contrast, those from Taiwan and the Philippines formulated their assessments by extrapolating beyond the geographic descriptors offered by the technical evidence.
The preceding observations derived from the above highlight the challenges associated with schematically driven decision-making. While the existence of a national preference ensures a ready response to conflict, the underlying assumptions necessary for the successful use of these may not hold true for cyberspace. Moreover, cyber operations interact with the saliency of underlying issues. At best actions in cyberspace are

[14] Deborah Welch Larson (1994), "The role of belief systems and schemas in foreign policy decision-making", *Political Psychology*, 17-33.

[15] Brandon Valeriano and Benjamin Jensen (2019), "What Do We Know about Cyber Escalation and Conflict? Observations from Cross-National Surveys", *Atlantic Council*, November, https://www.atlanticcouncil.org/wp-content/uploads/2019/11/What_do_we_know_about_cyber_escalation_.pdf.

[16] Brandon Valeriano, Benjamin Jensen and Ryan C. Maness (2018), *Cyber Strategy: The Evolving Character of Power and Coercion*, Oxford University Press.

[17] Miguel Alberto Gomez and Christopher Whyte (forthcoming), *Cyber Uncertainties: Observations from Cross-National Wargames*, in Myriam Dunn-Cavelty and Andreas Wenger (Eds.), *Cyber Security Politics: Dealing with Socio-Technological Uncertainty and Political Fragmentation*, Routledge.

perceived as a continuation of existing rivalry/adversarial behavior. At worst, it may trigger a security dilemma due to the appearance of an, until then unused, instrument[18].

The likelihood of misperception also increases with the introduction of unsubstantiated assumptions. As mentioned above, schemas represent knowledge of a given stimuli (i.e. a specific threat) and its corresponding attributes. The absence of specific attributes (e.g. intent) is thought to be unproblematic for schemas as these are simply assumed to exist to avoid cognitive dissonance on the part of a decision-maker. Real-world cases such as the cyber operation targeting the Pyeongchang Winter Olympics demonstrates this tendency. Barring additional evidence, it was initially assumed to have originated from North Korea given the underlying strategic environment on the Korean peninsula.

Conclusions: grappling with bias and misperception

While research on the effects of micro-level attributes on strategic choices in cyberspace continues, the above observations hint at the importance of moving beyond systemic and technological aspects of interstate cyber interactions. Although escalation resulting from cyber operations has not occurred, its absence ought not to invite complacency on our part. As cyberspace serves as a cornerstone of state power, its exploitation for strategic ends is likely to continue. Moreover, with the decision to exercise power ultimately resting in the hands of an individual or small groups, a better understanding of their cognitive processes is necessary if stability in cyberspace is to be achieved.

Given that the cognitive mechanisms discussed above operate below the level of consciousness, it would be difficult to explicitly restrict its usage among decision-makers. Instead, increasing domain expertise among elites is an appropriate step in mediating biased reasoning. For the studies mentioned, domain expertise appears to move decision-making closer to the normative expectations of rational choice. That is to say, while heuristics usage continues to persist, knowledge aligns these with empirical realities. This finding is consistent with the body of research in the field of cognitive psychology which confirms that individuals with greater domain expertise are those best suited to effectively employ heuristics.

The need to better educate elites, however, is not a novel idea[19]. However, more effort is necessary to increase domain expertise among elites given the use of cyber operations as an instrument of foreign policy. Alongside on-going efforts to develop cyber norms, taking this step is a move in the right direction in maintaining the stability of cyberspace and warding off unnecessary disputes due to misperception and biased reasoning.

[18] Ryan C. Maness and Brando Valeriano (2016), “The impact of cyber conflict on international interactions”, *Armed Forces & Society*, 42(2), 301-323, https://journals.sagepub.com/doi/abs/10.1177/0095327X15572997.

[19] Lene Hansen and Helen Nissenbaum (2009), “Digital disaster, cyber security, and the Copenhagen School”, *International studies quarterly*, *53*(4), 1155-1175, https://nissenbaum.tech.cornell.edu/papers/Digital%20Disaster.pdf.