When Less is More: Cognition and the Outcome of Cyber Coercion

Miguel Alberto Gomez

The rise of offensive interstate cyber interactions continues to fan interest in the coercive potential of cyber operations. Advocates of this revolutionary view insist that it signifies a shift in the balance of interstate relations; yet empirical evidence from past cases challenges these beliefs as actions often result in continued resistance rather than compliance. Regardless of its performance, the coercive potential of cyber operations cannot be readily dismissed. Consequently, the paper advances that the outcome of coercive cyber operations is better explained using heuristic decision-making strategies rather than normative approaches such as expected utility.

Keywords: Cognitive heuristics, expected utility, coercion, cyberspace

Introduction

On December 23, 2015, a cyber operation disabled over fifty power substations in western Ukraine leaving over 230,000 residents without electricity. This incident marked the first case of a cyber incident resulting in the disruption of a state's power grid.[1] With the Ukrainian-Russian conflict well into its third year, the notion that similar events serve as adjunctive coercive tools in times of dispute is further reinforced.[2]

Miguel Alberto Gomez is a senior researcher at the Center for Security Studies, ETH, Zurich and a PhD candidate at Cardiff University, Wales.

[1] Kim Zetter, "Inside the Cunning, Unprecedented Hack of Ukraine's Power Grid," *WIRED*, March 3, 2016, https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/.

[2] SANS-ICS, "Analysis of the Cyber Attack on the Ukrainian Power Grid," *SANS*, March 18, 2016, https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf.

Building on propositions from several authors, the rate at which politics, the economy, and the larger global society are increasingly dependent on cyberspace potentially magnifies the perceived threat.[3] This appears to empower the exercise of cyber coercion by increasing the potential cost of non-compliance with threats against the underlying cyber infrastructure. Yet despite these claims, such cases have performed poorly, with adversaries opting to resist rather than comply with an aggressor's demands.[4] Furthermore, even technically advanced operations have not resulted in significant policy shifts.[5] While critical voices attribute its lackluster performance to inherent domain limitations, the strategic utility of cyber coercion should not be readily dismissed. As noted by Gartzke and Lindsay, "the potential of cyberspace is more limited than generally appreciated, but is not negligible."[6] Thus, the continued use by states of coercive cyber operations merits further inquiry.

Consequently, this paper shifts away from the prevailing view that the success or failure of coercive cyber operations results from normative decision-making strategies through which the decision to comply or resist is a function of expected gains or losses. Instead, cognitive heuristics offers a clearer insight as to why states behave as they do contrary to the expectations of "more rational" strategies. While the parsimonious account offered by variants of the rational choice paradigm simplifies our understanding of this complex environment, the inclusion of a cognitive dimension reflects the need to seek narratives that better illuminate the phenomenon of cyber coercion. In so doing, the paper acknowledges the urgency raised by Dean

[3] Myriam Dunn-Cavelty, "From Cyber-Bombs to Political Fallout: Threat Representations with an Impact in the Cyber-Security Discourse," *International Studies Review* 15, no. 1 (2013): 105–122; Jon R. Lindsay, "Stuxnet and the Limits of Cyber Warfare," *Security Studies* 22, no. 3 (2013): 365–404; Jon R. Lindsay and Erik Gartzke, "Coercion through Cyberspace: The Stability-Instability Paradox Revisited," in *The Power to Hurt: Coercion in the Modern World*, ed. Kelly Greenhill and Peter Krause (New York: Oxford University Press, 2016).

[4] Benjamin M. Jensen, Brandon Valeriano, and Ryan Maness, "Cyber Victory: The Efficacy of Cyber Coercion," (Paper presented at the Annual Meeting of the International Studies Association, Atlanta, GA, 2016).

[5] Emilio Iasiello, "Cyber Attack: A Dull Tool to Shape Foreign Policy," in *Fifth International Conference on Cyber Conflict*, ed. Karlis Podins, Jan Stinissen, and Markus Maybaum (Tallinn: NATO CCDCOE, 2013), pp. 451–468.

[6] Lindsay and Gartzke, "Coercion through Cyberspace."

and McDermott that an understanding of state behavior in cyberspace rests on the interaction of factors across different operational levels.[7]

Therefore, the paper serves as a plausibility probe to demonstrate the suitability of cognitive heuristics as a valid decision-making strategy in response to coercive cyber operations. In so doing, the paper is divided into four key sections. The first provides a brief overview of coercion in the context of cyberspace. This is followed by a critique of the prevailing account that cyberspace is a domain of risk that results in the misaligned application of expected utility in interpreting state response to cyber coercion. In its place, cognitive heuristics is offered as a viable alternative with the understanding that decisions emerge from the exploitation of the unique statistical characteristics of cyberspace using frugal cognitive processes. The suitability of this approach is then explored through a plausibility probe of the Stuxnet campaign. Finally, the paper concludes with the possible limitations of this theoretical framework.

Coercion and Cyberspace

For the past two decades, strategic interest in cyberspace has been encouraged by the growth and pervasiveness of the underlying cyber infrastructure.[8] These developments, however, are overshadowed by fears of exploitable vulnerabilities within these systems and sub-systems that reinforce the belief of aggressors employing denial or punishment strategies with coercive intent.[9] This highlights the domain's inherent vulnerability relative to its socio-political and economic value, thus portraying a future in which exercising cyber power—manifested in cyber operations—serves as a principal coercive instrument for actors capable of employing it. As coercion is defined as the

[7] Benjamin Dean and Rose McDermott, "A Research Agenda to Improve Decision Making in Cyber Security Policy," *Penn State Journal of Law and International Affairs* 5, no.1 (2017).

[8] Stuart Starr, "Toward a Preliminary Theory of Cyberpower," in *Cyberpower and National Security*, ed. Franklin Kramer, Stuart Starr, and Larry Wentz (Washington DC: Potomac Books, 2009), pp. 43–88.

[9] Robert A. Pape, *Bombing to Win: Air Power and Coercion in War* (New York: Cornell University Press, 1996); John Stone, "Cyber War Will Take Place!" *Journal of Strategic Studies* 36, no. 1 (2013): 101–108.

use or threat of force to elicit a change in an adversary's behavior,[10] the above conditions validate the employment of cyber operations for this task. Given that the outcome is a function of possible losses or gains, threats to the underlying infrastructure that support a state's strategic interest lead to a re-evaluation of an adversary's position.

While the study of coercion in cyberspace has and continues to attract academic interest, the available literature remains scarce. Initial studies indicating the coercive potential of cyber operations reflect its purported offensive advantage. Saltzman writes that this advantage is enabled by the versatility and "byte power" of the cyber operations. He argues that versatility is the ability of actions in cyberspace to negatively impact a state's strategic interests.[11] Byte power, in turn, is the amount of damage inflicted by actions in cyberspace. Apart from these, the perceived absence of material constraints also grants cyber operations an asymmetric advantage. While access to advanced conventional (and nuclear) weapons is often constrained by economic considerations, the availability of tools via underground networks presumably offer materially deficient aggressors an advantage; yet, despite these arguments, the outcome of past cases calls into question the coercive potential of cyber operations.

Out of 164 past operations that were identified, only 64 percent resulted in observable changes of an adversary's behavior.[12] Furthermore, attempts to compel an adversary through denial were only successful approximately 1 percent of the time. If the underlying domain conditions—in conjunction with the offensive advantage offered by cyber operations—does indeed enhance the coercive potential of cyber operations, then what accounts for its dismal success rate?

Coercive Success or Failure

While the evidence suggests the limited potential of coercion through cyberspace, it does not completely discount its utility. Although the need to set expectations is merited, the factors that give rise to coercive success or

[10] Daniel Byman and Matthew Waxman, *The Dynamics of Coercion: American Foreign Policy and the Limits of Military Might* (New York: Cambridge University Press, 2002).

[11] Ilai Saltzman, "Cyber Posturing and the Offense-Defense Balance," *Contemporary Security Policy* 34, no. 1 (2013): 40–63.

[12] Jensen, Valeriano, and Maness, "Cyber Victory: The Efficacy of Cyber Coercion."

failure in cyberspace remain unidentified. For studies concerning the exercise of coercion, expected utility theory is routinely employed to evaluate state behavior. It posits that an actor's decision to resist or comply is based on the maximization and minimization of gains and losses relative to their net position. As states continue to invest in cyberspace to meet strategic objectives, coercive threats are increasingly being leveled against economic, political, social, or military goals, with the decision to comply or resist due to the (threat of) disrupting these goals.[13]

The prevailing factor supporting the coercive potential of cyber operations is the ability to exploit technological vulnerabilities.[14] A common threat representation within cyberspace is that of its vulnerabilities, unknowabilities, and inevitabilities exploited by cyber operations. Cavelty points to the conceptualization of threats originating from vulnerabilities and the extent to which systems deemed as "critical" are susceptible and adversely affected by them.[15] The interconnected nature of these systems allows individuals and organizations to continually innovate and extend their reach; however, it also magnifies the consequences in the event of exploitation. Given the complexity of these technologies and fundamental human limitations, eliminating these threats through improved product development and quality management is infeasible.

Consequently, these conditions introduce a chain of events that favors coercion through cyber operations. First is the loss of the sense of security. The complexity of the domain increases the possibility that an exploitable vulnerability exists. This fosters a notion of inevitability that an aggressor would discover this vulnerability and use it to its advantage. Finally, should this vulnerability be present in systems and sub-systems deemed as "critical," it potentially places a given society at risk.[16] In so doing, the application of expected utility theory to this scenario suggests that the likelihood of losses incurred due to coercive threats being exercised is relatively high and results

[13] Starr, "Toward a Preliminary Theory of Cyberpower."

[14] Ronald J. Deibert and Rafal Rohozinski, "Risking Security: Policies and Paradoxes of Cyberspace Security," *International Political Sociology* 4, no. 1 (2010): 15–32.

[15] Dunn-Cavelty, "From Cyber-Bombs to Political Fallout."

[16] Lene Hansen and Helen Nissenbaum, "Digital Disaster, Cyber Security, and the Copenhagen School," *International Studies Quarterly* 53, no. 4 (2009): 1155–1175; James Lewis, "National Perceptions of Cyber Threats," *Strategic Analysis* 38, no. 4 (2014): 566–576.

in compliance. The validity of this argument, however, rests not only on the recognition of this causal process and the probability of its realization but also on an adversary's ability to mitigate these threats. This presupposes that an actor in cyberspace exists in a risk-centric environment and possesses knowledge of threats, capabilities, and consequences.

The Cyberspace Environment

The literature on cyber coercion tends to conflate the notion of risk and uncertainty resulting in the inappropriate application of expected utility. Although the terms "risk" and "uncertainty" suggest a conceptual equivalency, each describes a unique information environment that influences the quality and processes of decision making. In adopting the terminology employed by Savage, risk refers to a "small world" in which the decision maker is aware of the probabilities of all possible outcomes and alternatives. In contrast, uncertainty reflects a "large world" where probabilities are not known or cannot be expressed with any mathematical certainty.[17]

If cyberspace is treated as a domain in which interconnectedness constrains the ability to predict possible points of failure and likely consequences, it then follows that decision makers operate in the context of uncertainty rather than in that of risk. In this respect, it has been shown that normative strategies (i.e., expected utility) employed in environments of uncertainty rather than risk often underperform. This issue is manifested through the bias-variance dilemma that is aggravated when normative strategies are applied to inappropriate environments.

The predictive accuracy of decision making is challenged by two important factors: bias and variance. The former refers to the extent to which a model deviates from the true state of the environment. As it is not possible to know the true state beforehand, a truly unbiased model cannot exist. The presence of bias, however, is mitigated by increasing variance through the addition of free parameters that accommodate a larger variety of true states. Doing so, however, risks overfitting and reduces predictive accuracy. Normative strategies such as expected utility offset bias with the inclusion of such parameters. This approach is suited to environments wherein exemplar cases

[17] Kirsten G. Volz and Gerd Gigerenzer, "Cognitive Processes in Decisions Under Risk are not the Same as in Decisions Under Uncertainty," *Frontiers in Neuroscience* 6, July 12, 2012.

are readily available or where these cases are not ambiguous. Barring these conditions, normative strategies may be able to accurately describe previous observation but fail in predicting future outcomes; in this case, cognitive heuristics may prove to be better suited to this task.

Heuristics are defined as "strategies that ignore part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods."[18] Compared with their normative counterparts, errors in this approach emerge solely from bias. While it seems counterintuitive to suggest that accuracy is achieved with less information, these heuristics outperform their more "rational" counterparts when exercised in uncertain environments. Take the case of investments as an example. Borges and others demonstrate that mere recognition of a company's name can be employed to build an investment portfolio with returns that are least 10 percent greater compared to other strategies.[19] In their research, there appears to be a strong positive correlation between the company and its performance in the market that is exploited by decision makers using their ability to recognize this relationship from memory (i.e., recurring media coverage of a well-performing company). Consequently, this serves as a cue to pick one company over another when building a portfolio.

Although an in-depth discussion of heuristics is beyond the scope of this paper, it is crucial to point out that the advantages exhibited by heuristics rest on the ability to exploit the statistical characteristics of an environment using inherent cognitive capabilities such as memory. In other words, heuristics are only as accurate as the extent to which they fit existing structures.[20] This is otherwise known as ecological rationality.

[18] Gerd Gigerenzer and Wolfgang Gaissmaier, "Heuristic Decision Making," *Annual Review of Psychology* 62 (2011).

[19] Bernhard Borges, Daniel G. Goldstein, Andreas Ortmann, and Gerd Gigerenzer, "Can Ignorance Beat the Stock Market," in *Simple Heuristics That Make Us Smart*, ed. Gerd Gigerenzer, Peter M. Todd, and the ABC Research Group (New York: Oxford University Press, 1999).

[20] Laura Martingnon and Ulrich Hoffrage, "Why Does One-Reason Decision Making Work?" in *Simple Heuristics That Make Us Smart*, ed. Gerd Gigerenzer, Peter M. Todd, and the ABC Research Group (New York: Oxford University Press, 1999).

The Ecological Rationality of Cyberspace

Environments in which heuristics are well suited to are characterized by uncertainty, redundancy, sparseness of data, and variability.[21] While earlier sections have touched upon the uncertain nature of cyberspace, this requires further elaboration. Extending Perrow's work on "normal accidents," it is argued that the connectivity and interdependency that cyberspace enables simultaneously curtails attempts to predict both the causes and effects of disruptive events. The possibility of a cascading disaster upon which the coercive potential of cyber operations is grounded would not exist without this paradoxical relationship.[22] Take, for instance, the case of a word processor. As a standalone application, security professionals are able predict the number of vulnerabilities per thousand lines of code based on their experience with similar software. In this situation, one operates in an environment of risk given the knowledge of possible vulnerabilities obtained from direct access to the underlying code and/or experience. To enhance productivity, however, users could interconnect their word processors to engage in collaborative work. In so doing, previous knowledge with respect to vulnerabilities is devalued since the state of other systems with which they connect are unknown. Consequently, it becomes difficult to predict where, when, or how failure could occur, thus placing users in an environment of uncertainty.

When applying this logic to the question of coercion, states that depend on these systems cannot predict the true extent or damage that aggressors may inflict. This inhibits an accurate assessment of the consequences of either complying or resisting coercive demands. While some argue that this, in fact, challenges the utility of coercion in cyberspace, this paper claims that this does not necessarily diminish the feasibility of cyber coercion; rather, it suggests instead that this lack of information influences the selection of an appropriate decision-making strategy when viewed in the context of other events.

Coercion in cyberspace does not exist in a vacuum and the underlying uncertainty is tempered by existing redundancies. Redundancy is the correlation between informational cues used in decision making. It is important to

[21] Peter M. Todd, Gerd Gigerenzer, and the ABC Research Group, *Ecological Rationality: Intelligence in the World* (New York: Oxford University Press, 2011).

[22] Charles Perrow, *Normal Accidents: Living with High-Risk Technologies* (Princeton: Princeton University Press, 1999).

note that coercive cyber operations often involve established rivals with a history of aggressive behavior toward one another.[23] As such, certain actions between parties are expected whether they manifest in the physical or virtual domain. Chinese cyber espionage toward the United States, for instance, is not particularly surprising and is positively correlated with China's interest in gaining an informational advantage. The WannaCry ransomware attack attributed to the North Korean regime, in contrast, does not appear to be related to their current strategic or political objectives. This demonstrates that certain events in cyberspace are framed by established interstate relations. Consequently, decision makers may exploit this relationship and their familiarity with these issues to evaluate coercive cyber operations and their consequences.

While cyberspace may be perceived as an extension of the physical domain where pre-existing relationship are continuously expressed, these events are quite rare. Therefore, information pertaining to the overall efficacy of coercive cyber operations, preferred tools and tactics, and other relevant information are sparse. Although advancements in forensic techniques have allowed a better analysis of technical characteristics, they alone provide limited strategic insight.[24] Consequently, the uncertainty that exists at the technological level is further compounded by uncertainty at the strategic/political level, thus casting greater doubt on the usefulness of coercion through cyberspace. This only appears to be the case, however, if viewed through the lens of normative approaches such as expected utility. Since decisions are made based on gains and losses, the scarcity of information should not confirm the absence of future losses nor the continued success of initial compromise since decision makers are not privy to all possible outcomes and alternatives.

Finally, the performance of heuristics depends on the weight or validity of cues within the environment. Validity is the rate by which cues can correctly discriminate between choices. For instance, has the forward deployment of ground forces in the past resulted in the compliance of the threatened state? Linking this to key tenets of coercion theory, the outcome of coercion is

[23] Brandon Valeriano and Ryan C. Maness, "The Dynamics of Cyber Conflict Between Rival Antagonists, 2001–11," *Journal of Peace Research* 51, no. 3 (2014): 347–360.

[24] Thomas Rid and Ben Buchanan, "Attributing Cyber Attacks," *Journal of Strategic Studies* 38, no. 1 (2015): 4–37.

dependent on both the ability of a coercer to exact costs on an adversary by threatening its assets and how the latter valuates those said assets. While the literature correctly assesses the first point of this argument, it rarely recognizes variations across adversaries with respect to their perception of cyberspace, which, in turn, influences the valuation of assets.[25] In other words, what may be a valid cue that predicts compliance in one case may not be the same with another, thus increasing uncertainty.

Heuristic Selection

The preceding section has established a case in which heuristics appear a viable alternative in explaining the outcome of coercion in cyberspace, given the domain's ecological rationality. First, uncertainty denies decision makers the ability to empirically assess all possible outcomes and alternatives. Second, the correlation between events in cyberspace and existing rivalries compensates for extant uncertainties and enables the use of similar cross-domain experiences to inform decisions. Third, the rarity of coercive cyber operations further inhibits the use of normative strategies as these deny decision makers points of references upon which to base their decisions on. Finally, the inability to recognize variations in cue validity results in an incorrectly specified approach. With these points in mind, the question that remains is which heuristic can best exploit these environmental structures.

The paper posits that one-reason heuristics provide insight regarding the outcome of cyber coercion. This family of heuristics performs well in cases where cue validities vary highly, significant redundancy exists, and data is scarce.[26] If this family of heuristics is employed in deciding whether to comply or resist coercive demands, the decision-making process proceeds in accordance with search, stopping, and decision rules. These rules govern the search of appropriate cues, the conditions that leads to the cessation of the search, and the way these cues are employed resulting in a specific decision.

As simple as heuristics may be, these have been shown to outperform more complex strategies such as multiple regression, neural networks, and so forth. However, it is important to establish that this strategy is non-

[25] Forrest Hare, "The Cyber Threat to National Security: Why Can't We Agree," in *Conference on Cyber Conflict Proceedings*, ed. Christian Czosseck and Karlis Podins (Tallinn: CCD COE, 2010).

[26] Gerd Gigerenzer, "Why Heuristics Work," *Perspectives on Psychological Science* 3, no. 1 (2008): 20–29.

compensatory in that it avoids looking for conflicting evidence and relies on a subjective rather than objective assessment of a given situation. This may prove to be troublesome, if not dangerous, in certain environments. For instance, a false flag operation by a third party that mimics the behavior of one rival may result in unintended escalation under the right circumstances.

The Viability of Heuristics: Stuxnet

To support the preceding theoretical arguments, the feasibility of heuristics is demonstrated with a plausibility probe. Although several events since 2007 may serve this purpose, the paper employs the often-used case of Stuxnet that has been attributed to both the United States and Israel. The decision to do so is due to the availability of information pertaining to this case that allows for a comparison of the two decision-making strategies to be made.

The interaction between the United States and Iran in cyberspace is characterized as a series of coercive acts of varying intensity, severity, and scope.[27] Of these, Stuxnet remains the most prominent case of cyber coercion. The existence of Stuxnet first came to light in June 17, 2010 when the Belarusian anti-virus company VirusBlokAda was approached to respond to unknown system reboots occurring in Iran.[28] Despite its "initial" discovery in 2010, analysts believe that it had been operational as early as June 2009 with ten initial infections affecting five organizations within Iran and resulting in a total of 12,000 infections by the time it was identified in 2010. Its advanced feature set suggests the involvement of state or state-funded organizations in its development and eventual release. This gave it the recognition as being the first "weaponized" malware in history. Moreover, its feature set and targets (Industrial Control Systems) signaled a shift in capability, complexity, and intent of actors within cyberspace.[29] By the time the infection had been contained, over 1,000 nuclear centrifuges used for

[27] Jason Healey, "Winning and Losing in Cyberspace," in *Eighth International Conference on Cyber Conflict*, ed. Nikolaos Pissanidis, Henry Roigas, and Matthijs Veenendaal (Tallinn: CCD COE, 2010).

[28] Kim Zetter, "An Unprecedented Look at Stuxnet, the World's First Digital Weapon," *Wired*, March 11, 2014, https://www.wired.com/2014/11/countdown-to-zero-day-stuxnet/.

[29] Jon R. Lindsay, "Stuxnet and the Limits of Cyber Warfare," *Security Studies* 22, no. 3 (2013): 365–404.

uranium enrichment had been damaged, and the discourse regarding the use of cyber weapons had entered a new era.

Yet despite its operational characteristics, Stuxnet failed to coerce the Iranian regime in ending its nuclear enrichment program. The prevailing sentiment is that Stuxnet's failure stemmed from the limited damage inflicted against Iran's enrichment infrastructure. Post-incident analysis revealed that the number of centrifuges affected did not exceed normal operational wear-and-tear, and this account appears consistent with our understanding of coercion viewed through the lens of expected utility theory. In other words, the damage did not reach disruptive or debilitative levels that would prompt a reassessment of policy. Yet for this argument to hold, one must allow for one crucial assumption: that the Iranian regime had adequate knowledge of their capabilities and vulnerabilities in cyberspace, providing the confidence to risk further attempts against their cyber infrastructure. If true, this implies that the decision to resist was made in an environment of risk. The Iranian response, however, challenges this at an empirical and theoretical level.

While it is unreasonable to assume that those responsible for Iranian cyber security had perfect knowledge of all the possible attack vectors, a suitable security program would at least have taken steps to mitigate viable threats as informed by both first-hand experience and publicly available knowledge. Without direct involvement or insight into their internal processes, this readiness is deduced from behavior once a threat is realized. In the case of Stuxnet, reports that Iranian authorities had resorted to external third parties to better understand the unusual behavior of their systems suggests that a Stuxnet-like event had not been anticipated nor its consequences considered. Through no fault of their own, the complexity of Stuxnet had no precedence from which computations of possible losses could be derived.

Although it may be argued that additional information regarding the capabilities and damage potential of Stuxnet could have surfaced as the investigation proceeded, this implies the existence of both technological expertise and established organizational structures in support of such endeavors. Organizations require a mechanism that enables the synthesis of information across different units to understand the full implications of these events. Furthermore, the existence of such a structure cannot be assumed across states nor is their efficacy a foregone conclusion.[30] Iranian dependence on external aid during the incident, along with earlier reports of their cyber capabilities, calls into question their ability to fully comprehend the consequences of Stuxnet and further challenges the applicability of normative strategies that explain their decision to resist.

Finally, if the Iranian regime was indeed confident in their ability to defend against Stuxnet or further acts of coercion, then why had there not been a stronger response? Both Gartzke and Lindsay argue that operations that result in compromise but are eventually contained end in an escalatory spiral.[31] Though less extreme, the game theoretic model of Edwards and others suggests that those aware of their vulnerabilities and who have mitigated them should at least publicly attribute coercive acts to their rivals.[32] Neither had transpired with respect to Stuxnet. Although some analysts claim that later Iranian cyber operations were such a response, their operational characteristics do not appear to be proportionate nor tailored to serve as a reply to Stuxnet.

The prevailing account of Stuxnet's failure, while seeming to confirm the usefulness of normative strategies, stands on unstable ground upon closer inspection. Although speculative without first-hand information, it appears that the Iranian regime did not have a full understanding of their own vulnerabilities. Consequently, it would not have been appropriate for decision makers to rely on expected utility or its related strategies to frame their response given that information regarding the possible consequences of resisting or complying were either incomplete or unavailable. Furthermore, the feasibility of normative strategies is challenged further in other cases of cyber coercion. The "BoxingRumble" operation against Chinese cyber espionage, for instance, did not result in significant damage either; nonetheless, Chinese operations were halted for the time being in response.[33] This apparent

[30] Rebecca Slayton, "What is the Cyber Offense-Defense Balance?" *International Security* 41, no. 3 (2017): 72–109.

[31] Erik Gartzke and Jon R. Lindsay, "Thermonuclear Cyberwar," *Journal of Cybersecurity* 3, no. 1 (2017): 47–48.

[32] Benjamin Edwards, Alexander Furnas, Stephen Forrest, and Robert Axelrod, "Strategic Aspects of Cyber Attack, Attribution, and Blame," in *Proceedings of the National Academy of Sciences* (forthcoming).

[33] Sean Gallagher, "NSA secretly hijacked existing malware to spy on N. Korea, others," *arsTechnica*, January 19, 2015. https://arstechnica.com/information-technology/2015/01/nsa-secretly-hijacked-existing-malware-to-spy-on-n-korea-others/.

contradiction appears to question the validity of conclusions established through normative strategies.

Filling the Gap

If normative strategies such as expected utility are not suited for this environment, would there be reason to believe that heuristics could do better? Extending the argument that cyber coercion occurs between rivals and that the environmental structures favor one-decision heuristics, this assumption is demonstrated using the "Take The Last" (TTL) heuristic.

The TTL heuristic functions by employing a strategy known as an *Einstellung* set. Psychologists since the 1930s have observed that individuals solve seemingly related problems with strategies that had worked in the past.[34] This assumes that the TTL heuristic is invoked in environments where decisions are frequently made about events that are correlated with one another in some form. This correlation is indirectly manifested in the ability of the decision maker to recognize similarities between different tasks; however, recognition in this case is not necessarily equivalent to memory but rather refers to the intuitive characteristics of events that are reinforced through constant exposure.

Since coercive cyber operations involve established rivals that routinely interact with one another, TTL is an ideal strategy not only because of environmental structures but also of its efficiency. Unlike expected utility that requires intensive computation, which increases cognitive load, TTL relies merely on recognition to identify alternatives. Furthermore, in time-critical situations such as interstate disputes, the speed with which TTL is exercised makes it a preferable choice over alternative strategies. Thus, TTL proceeds as follows: search for the cue that stopped the search during the last related problem; compare the validity of the cue relative to the alternatives. If it discriminates, use the cue; otherwise, go back to the problem before the last and determine which cue stopped that search.

In explaining the outcome of Stuxnet using the TTL heuristic, the process begins by building a repository of all the similar events in the past. Since

[34] Gerd Gigerenzer and Daniel G. Goldstein, "Betting on One Good Reason: Take the Best Heuristic," in *Simple Heuristics That Make Us Smart*, ed. Gerd Gigerenzer, Peter M. Todd, and the ABC Research Group (New York: Oxford University Press, 1999).

the target of Stuxnet had been systems-controlling nuclear centrifuges responsible for enrichment, the repository most likely contained previous attempts to coerce Iran into stopping its nuclear program. This assumption is not necessarily tenuous given the amount of effort invested by its rivals who achieve just that. Furthermore, the fact that this occurred in cyberspace should not challenge the ability of decision makers to recognize similarities since the objective in question remains the same (i.e., ending the nuclear program).

Once this mental repository is constructed, the decision maker needs to identify the last instance when the cue discriminated between alternatives. Since first-hand accounts are unavailable, this paper turns to a timeline of coercive events prior June 2010. Despite the existence of on-going talks between 2006 and 2010, the United Nations Security Council imposed a total of six sanctions intended to disrupt the nuclear enrichment program. Apart from this, the United States had also begun to seriously consider air strikes while Israel threatened military action. While it is impossible to determine which of these events was used as a reference point, it should not matter since the outcome had been the same on the part of Iran: resist.[35]

Given that the context that framed Stuxnet and a similar event in the past, it is likely that decision makers opted to remain consistent with their defiant behavior. The characteristics of Stuxnet would have limited the accuracy of more complex decision-making strategies given the lack of information regarding its true capabilities and the extent of compromise. Furthermore, if resistance had worked when the threat was greater (i.e., thoughts of actual physical confrontation), then it should also suffice in this less extreme situation.

The Way Forward

Over the course of several pages, this paper has built an argument in support of cognitive heuristics as an analytical tool to evaluate the outcome of coercive cyber operations. Although normative strategies remain the mainstay for evaluating state behavior, the unique characteristics of cyberspace calls its adequacy into question. Whereas experience in the physical domain permits the objective evaluation of gains and losses, the uncertainty endemic to

[35] Shreeya Sinha and Susan Campbell Beachy, “Timeline on Iran’s Nuclear Program,” *New York Times*, April 2, 2015, https://www.nytimes.com/interactive/2014/11/20/world/middleeast/Iran-nuclear-timeline.html?_r=0#/#time243_10809.

cyberspace limits the predictive accuracy of expected utility and related strategies. In its place, fast-and-frugal strategies such as Take The Last heuristic provide a more robust account of coercion in this virtual domain.

Depending on heuristics, however, is not a foregone conclusion. As there is no such thing as a one-size-fits-all decision-making strategy, the performance of either heuristics or normative strategies is a function of both environmental structures and individual cognitive capacities. This interdependence is best expressed in Herbert Simon's analogy of rationality as scissor blades where one blade represents the cognitive limitations of individuals while the other represents environmental structures and conditions. In as much as a pair of scissors cannot work with just one blade, our understanding of rationality cannot be limited to one aspect or the other.

Consequently, three important points are raised. The first is that the use of cognitive heuristics in the domain of interstate relations need not be framed as a failure of rationality. Despite recent findings in cognitive psychology, scholars in international relations and political science continue to frame cognitive heuristics as low-cost strategies that result in sub-optimal decisions. This paper instead has highlighted the importance of fitting strategies to the appropriate environment, as even complex approaches can result in poor outcomes if used incorrectly.

Second, despite the performance of heuristics in evaluating coercive outcomes in cyberspace, these results are not generalizable across all forms of cyber interactions. While it does appear that heuristics perform better when explaining cyber coercion, it does so because environmental structures are efficiently (and correctly) exploited by underlying cognitive processes. These conditions, however, may not exist in cases of disruptive cyber operations that form much of the interactions in cyberspace. For these, the environment of uncertainty gives way to one of risk due to the well-documented effects of the tools and tactics employed. This consequently enables the use of normative strategies that can better exploit the available information.

Third, decisions in the face of crisis cannot be assumed to emerge from the thoughts of a single individual; unique organizational dynamics contribute to the nature of the decisions made. Furthermore, other factors, such as audience costs that are not addressed in this paper, may also be significant with respect to responding to coercion. This is worth noting given the salience of issues that color various interactions in cyberspace.

The field of cyber security is still in its infancy. Yet with threats evolving both in terms of complexity and scope, there is urgency for academics and policy makers alike to understand state behavior in response to events within cyberspace. This paper contributes to this endeavor by offering an avenue of analysis that has rarely been considered by those in the field but whose insight can assist in maintaining stability within this virtual domain.