ARI 102/2018
11 September 2018

Coercion and Cyberspace

Miguel Alberto Gomez | Senior researcher at the Center for Security Studies, ETH Zurich | @mgomez85

Theme

Cyberspace is a new domain for coercive operations in support of foreign policy and security with advantages for offensive actions and hindrances to its success.

Summary

This ARI provides an overview of factors crucial in our understanding of coercive cyber operations as the exercise of power through cyberspace in order to coerce an adversary into a particular course of action. It its focused on the compellent actions of the state actors though they, and non-state actors, may carry out deterrent actions as well. The first section presents the fundamentals of coercion. The second frames coercion in the context of cyberspace and surfaces the characteristics of the domain that enables it. Finally, the third establishes the causes behind coercive failure and, inversely, success.

Analysis

Over the past decade, cyber operations are increasingly employed as coercive instruments of foreign policy. From the Bronze Soldier incident between Russia and Estonia in 2007 to the long-standing dispute on the Korean peninsula, cyber operations are exercised in hopes of altering an adversary's behavior. Yet despite such optimism, less than 5% of these have achieved their intended objectives.[1] Paradoxically, states continue to engage in coercive behavior in cyberspace despite its seeming inefficacy. This raises two important questions. First, how are cyber operations instruments of coercion? Second, what accounts for their limited outcomes?

Coercive cyber operations are not exempt from principles that enable coercive interstate behavior. Commonly understood as *"the threat of damage, or of more damage to come that can make someone yield or comply"* (Schelling, 1966). Unfortunately, the concept is muddled by a lack of a clear operational definition. Typically, characterizations proposed by either Schelling or George (1991) are often adopted.[2] And while most agree that deterrence refers to the use of threats to coerce an adversary from engaging in an

[1] Iasiello, E. (2013). Cyber attack: A dull tool to shape foreign policy. In K. Podins, J. Stinissen & M. Maybaum (Eds.), *2013 5th International Conference on Cyber Conflict (Cycon)*, 451-470.

IEEE. Jensen, B., Maness, R. C., & Valeriano, B. (2016). Cyber Victory: The Efficacy of Cyber Coercion *Annual Meeting of the International Studies Association*. Valeriano, B., & Maness, R. C. (2014). The dynamics of cyber conflict between rival antagonists, 2001-11. *Journal of Peace Research 51*(3), 347-360.

[2] Schelling, T. C. (1966). *Arms and Influence*. New Haven, CT: Yale University Press. George, A. L. (1991). *Forceful persuasion: Coercive diplomacy as an alternative to war*. US Institute of Peace Press.

undesired action, the debate centers on whether the threat or limited use of force to alter an adversary's behavior ought to be referred to as compellance or coercive diplomacy.

Schelling treats compellence as *"a threat intended to make an adversary do something"* and does not distinguish between a reactive or proactive use of force in order to influence an adversary's behavior. He assumes the presence of a unitary rational actor behaving in a manner that maximizes gains while minimizing losses. George, in contrast, frames coercive diplomacy as a narrower and reactive response to an adversary's actions. Whereas Schelling offers a parsimonious account grounded in rational choice, George offers a more nuanced and context-dependent explanation of the phenomenon. In recent years, a growing number of studies have started to use the term (military) coercion in place of either compellence or coercive diplomacy.[3]

With respect to coercive cyber operations, the umbrella term of coercion suits this phenomenon for three reasons. First, the proactive or reactive nature of compellence fits the image of cyber tools being preemptively deployed on an adversary's system. Fears of such may convince an adversary to reconsider its actions. Second, coercive cyber operations often take place during on-going regional disputes.[4] Its employment as one in a handful of instruments (i.e. military threats, economic sanctions, etc.) highlights the primacy of strategy in its use and, consequently; the importance of context as suggested by George. Finally, the restraint with which cyber capabilities are exercised reflects a degree of rationality on the part of coercers.

Yet despite its conceptual simplicity, coercive success is difficult to achieve. The outcome of coercion is contingent on the clear communication of a threat, suitable cost-benefit calculations, the credibility of the coercer, and reassurances from the coercer upon compliance. Although George identifies a host of other factors that contribute to the outcome of coercion, these may be consolidated into the above.

Unambiguous communication is the cornerstone of successful coercion. Adversaries must know what behavior needs to be modified, the time in which these needs to occur, and the costs/threats associated with compliance or resistance. Yet reality poses difficulties in clearly communicating threats. Systemically, the anarchic nature of the international system can result in misperception between states. Fearon (1995) posits that fragmentary information encourages misrepresentation and an excess in confidence during periods of conflict that increases the possibility of war and, consequently, coercive failure.[5] Complementing this, cognitive biases may also encourage a breakdown in communication. Research demonstrates the use of pre-existing schemas in the formation of decisions regarding the behavior of other states.[6] And while this tool serves

[3] Jakobsen, P. V. (2006). Coercive Diplomacy. In Collins, A. *Contemporary Security Studies*. Oxford: Oxford University Press, 225-247.

[4] Valeriano, B., & Maness, R. C. (2015). *Cyber war versus cyber realities: cyber conflict in the international system*. Oxford; New York: Oxford University Press.

[5] Fearon, J. D. (1995). Rationalist Explanations for War. *International Organization, 49*(3), 379-414.

[6] Herrmann, R. K., Voss, J. F., Schooler, T. Y. E., & Ciarrochi, J. (1997). Images in international relations: An experimental test of cognitive schemata. *International Studies Quarterly, 41*(3), 403-433.

(cont.)

to mitigate cognitive limitations, it increases the possibility of bias that results in misperception sub-optimal judgements.

Successful coercion assumes the presence of a rational actor capable of evaluating the costs and benefits associated with resisting or conceding to a coercer. Although the importance of costs and benefits in determining the outcome of coercion is straightforward, a number of factors can influence a breakdown of this process. Systemically, two (2) complementary factors that result in such a failure are conspicuous compliance and the possibility that this invites further demands.[7]

Initially forwarded by Schelling, conspicuous compliance is rooted in the argument that *"the very act of compliance – of doing what is demanded – is more conspicuously compliant, more recognizable as submission under duress, than when an act is merely withheld in the face of a deterrent threat. Compliance is likely to be less casual, less capable of being rationalized as something that one was going to do anyhow."* Phrased simply, the act of conceding signals the weakness of an actor. Within an anarchic system in which each state is poised to ensure its own survival, such a situation is not beyond reason and leads to the second point – complying with a previous demand can invite additional demands in the future.

As Schelling argues, *"compellent threats tend to communicate only the general direction of compliance, and are less likely to be self-limiting, less likely to communicate in the very design of the threat just what, or how much, is demanded.... The assurances that accompany a compellent action— move back a mile and I won't shoot (otherwise I shall) and I won't then try again for a second mile—are harder to demonstrate in advance [than with deterrence], unless it be through a long past record of abiding by one's own verbal assurances."* Although this statement highlights key differences between compellence and deterrence, its core argument continues to cite the possibility that compliance with earlier threats does not guarantee the absence of future threats. Other actors may perceive previous concessions as an opportunity to improve their current standing with the international.

Apart from systemic factors that impinge on cost-benefit considerations, individual cognitive processes similarly affect the outcome of coercive threats. Prospect Theory which posits that losses are valued more than gains cause decision-makers to resists rather than comply even if the cost of doing the former is much higher than the latter.[8] Additionally, coercion may also fail when the coercing actor incorrectly recognizes an adversary's values and thus fails to impose a credible threat that results in the require cost-benefit calculations.

Besides clear communication and the imposition of costs, the outcome of coercion is further determined by the capability and resolve of the coercer to follow through. Talk is cheap and coercers must be able to demonstrate their ability to carry out threats should their demands not be met. While both capability and resolve are difficult to assess, the latter is particularly challenging. A coercer may fail to follow through with a threat for a

[7] Schaub, G. (2004). Deterrence, compellence, and prospect theory. *Political Psychology, 25*(3), 389-411.

[8] Kahneman, D. (2011). *Thinking, fast and slow* (1st ed.). New York: Farrar, Straus and Giroux.

number of reasons. These include, but are not limited to, grandstanding, lack of domestic support, or past failures to carry out threats. To demonstrate resolve, coercers resort to costly signaling that binds them to follow through with their intended actions.

Costly signals can be done in one of two ways. First, states may choose to tie their hands and force themselves into a specific course of action should their demands not be met. Second, states can incur sunk costs. Examples of which include the forward deployment of armed forces to the border or severing diplomatic relations with their adversaries. Either method, however, is not without risk. Costly signaling increases the possibility of armed conflict by forcing states into an inflexible course of action.[9] The idea being that the adversary realizes this possible outcome and would, in a timely manner, concede. This is predicated, however, on how well these signals are interpreted and the outcome of the cost-benefit analysis.

Lastly, the coercer must be able to reassure an adversary that compliance results in the threat being rescinded. Relatedly, coercers must be able to provide an adversary a means with which to comply that minimizes damage to its reputation. Great powers, however, find this last requirement challenging given their inherent capabilities as these, paradoxically, reduce their credibility in the eyes of weaker adversaries. Power imbalances in favor of the coercer may be interpreted as a justification for further demands despite previous concessions. Thusly, an adversary may find that resistance is a better course of action in the face of coercive threats.

Cyber Coercion: An Overview

If coercion is the exertion of pressure on an adversary by threatening something of value, then cyberspace is an ideal medium given its growing strategic value.[10] Over the past decade, (broadband) connectivity has nearly tripled globally. Similarly, Information Communication Technology (ICT) usage has grown rapidly over the past decade (ITU, 2016). Although greater awareness, education, and improvements in developmental processes have mitigated certain vulnerabilities, these continue to persist within critical systems. Fortunately, contextual factors such as the unique implementation of cyber infrastructure across states and the resources required to inflict persistent damage tempers such concerns. Yet regardless of such reassurances, the fundamental structure of cyberspace assists, if not enables, coercive behavior.

Cyberspace is treated as consisting of three key layers: physical, syntactic, and the semantic layer.[11] The physical layer consists of hardware components that store, process, and transmit electrical, optical or radio signals. Within this layer, vulnerabilities are subject to physical and environmental constraints such as the susceptibility to theft or the susceptibility to noise within the electromagnetic spectrum. A step above this is the syntactic layer through which the representation, processing, storage, and transmission of data is governed by pre-defined rules or protocols. These serve to

[9] Fearon, J. D. (1997). Signaling foreign policy interests: Tying hands versus sinking costs. *Journal of Conflict Resolution, 41*(1), 68-90.

[10] Kuehl, D. T. (2009). From Cyberspace to Cyberpower: Defining the Problem. In F. D. S. Kramer, Stuart H.; Wentz, Larry (Ed.), *Cyberpower and National Security*. Dulles: Potomac Books, 24-42.

[11] Libicki, M. C. (2009). *Cyberdeterrence and cyberwar*. Santa Monica: Rand Corporation.

provide the desired functionality and to ensure interoperability between manufacturers. Vulnerabilities exist through flaws in the implementation of these protocols that may lead to unplanned and undesired outcomes. Finally, the semantic layer presents the data in a form that is interpretable and useful to users. At this layer, the conceptualization of cyberspace varies greatly.

From what has been termed as the "western consensus", cyberspace ceases when the information serves defined strategic goals such as economic growth. On the other hand, other actors extend cyberspace to include the mental processes of individuals such that both perception and behavior are influenced by available information, thus introducing another source of vulnerability.[12] Yet regardless of this variation, it is important to note that each layer is dependent on the other for cyberspace to function. Consequently, this interdependence enables the exploitation of cyberspace to meet strategic objective.

For advocates of coercive cyber operations, arguments are often grounded on the offensive advantage offered by the domain. An offensive advantage is defined as an instance in which new technologies skew the balance of the difficulty between conducting offense or defense in favor of the former. Specifically, new technologies are thought to increase the mobility and damage potential of offensive weapons vis-à-vis defensive ones. For instance, the creation of the machine gun or the development of combat aircraft are thought to provide aggressors with the above advantage. The interconnectivity between the components of cyberspace conceptually grants these advantages. The linkage between the physical, syntactic, and semantic layer results in the disruption of a lower layer to adversely affect those above it. Cutting an undersea cable, for instance, prevents the transmission, processing, and receipt of information at the higher levels. Similarly, the corruption of data at the syntactic layer prevents the proper use of it at the semantic level.

In parallel to this cascading effect, the consequences are also magnified from layer to layer. The loss of communication from a cut cable would immediately result in the disruption of communication, at the first two layers. But at the semantic layer, the loss of information may adversely affect specific strategic objectives, the severity of which increases over time. Consequently, the coercive potential of cyber operations is contingent upon its ability to (1) cascade damage across layers, (2) the magnification of consequences, and (3) the persistence of the threat. And while offensive tools are accessible, those meeting these criteria requires organizational maturity and significant economic resources.

While a standardized taxonomy of cyber operations remains elusive, actions in cyberspace may be categorized based on intent: disruptive, espionage, and degradative. As implied by its name, disruptive cyber operations aim to disturb the routine functions of its target. Examples of these include website defacement and (Distributed) Denial-of-Service. These operations do not require a significant amount of expertise or resources to execute as the tools required are readily available. Consequently, its ease of use comes at the cost of its reduced severity and lack of persistence as these threats are

[12] Giles, K., & Hagestad, W. (2013). Divided by a Common Language: Cyber Definitions in Chinese, Russian and English. *2013 5th International Conference on Cyber Conflict (Cycon)*

easily identified and contained. In contrast, espionage operations are meant to be persistent so as to allow the exfiltration of privileged information. As with its real world namesake, these provide aggressors with an informational asymmetry over adversaries that may result in a strategic advantage in times of conflict. The use of this information to threaten an adversary, however, has a relatively long time horizon that limits its coercive value. Finally, degradative cyber operations are those intended to damage or destroy an adversary's cyber infrastructure for the purpose of inhibiting their strategic interests. These operations rely on the growing importance of cyberspace in sectors such as the military, the economy, and other public services. Actions within this category are designed to cause cascading effects with both technical and strategic consequences. Consequently, degradative cyber operations are ideal for coercion. The case of Stuxnet proves this point.

The features of Stuxnet allowed it to meet the 3 criteria previously established. While it operated within the syntactic layer of the systems controlling Iranian nuclear centrifuges, it managed to affect both the physical and semantic layers as well. By manipulating the rate with which these devices spun, it was able to inflict physical damage on the hardware. Similarly, by manipulating the protocols within the system it was able to send false information (semantic) to operators suggesting that all was well; thus allowing it to persist. Strategically, the physical damage inflicted on the nuclear centrifuges limited the amount of weapons grade fissile material that was produced that, in turn, affected the nuclear weapon's program of the Iranian regime. These make Stuxnet, and its related operations, viable coercive tools – at least in theory.

In reality, however, Stuxnet and other similar operations have resulted in coercive failure despite meeting the aforementioned criteria. Despite growing technical sophistication alongside a vulnerable cyberspace, coercive cyber operations are far less successful than expected. Yet its dismal performance may have less to do with technological constraints and more with the organizational and strategic considerations associated with its execution.

Coercive Failure and Cyberspace and Its Future

To better understand the root causes of coercive failure of cyber operations, the attributes for successful coercion need to be revisited. In summary these are: clear communication of a threat, suitable cost-benefit calculations, the credibility of the coercer, and reassurances from the coercer upon compliance. While technological advancements allow aggressors to meet the contingent technological requirements for success, the above requirements are either infeasible or poorly understood in the context of cyberspace.

In order for coercion to be successful, an aggressor need to be able to clearly communicate this threat. In cyberspace, this is easier said than done. Unlike conventional means, cyber operations do not come with a return address. The attribution problem associated with cyberspace limits the ability of targets to assess the source of the operation. While cyber operations are more frequently observed in the context of the on-going dispute, uncertainty as to the identity of the aggressor muddies the message. What action should be stopped on the part of the target? Are we even certain that X is the source of the operation? Questions such as these hinder the communicative exchange between the coercer and the target that, in turn, limit the efficacy of coercion as a whole. And while experience now allows targets to move beyond the question of “who was behind it?” to “what do we do about it?” the consequences of conceding or resisting remains a pressing issue. That is to say, the identity of the coercer does not alleviate other considerations with respect to coercion.

The decision to comply or resist depends on the costs and benefits associated with either course of action. In the context of cyberspace, this decision is predicated on (1) how a target perceives the domain (2) and the larger strategic picture. As previously mentioned, there is no unified definition as to what cyberspace is. Available research suggests that the value of cyberspace is based on existing worldviews.[13] Liberal regimes treat cyberspace as an enabler of economic growth and democratic values. In contrast, illiberal regimes perceive it as threat to their legitimacy. Consequently, the outcome of coercive cyber operations are contingent on the recognition and exploitation of these variations. At one end of the spectrum, cyber operations that threaten the banking sector of a target interested primarily in controlling online content will not generate sufficient cost that results in compliance. On the other end, threatening physical/bodily harm in order to limit freedom of speech in a society that values such would incur significant resistance. Over the past decade, the majority of coercive cyber operations appear to have fallen into either one of these extremes; thus resulting in failure.

Assuming that threats are clearly communicated and aligned correctly, coercers are still required to demonstrate their resolve. In the physical domain this is easily done via clearly worded threats or demonstrations of force. Within cyberspace, demonstrating these capabilities affords targets the opportunity to develop the necessary countermeasures. Although Smeets and Lin (2018) argue that signaling resolve in this manner is unnecessary and that past actions should serve as demonstrations of capability, this is not sufficient for coercers that have just begun to use the domain for this purpose.[14] Apart from burning cyber capabilities through demonstrations, other resources may be imperiled as well. Stuxnet, for instance, required not only advanced engineering skills but also demanded an existing espionage network capable of delivering the malware over an air-gaped network. Its discovery and analysis would have certainly tipped the Iranian regime of the presence of this network.

Finally, the success of coercion hinges on the ability of the coercer to provide guarantees that compliance results in the cessation of threats. While a coercer may indeed decide to stop coercive operations in exchange for compliance, this does not necessarily mean that other non-coercive operations will cease. In light of the growing importance of cyberspace, cyber espionage appears to have become a routine occurrence between states. While the activity is routinely accepted as normal interstate behavior, the tools and techniques required for both espionage and coercion (degradative cyber operations) are quite similar to one another. Consequently, discovery can result in a belief that the

[13] Hare, F. (2010). The Cyber Threat to National Security: Why Can't We Agree? *Conference on Cyber Conflict, Proceedings 2010*, 211-225. Rivera, J. (2015). Achieving Cyberdeterrence and the Ability of Small States to Hold Large States at Risk. *2015 7th International Conference on Cyber Conflict - Architectures in Cyberspace (Cycon)*, 7-24.

[14] Smeets, M., & Lin, H. S. (2018). Offensive cyber capabilities: To what ends? *2018 10th International Conference on Cyber Conflict (CyCon)*, 65-71.

target is part of a new coercive campaign despite previous concessions. The inability to discern intent from the mere presence of these tools along with previous coercive behavior fosters the belief of malicious intent on the part of the target and reduces that chances of coercive success in the future.

Conclusion

Despite advances in capabilities and its growing frequency, the success of coercive cyber operations is not a forgone conclusion. Although states are increasingly dependent on the domain in order to achieve its strategic objectives, the exercise of coercion remains subject to previous strategic considerations. While certain scholars and pundits continue to espouse its revolutionary potential, cyber operations are fast becoming perceived as an adjunctive foreign policy instrument. Rather than its independent exercise, the coming years will see cyberspace as one of many means with which states are able to achieve their stated strategic objective via coercive means.