Coercion through Cyberspace: The Stability-Instability Paradox Revisited

25 October 2014

Jon R. Lindsay **Erik Gartzke**

Prepared for Kelly Greenhill and Peter Krause, eds., *The Power to Hurt: Coercion in the Modern World* (working title)

1 Introduction

Information technology is the nervous system of the global economy. Critical infrastructure for banking, power, transportation, and industry increasingly depends on embedded computers connected to the internet. Firms and citizens entrust vital personal, medical, and financial data to distant servers in return for more convenient and efficient services. Military command and control relies on digital networks to connect far-flung surveillance and strike systems and to project power rapidly and precisely. Yet this vital interconnectivity also facilitates new modes of crime, protest, espionage, and warfare. Ubiquitous computer networks both provide access to valuable targets and become targets themselves. Protecting and influencing cyber infrastructure has thus become a major priority for governments and other political actors around the world.

The very ubiquity of information technology makes the danger of cyber threats easy to exaggerate. In contemporary defense policy discourse there are three influential narratives of mounting cyber peril. The most dangerous envisions the paralysis of industrial control systems of military command and control through surprise attack by anonymous hackers. This scenario is often described as a "digital Pearl Harbor" or a "cyber 9/11" depending on whether the imagined aggressor is a revisionist state like China or Iran or a non-state anarchist or terrorist empowered

by the information revolution. A second narrative offers an alternative to the shock of sudden catastrophe by warning of the long term erosion of economic and military competitiveness. The relentless theft of vital secrets stored on corporate and government networks is thus thought to cause a prolonged "death by a thousand cuts." In both of these scenarios, weaker states and terrorists gain increasing access to powerful hacking tools while technology-dependent advanced industrial states become increasingly vulnerable to cyber attack and exploitation.[1] A third category of threat narrative concerns the transformation of internet architecture to decisively benefit one political group at the expense of the other. At one extreme, the growth of flexible social media enables connected protesters to overwhelm and overthrow authoritarian regimes.[2] At the other extreme governments censor and reconfigure the internet to undermine innovation and freedom. State paranoia about paralysis and erosion thus leads to digital lockout or "the end of the internet" as we know it.[3]

National security officials, the defense industry, and media pundits all have incentives to exaggerate the cyber threat.[4] The secrecy of cyber operations further complicates assessment, even as states make major investments in cyber defense. Each of the three narratives above are indeed exaggerations, but they point toward more plausible scenarios using cyber operations as subtle complements to or even substitutes for more traditional forms of aggression.

---

[1] Richard A Clarke and Robert K Knake, *Cyber War: The next Threat to National Security and What to Do about It* (New York: Ecco, 2010); Joel Brenner, *America the Vulnerable: Inside the New Threat Matrix of Digital Espionage, Crime, and Warfare* (New York: Penguin Press, 2011); Lucas Kello, "The Meaning of the Cyber Revolution: Perils to Theory and Statecraft," *International Security* 38, no. 2 (2013): 7–40.

[2] Larry Diamond, "Liberation Technology," *Journal of Democracy* 21, no. 3 (2010): 69–83.

[3] Jonathan L. Zittrain, "The Generative Internet," *Harvard Law Review* 119, no. 7 (May 1, 2006): 1974–2040.

[4] Myriam Dunn Cavelty, "Cyber-Terror—Looming Threat or Phantom Menace? The Framing of the US Cyber-Threat Debate," *Journal of Information Technology & Politics* 4, no. 1 (2008): 19–36; Jerry Brito and Tate Watkins, *Loving the Cyber Bomb? The Dangers of Threat Inflation in Cybersecurity Policy* (Arlington, VA: Mercatus Center at George Mason University, 2011); Sean Lawson, "Beyond Cyber-Doom: Assessing the Limits of Hypothetical Scenarios in the Framing of Cyber-Threats," *Journal of Information Technology & Politics* 10, no. 1 (2013): 86–103.

Understanding the dynamics, magnitude, and likelihood of aggression online requires greater attention to the operational requirements for staging various types of cyber operations, the strategic benefits actors hope to gain through them, and the risks of unintended consequences. Too often defenders of the cyber revolution focus narrowly on the technological possibility for harm but discount operational and institutional obstacles to effectiveness and ignore the strategic utility of cyber harm or threats of harm.[5]

A realistic appraisal of cyber threats must take not only technological but also strategic logic into account. Thomas Schelling distinguishes brute force, which is needed in a contest of strength, from coercive threats, which are useful in a contest of resolve.[6] Both require the power to inflict harm, but brute force exercises it while coercion holds (at least some of) it in reserve. Likewise, actors might use cyber operations to attempt to change the balance of power directly or they might use them to provide information about their intentions and commitment. To paraphrase Clausewitz, cyberwar is politics by other means. As a result of technical and political constraints, the coercive potential of cyberspace is more limited than generally appreciated, but it is not negligible, especially when exploited in conjunction with other forms power such as military force.

In this chapter we lay out a typology of cyber operations, distinguishing the skills and resources needed to cause different types of harm. Not all cyber options are equally available to all actors because of varying requirements in organizational capacity, intelligence support, and risk sensitivity. For each of the exaggerated myths mentioned above, there are low-cost, low-

[5] Thomas Rid, "Cyber War Will Not Take Place," *The Journal of Strategic Studies* 35, no. 5 (2012): 5–32; Erik Gartzke, "The Myth of Cyberwar: Bringing War in Cyberspace Back Down to Earth," *International Security* 38, no. 2 (2013): 41–73.

[6] Thomas C. Schelling, *Arms and Influence: With a New Preface and Afterword* (New Haven: Yale University Press, 2008).

payoff irritants widely available as well as higher-cost, potentially higher-payoff adjunct capabilities available to a more restricted set of predominantly nation-state actors. Next we evaluate the coercive utility of these various harms, or threats of these harms, by taking into consideration the interaction of cyberspace with other domains. Finally we ask, what types of cyber coercion are most likely? We argue that there exist two important bounds on the distribution of cyber harm. First, because voluntary connections to the internet make cyber harms possible in the first place, aggressors must be careful not to provoke their victims to disconnect. Second, the availability of military instruments beyond the cyber domain, creates potential for retaliation for unacceptable harms. These constraints combine to make small-scale cyber aggression relatively appealing and thus more likely while making large-scale aggression difficult and undesirable for initiators and thus less likely.

The finding of this chapter extends the logic of the stability-instability paradox pioneered in the 1960s. While nuclear weapons can deter nuclear war, they can fail to deter, and even encourage, conventional or peripheral war. Mutually assured destruction restrained the superpowers from engaging in direct confrontations during the Cold War, even as this restraint encouraged and facilitated the exercise of proxy wars throughout the Third World. The mechanisms of restraint in the cyber domain are slightly different than in the nuclear world—the risk of voluntary disconnection and military retaliation vs. mutual Armageddon—but the results are similar: little truly dangerous behavior and a lot of provocative friction. The social and economic value of the internet expands the scope for minor aggression like espionage, covert influence, and symbolic protest. Cyber operations also act as valuable adjuncts for battlefield operations akin to signals intelligence and electronic warfare for states who are willing and able to go to war for other reasons. However, there are diminishing incentives to “go big” with cyber

warfare alone given the incentives targets have to identify even a hard-to-identify attacker and shift domains to punish cyber aggression. Although the attribution of the attacker's identity is widely thought to be hard problem in cyberspace, anonymity is never guaranteed and might not even be useful for some forms of coercion. A nonzero risk of attribution opens the door to retaliatory punishment, which encourages attackers to exercise restraint in cyber aggression. Ironically enough, the instability we perceive in cyberspace is indicative of the stability of deterrence of the most dangerous cyber threats.

2 The Power to Hurt Online

The US Defense Department describes "cyber" as a separate "war fighting domain" analogous to the traditional air, sea, land, and space domains.[7] However, nothing useful happens in the traditional domains without the ability to monitor and control behavior in them, and information technology has long been critical for these purposes. The internet, and digital devices more generally, facilitate communication, computation, and control within and across organizational and national boundaries. Accordingly, the tactical goals of cyber operations are to degrade or take control of an adversary's computing systems while maintaining control of one's own networks.

Cyber operations (also called "computer network operations" in U.S. military doctrine) are usually divided into three functions: attack, exploitation, and defense. Cyber attack and exploitation leverage similar techniques to gain access to a target system and take advantage of vulnerabilities in its design in order to send instructions to a machine. Both types of intrusion can use technical hacking and malware (viruses, worms, Trojan horses, rootkits, *etc*.) as well as

[7] William J. Lynn III, "Defending a New Domain," *Foreign Affairs*, September 1, 2010.

"social engineering" to trick users into revealing sensitive data and passwords, and most types of intrusions rely on command and control servers to coordinate the attack and pass data back to the attacker. The most important technical difference between "cyber exploitation" and "cyber attack" lies in the behavior of "payload" code in the malware package. Exploitation payloads seek to preserve the illusion of normal functioning in the target system while illicitly stealing data and using system resources. Attack payloads, by contrast, can cause servers to shut down, alter important data, or create malfunctions in computer-controlled industrial equipment. Disruptive attack invariably requires some supporting exploitation for preparatory reconnaissance and performance feedback. Further adding to the confusion, most intrusions described as "attacks" in media reports are actually exploitations—hacking to spy rather than destroy. It is commonplace to hear any sort of cyber intruder described as an "attacker" as in the oft-repeated claim that the U.S. Department of Defense is attacked ten million times per day.[8] In reality such "attacks" are generally routine automated port scans from cyber criminals trolling for low-hanging fruit. This ambiguity contributes to something of a siege mentality in popular accounts of cybersecurity.[9]

Cyber attack and exploitation ultimately use logical code rather than kinetic force to get into the target system. No amount of knocking on a closed door will cause it to open. As Martin Libicki rightly points out, "There Is No Forced Entry in Cyberspace."[10] Even though an attack payload certainly can cause damage downstream, the term "brute force" is a misleading

---

[8] E.g., Zachary Fryer-Briggs, "U.S. Military Goes on Cyber Offensive," *Defense News*, March 24, 2012.

[9] A good primer on offensive cyber operations is William A. Owens, Kenneth W. Dam, and Herbert S. Lin, eds., *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities* (Washington, D.C: National Academies Press, 2009).

[10] Martin C Libicki, *Conquest in Cyberspace: National Security and Information Warfare* (New York, NY: Cambridge University Press, 2007), 31.

regarding the mechanism.[11] Because victims must be made an accomplice in their own exploitation, all intrusions depend on deception. Tactical interaction in cyberspace (like any intelligence activity) is a battle of wits rather than brawn. Whether it can also be a battle of resolve is a question we shall address in the next section. The ubiquity of deception in cyber operations also raises the possibility of its use by the defense. An attacker who walks through an open door cannot be sure it does not lead to a trap.[12]

Cyber defense includes all measures taken to protect networks from adversarial attack and exploitation. This includes the use of firewall and physical boundaries, monitoring network activity, keeping software patches and antivirus definitions up to date, persuading users to improve their security hygiene, and coordination with law enforcement. Sophisticated intruders prize "zero day" vulnerabilities—engineering flaws that have not yet been patched by vendors– as well as many other tricks for eluding detection during infiltration and exploitation.[13] The witting or unwitting insider threat remains an organization's weakest link: users routinely

[11] Schelling uses "brute force" colloquially as physical pounding to weaken an adversary. This usage should definitely not be confused with the computer science notion of a "brute force" attack, which involves running through a rote list of passwords, keys, or vulnerabilities in a simpleminded hope of finding a key to unlock a door. Schelling's "brute force" breaks down a door while a hacker's "brute force" attack looks for an open door. There is no kinetic force in cybersecurity, only deception, even though malware can, via deception, cause kinetic damage in peripherals controlled by computers.

[12] Erik Gartzke and Jon R. Lindsay, "Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace," *Security Studies*, Forthcoming 2015.

[13] Discovering "zero-day vulnerabilities" or novel bugs is hard technical work, and as a result there is a gray market for this skilled labor supported by vendors looking to identify and secure their products (Google, for example, offers a bug bounty), governments looking to enhance their cyber arsenals or protect their networks, and various malicious actors. The use of multiple precious zero days in a single cyber campaign (like Olympic Games/Stuxnet by the US or the so-called Elderwood group in Beijing) is often indicative of skilled and resourced state actor. Some zero-day vulnerabilities have been maliciously exploited long before discovery by vendors, more are exploited in the window between discovery and patching, and even more are exploited well after patching because users fail to patch for a number of reasons (e.g., upgrading to a new version of software may disrupt interfaces with other systems). On the distribution of zero-day vulnerabilities see Bilge Leyla and Dumitras Tudor, "Before We Knew It: An Empirical Study of Zero-Day Attacks in the Real World," *Proceedings of the ACM Conference on Computer and Communications Security*, 2012, 833–44.

disregard prophylactic advice or get fooled by creative phishing scams.[14] Cyber defense is often thought to be relatively more difficult than cyber attack or exploitation (i.e., cyberspace is supposedly "offense dominant"), but it is not clear that this is categorically true.[15] Sophisticated offensives require expertise and preparation and are more likely to trigger robust attribution efforts. Moreover, defensive deception can undermine the attack through active counterintelligence measures like bait or "honeypots" which draw intruders in for observation and quarantine or defensive counterattack (also called "active defense" or "hack back") to retaliate against intrusion infrastructure. Forensic investigations of attacks often turns up clues in technical code artifacts and other intelligence sources which help to attribute the identity of the intruder(s). Attackers must be careful against resourced and resolved defenders.

These three tactical functions can be exaggerated into the three cyber myths mentioned in the introduction. Catastrophic attacks, omniscient exploitation, and impenetrable defense (i.e., *paralysis*, *erosion*, and *lockout*) are myths because they imagine large rewards for little cost. However, in cybersecurity as elsewhere in life, there is no free lunch. The rewards of any given cyber operation are rarely so great and the costs are rarely so trivial. The potential benefits of attack are discounted by uncertainty about the true value of the target to the adversary and the ability for the attacker to take advantage of it. Operative costs include the organizational

[14] "Social engineering" approaches rely on deception of unwitting users to subvert technical defenses. Techniques include "spear phishing" emails disguised to look very convincing to targeted audiences, "watering hole" websites disguised to simulate legitimate sites, or providing infected physical media such as USB drives through "candy drops." Compromise of the account of a trusted insider can enable an attacker to move laterally through a network and escalate privileges, for instance leveraging a secretary to gain access to a chief executive's accounts. RSA, *Social Engineering and Cyber Attacks: The Psychology of Deception*, White Paper, (July 2011), http://www.rsa.com/products/consumer/whitepapers/11456_ SOCENG_WP_0711.pdf.

[15] The logic of offense dominance in cyberspace is critically evaluated by Jon R. Lindsay, "Stuxnet and the Limits of Cyber Warfare," *Security Studies* 22, no. 3 (2013): 365–404; Gartzke and Lindsay, "Weaving Tangled Webs."

resources and human capital required to plan and run an effective cyber campaign.[16] Setting aside the myths of low costs and high rewards, variation in operative costs and benefits allows us to describe more realistic cyber operations. We can identify a set of higher cost, potentially higher reward *adjuncts* that enhance the extant capabilities of stronger actors. There is also a much larger set of low cost, low reward *irritants* that weaker actors or even solitary individuals might be able to access and afford. Each of these categories can be further parsed into modes of attack, exploitation, and defense. Table 1 summarizes our typology of cyber harms, from the mythic free-lunch varietals to more realistic adjuncts and irritants.

Table 1: Types of Cyber Operation

Cyber Operation	**Irritants**	**Adjuncts**	**Myths**
Attack	Hacktivism	Disruption	Paralysis
Exploit	Cybercrime	Espionage	Erosion
Defend	Mobilization	Control	Lockout

2.1 Adjuncts: High costs, (Potentially) High payoffs

Cyber adjuncts are force-multipliers. They amplify the power of actors who have enough resources and expertise to figure out how to manage the complexity and uncertainty associated with ambitious intrusions. As the name suggests, they are not particularly useful as stand-alone operations; that is, they are complements to power not substitutes for it. Adjuncts can usefully augment the effectiveness of military or intelligence activity in other domains. Cyber *disruption* includes cyber attacks against industrial control systems (ICS, to include Supervisory Control and Data Acquisition [SCADA] subsystems) or military command and control and intelligence

[16] Michael C. Horowitz, *The Diffusion of Military Power: Causes and Consequences for International Politics* (Princeton, N.J: Princeton University Press, 2010)demonstrates the importance in general of financial costs and organizational capital in the adoption of military innovation. These two factors are regularly underestimated in technologically determinist narratives of military performance. For cyber in particular, while software appears inexpensive or even free, the supporting costs of management, development, debugging, intelligence, and control can be significant.

systems (C4ISR). The most famous historical example is the Stuxnet attack on Iran's nuclear program, used by the U.S. and Israel in conjunction with diplomatic, sanctions, and intelligence pressure as well as the tacit threat of airstrikes. The sophisticated worm required considerable intelligence, preparation, and technical expertise yet it caused only a temporary loss of Iranian enrichment efficiency—it emphatically was not designed to paralyze the Iranian program.[17] Cyber attack can potentially substitute for electronic warfare in the suppression or destruction of enemy air defenses (SEAD/DEAD), allegedly used, for instance, by Israel to facilitate a 2007 raid on a Syrian nuclear complex.[18] Similarly, the Russians coordinated cyber attacks against Georgian government websites and communications in conjunction with its land invasion of South Ossetia.[19] However, planning complications, intelligence gaps, and uncertainties about unintended consequences (such as encouraging the propagation of cyber weapons by establishing a precedent for their use) have lead U.S. cyber planners to exercise restraint in considering cyber attacks against Libya and Syria.[20] Disruption can play a useful role in a coordinated military operation, but it takes significant organizational skill and effort to integrate cyber operations.

Less cost-intensive but hardly inexpensive, cyber *espionage* is the use of computer network exploitation to access the secrets of an economic competitor or political adversary. Espionage creates a potential "brute force" advantage by altering the balance of power over time, but realizing this advantage requires an actor to leverage complementary strengths.[21] So-called

---

[17] Lindsay, "Stuxnet and the Limits of Cyber Warfare."
[18] David A. Fulghum, "Why Syria's Air Defenses Failed to Detect Israelis," *Aviation Week, Ares Blog*, October 3, 2007. The cyber explanation has been disputed in this particular case, but the general concept is certainly feasible.
[19] Deibert R.J, Rohozinski R, and Crete-Nishihata M, "Cyclones in Cyberspace: Information Shaping and Denial in the 2008 Russia-Georgia War," *Security Dialogue* 43, no. 1 (2012): 3–24.
[20] Ellen Nakashima, "U.S. Accelerating Cyberweapon Research," *Washington Post*, March 18, 2012; David E. Sanger, "Syria War Stirs New U.S. Debate on Cyberattacks," *The New York Times*, February 24, 2014.
[21] We are using the term "brute force" liberally here to describe actions taken to alter the balance of power rather than communicate resolve—conquest vs. coercion. A potential change in the balance of power due to espionage involves no kinetic "brute force" attacks.

"advanced persistent threat" (APT) intrusions target specific organizations and data; therefore,they require a much higher level of preparation and support than generic untargeted cybercrime (where any credit card will do). Chinese APTs have received great notoriety, with ongoing reports of penetrations of Western firms, governments, and non-governmental organizations. However, the discovery of malware like Flame and Duqu, as well as leaks from the U.S. intelligence contractor Edward Snowden, provide ample evidence of Western cyber espionage too. China and the U.S. have both invested heavily in intelligence infrastructure and have long experience in spycraft. Cyberspace creates new opportunities for these states and enables the exfiltration of more data relative to traditional human intelligence (HUMINT), but the organizational infrastructure for intelligence collection and processing remains vital in the digital age.[22]The theft of secret data is only the first step in converting espionage into competitive advantage or realizing, as former National Security Agency (NSA) Director General Keith Alexander said of Chinese activity, "the greatest transfer of wealth in history." The spy must extract valuable "needles" from a petabyte-scale "haystack" of junk data and then successfully disseminate the take to a customer who can make sense of it. The customer must be able to absorb the stolen data into its production or decision processes and to use the result to enhance its advantage in market or political competition. While China has invested large sums in improving its capacity to absorb foreign technology , it is still playing catchup to Western innovation.[23]

[22] Nigel Inkster, "Chinese Intelligence in the Cyber Age," *Survival* 55 (2013): 45–66, doi:10.1080/00396338.2013.767405.

[23] Jon R. Lindsay and Tai Ming Cheung, "From Exploitation to Innovation: Acquisition, Absorption, and Application," in *China and Cybersecurity: Espionage, Strategy, and Politics in the Digital Domain*, ed. Jon R. Lindsay, Tai Ming Cheung, and Derek S Reveron (New York: Oxford University Press, Forthcoming).

If a corporation, military, or other political entity wants to capture the positive network effects and efficiencies of internet communication, then the technology must be defended against intrusion and misuse. Cyber *control* assures the positive use of computing networks for their intended purposes, but it is neither cheap nor absolute. Indeed, to the degree that disruption and espionage are possible (or hacktivism, fraud, or mobilization, for that matter), perfect control or lockout is not. This claim is generally uncontroversial, as cyber defense is widely held to be a very difficult coordination problem in an offense-dominant medium. Some aspects of cyber defense have private goods characteristics such as firewall and intrusion detection systems protecting the owner's network perimeter, but cybersecurity is also beset by public goods problems.[24] Examples include users who opt not to patch their systems and end up hosting botnets that attack other users and software venders who neglect the development security features in the rush to get their products to market. Meanwhile, offensive cyber threats can change their signatures faster than defenders can keep up.[25] Even authoritarian governments cannot achieve absolute advantage in the arms race with technologically-savvy dissidents, at least as long as those states also desire digital access to international economic transactions. Yet weak dissidents face even greater challenges assuring control of their social networks, since regime actors can use not only espionage but also more draconian enforcement measures against them.[26] The relationship between internet control, innovation, and freedom is too complicated to

[24] Paul Rosenzweig, *Cybersecurity and Public Goods: The Public/Private "Partnership,"* Koret-Taube Task Force on National Security and Law, Emerging Threats Essay (Stanford University Hoover Institution, December 2011).
[25] Dale Peterson, "Offensive Cyber Weapons: Construction, Development, and Employment," *Journal of Strategic Studies* 36, no. 1 (2013): 120–24.
[26] Sarah McKune, "'Foreign Hostile Forces': The Human Rights Dimension of China's Cyber Campaigns," in *China and Cybersecurity: Espionage, Strategy, and Politics in the Digital Domain*, ed. Jon R. Lindsay, Tai Ming Cheung, and Derek S Reveron (New York: Oxford University Press, Forthcoming).

analyze further in this chapter, but that very complexity makes either version of lockout an unlikely possibility.

2.2 Irritants: Low costs, Low payoffs

Cyber adjuncts can make the strong even stronger, if coordinated with other sources of strength. Contrary to a common belief about the asymmetric nature of cyber warfare, adjuncts tend to advantage nation state actors, especially great powers with institutional resources to overcome the planning and intelligence obstacles and to price in the risk of failure. Cyber irritants, by contrast, are widely affordable for all types of actors, weak or strong, and they can be employed with lower risk of adverse consequences. By the same token, the expected rewards or irritants in competition are low. The risks are low mainly because those who have the power to intervene to stop or punish irritant behavior often do not have the motivation to do so. While irritants are often illegal, law enforcement authorities often do not launch sufficiently aggressive investigations, for want of resources or authorization, to discover and sanction the perpetrators. Irritant attackers can thus hide safely behind their digital anonymity, whereas adjunct attackers would provoke a more concerted investigation and response.

The vast majority of empirically-observable cyber attacks can be grouped into the category of *hacktivism*, which includes distributed denial of service (DDoS) attacks that knock servers offline, website defacements, defamation, and other forms of online protest. The most famous example in this class is the barrage of DDoS attacks and defacements from Russian nationalists which wracked Estonia in 2007 following the removal of a Soviet memorial in Tallinn. Similar activity from Chinese nationalists is also common in China's periodic tensions with Taiwan and Japan. The anarchist group "Anonymous" has embarrassed several firms and even government agencies in the U.S. by illicitly acquiring and then publicly posting confidential

data. Hacktivist attacks can grab headlines and be embarrassing to those they target, but they usually subside within a few news cycles. DDoS and defacements have become prominent during almost any period of political tension as a form of nationalist protest, and there is often some ambiguity behind whether government actors or nationalist citizens are responsible.[27] At the technical level, most hacktivism is technical and reversible, and mitigation techniques for things like DDoS are readily available. Colloquially "hacktivism" is sometimes used to cover what we call *mobilization* in our typology (i.e., when hacktivists seek to encourage one another and create support for a cause). Also, the same technical methods used for criminal exploitation or what we categorize as *fraud* can also be considered hacktivism if conducted for the purposes of defamation or political influence rather than financial gain (e.g., when hackers steal and expose confidential information after the manner of Wikileaks).

Cyber *mobilization* is the use of social media to coordinate protest activity online and on the ground. We categorize this as a mode of defense because dissidents seek to use social media as designed to realize the positive network benefits of information technology rather than to impose a cost on other's use of information technology in the manner of attack and exploitation. Many recent large scale political protest movements associated with successful revolutions like the Arab Spring or Ukrainian Maidan leveraged social media to mobilize and organize dissidents.[28] However, their success owed more to the tenacity of the protestors, their physical presence in number, and the restraint of government security forces. The limits of mobilization were highlighted in the abortive Iranian "Green Revolution" of 2009 and the Chinese "Jasmine Revolution" of 2010, where government security forces exploited the use of social media to

[27] A wave of DDoS attacks against Georgia in 2008 were initially thought to be directed by the Russian government but may have been more spontaneous: Deibert R.J, Rohozinski R, and Crete-Nishihata M, "Cyclones in Cyberspace."
[28] Diamond, "Liberation Technology."

identify and punish protesters.[29] The defenses of mobilization are weak because software products must be distributed wholesale to interconnected users with uneven technical skills. State-sponsored attackers can exploit this vulnerability, as the Chinese state sought to provide spyware to protesters in Hong Kong in 2014.[30] Mobilization can also take on virulent or socially dysfunctional forms such as cyber bullying and its industrial scale Chinese analogue known as "human flesh search," a virtual lynch mob defaming corrupt officials, disgraced celebrities, and unfortunate citizens alike.[31]

Cybercrime or computer *fraud* is the primary empirical manifestation of cyber exploitation. The vast majority of threats reported by cybersecurity firms fall into this category. Unlike most of the other harms detailed above, which all have a political element, fraud is motivated primarily by financial gain. There is a complex, segmented, global market for cybercrime divided into interdependent advertising, theft and fraud, and technical support rackets. This infrastructure also supports more sophisticated adjuncts by providing malware and compromised hosts to facilitate intrusion. However, as discussed above, APTs require much more skill and effort per target and assume greater risk. Unlike APTs, retail cybercrime is untargeted and scales more easily to exploit millions of potential victims, since it only has to be successful a fraction of a percent of the time to be profitable. Underground revenues total hundreds of billions of dollars annually, although the vast majority of cyber criminals actually make very little money because of rampant dishonesty in the underground economy and nontrivial law enforcement risks. Even spectacular compromises of millions of credit card

---

[29] Evgeny Morozov, *The Net Delusion: The Dark Side of Internet Freedom*, Reprint edition (New York: PublicAffairs, 2011).
[30] Shannon Tiezzi, "China's Cyber War Against Hong Kong Protestors," *The Diplomat*, October 1, 2014, http://thediplomat.com/2014/10/chinas-cyber-war-against-hong-kong-protestors/.
[31] Saul Levmore and Martha C. Nussbaum, eds., *The Offensive Internet: Speech, Privacy, and Reputation* (Cambridge, MA: Harvard University Press, 2011).

accounts do not readily translate into handy profits because translating that data into a useable monetary instrument is difficult.[32]

To sum up, the tactics in a contest for control in cyberspace are cyber attack, exploitation, and defense. These can be used as costly adjuncts to potentially enhance other advantages, or as inexpensive irritants for minor gain. Myths of grave harms are based on an unrealistic assessment of the operational costs involved. The myths also underestimate the coercive context of supposed harms: even if the operational barriers are overcome (i.e., if adjunct capabilities are perfected), how will the adversary react? The strategic logic of cybersecurity must take into account not only the technical possibility of harm, but also the coercive potential of threats toward and received by the target of harm. For the most part the discussion above has focused on the ways in which cyber means can provide some direct benefit through via support to military operations, intelligence advantage, information system control, symbolic protest, political mobilization, financial gain, etc. Some mention of deterrent threats (or their absence) has been unavoidable even in the discussion of "brute force" cyber harms, such as the unwillingness of law enforcement to pursue all irritants, or the willingness of strong actors to use cyber adjuncts when backed up by other capabilities. We now turn from harmful means to coercive ends, and from the power to hurt to the power to persuade.

[32] Kirill Levchenko et al., "Click Trajectories: End-to-End Analysis of the Spam Value Chain," in *Proceedings of the 2011 IEEE Symposium on Security and Privacy*, SP '11 (Washington, DC, USA: IEEE Computer Society, 2011), 431–46; Ross Anderson et al., "Measuring the Cost of Cybercrime," in *The Economics of Information Security and Privacy*, ed. Rainer Böhme (Berlin: Springer-Verlag, 2013), 265–300; Cormac Herley, "When Does Targeting Make Sense for an Attacker?," *IEEE Security & Privacy* 11, no. 2 (2013): 89–92; Jianwe Zhuge et al., "Investigating the Chinese Online Underground Economy," in *China and Cybersecurity: Espionage, Strategy, and Politics in the Digital Domain*, ed. Jon R. Lindsay, Tai Ming Cheung, and Derek S Reveron (New York: Oxford University Press, Forthcoming).

3 Cyber Coercion

Coercion is different from pure harm, even though coercing may require causing some harm in the process of creating credible threats of more harm. Harm pure and simple, or brute force, aims to change the balance of capabilities between adversaries in the present, while coercion uses harm or threats of harm to influence an adversary's decision-making in the future. Coercion is a signaling process which attempts to link particular behaviors to unpleasant consequences in the mind of the opponent. It targets the willingness of the opponent to endure suffering or comply with demands and can include *deterrence*, to prevent something from happening, or *compellence*, to cause something to happen, as well as more complicated forms of signaling which we will not address in this chapter.

Yet the future-directedness of all forms of coercion seems to create a problem in cyberspace. How can an adversary be made to understand a credible threat of future harm via network connection and yet voluntarily maintain the connections on which that future harm depends? As mentioned above, cyber intrusions depend on deception because logical, massless code cannot kinetically force its way through anything. If software vulnerabilities are highlighted by an explicit threat to exploit them, then the target can patch or otherwise neutralize the threat. Therefore, if attackers rely on the difficulty of attribution to protect themselves, then they cannot make coercive demands which would reveal their identity. Moreover, if bad intelligence or buggy malware in the threatened cyber attack causes unexpectedly high or low damage when exercised, then the punishment may have little resemblance to the threat. Vague threats from anonymous sources lack credibility, fail to precisely specify the action proscribed or demanded, and offer little reassurance that the coercer will withhold punishment if the target complies.

Furthermore, a necessary reliance on deception and ambiguity creates considerable potential for the misperception of coercive signals.

Nevertheless, cyber operations are not completely devoid of coercive potential, even if they are more limited compared to more traditional means of aggression. Cyber can be employed in some circumstances for deterrence and compellence, especially if expectations for success are limited. The most promising coercive cyber tools are adjuncts (rather than irritants) because they have the potential to impose higher costs in conjunction with other (non-cyber) tools. Table 2 summarizes strategies for the coercive threat of "pure" cyber and "cross domain" harms, which are threats of means other than cyber, especially military force, to influence cyber behavior.[33] We will discuss potential applications and complications associated with each of these strategies.

Table 2: Cyber Coercion

	Deterrence	**Compellence**
Cyberspace	Detection Denial Deception	Latency Extortion Seduction
Cross Domain	Retaliation Disconnection	Escalation Protection

3.1 Cyber Deterrence

The great ambiguity of attribution in cyberspace is generally thought to undermine the credibility of retaliatory threats and thus cyber deterrence in general.[34] However, these assumptions are not necessarily true once the entire range of cyber operations is taken into account. Pervasive surveillance or the threat of detection augments general deterrence, public investment in cyber

[33] Cross-domain deterrence (or coercion more generally) is the use of one set of means to influence an unlike set of means and behaviors. See Erik Gartzke and Jon R. Lindsay, "Cross-Domain Deterrence: Strategy in an Era of Complexity" (presented at the International Studies Association Annual Meeting, Toronto, 2014).

[34] For extensive discussion of the problems of deterring cyber attacks see Martin C. Libicki, *Cyberdeterrence and Cyberwar* (Santa Monica, CA: RAND, 2009); National Research Council, ed., *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy* (Washington, D.C.: National Academies Press, 2010).

defense improves deterrence by denial, and defensive deception improves both punishment and denial strategies.

The chapter by Austin Long in this volume discusses ways in which intelligence capacity can augment and enhance coercion. Insofar as cyber operations are particularly well suited for intelligence, then they should also be expected to enhance the intelligence dimensions of coercion. A reputation for skill at surveillance (by any means) can dissuade targets from planning and executing harmful operations. Intelligence collection can also provide the direct benefit of targeting data and it can help, potentially, to shift the balance of power over time by transferring knowledge and evening the playing field (with all the caveats regarding absorption mentioned above). Moreover, the fact or fear of surveillance can also encourage paranoia and force a target to adopt onerous security measures. Extensive signals intelligence (SIGINT) monitoring of underground groups forces many of them to rely on couriers who are slower, more expensive, and less efficient than mobile phones. Battlefield targets that adopt debilitating "emissions control" postures effectively commit "EMCON suicide." Paranoia that detection will be followed by swift precision strikes or law enforcement action may cause targets to forgo misbehavior altogether. Michel Foucault introduced the metaphor of the "panopticon" to describe how pervasive state surveillance deters social deviance.[35] Likewise, citizens subject to continuous monitoring via public cameras and internet surveillance tend to internalize obedience to the state or at least curtail observably deviant behavior.

[35] The panopticon was a design for a circular prison in which guards posted at the hub could observe prisoners in cells located all along the circumference without themselves being observed. The prisoners could never be certain whether they were being watched or whether the guards were on break, so to be safe they would tend to act as if they were always being watched. See Michel Foucault, *Discipline & Punish: The Birth of the Prison*, trans. Alan Sheridan (New York: Random House, 1977).

A reputation for extensive cyber exploitation similarly exercises a deterrent effect on would be conspirators, enemy hackers, and terrorists. While Edward Snowden's leakage of top secret NSA documents has certainly compromised technical intelligence sources, it has also helped the U.S. to advertise the extent and technical skill of NSA penetration of the internet.[36] This deterrent signal is credible because it is also costly in terms of lost sources and, potentially, lost market share for US firms.[37] Deterrence by detection works best when the target knows or fears that the probability of detection is high. Attackers must be concerned that defenders will invest considerable effort into attributing identity for damaging attacks, even as the complexity of consequential attacks and the context of crisis make it more likely that the attacker will leave clues that aid detection.[38] Deterrence by detection is enhanced by cultivating a reputation for skilled cyber exploitation, which in turn improves the credibility of threats of punishment by whatever means, cyber or cross-domain.

A reputation for skilled cyber defense, by contrast, enhances deterrence by denial.[39] Cyber defense can block, parry, or redirect intrusions. Public knowledge of the robustness of

---

[36] The deterrent effect of Snowden's revelations seems to have been acknowledged by Admiral Michael Rogers, Director of the NSA, in his 11 March 2014 confirmation testimony to the Senate Armed Services Committee (http://www.armed-services.senate.gov/imo/media/doc/Rogers_03-11-14.pdf): "I believe the recent disclosures of a large portion of our intelligence and military operational history may provide us with opportunity to engage both the American public and our international partners in discussion of the balance of offense and defense, the nature of cyber warfare, norms of accepted and unacceptable behavior in cyberspace, and so forth."

[37] The benefits of deterrence must be weighed against the costs of compromise. Snowden's revelations incentivizes intelligence targets to search for better tradecraft to restore anonymity and defeat deterrence. In the U.S. case this has included greater reticence of American internet firms to collaborate with the NSA for fear of losing international market share, but the NSA surely has other avenues to exploit to maintain situational awareness.

[38] For similar argument about attribution for nuclear terrorism, a threat supposedly magnified by the anonymity of the perpetrators, see Keir A. Lieber and Daryl G. Press, "Why States Won't Give Nuclear Weapons to Terrorists," *International Security* 38, no. 1 (July 1, 2013): 80–104.

[39] Deterrence theorists distinguish deterrence by punishment from deterrence by denial. The former imposes a costly penalty if the target successfully crosses the line, while the latter raises the costs of attempting to cross it. Police deter burglars through the unsavory thought of prison, while safes and alarms raise the costs of burglary itself to make it prohibitive to all but the most skillful and resolved of thieves.

defense makes attackers worry that their efforts might be futile, even dangerous. Defensive aptitude can be signaled through costly investment in cyber intelligence, law enforcement, and regulatory agencies and the advertisement of success in detecting and thwarting attacks. Creation of U.S. Cyber Command to defend military networks, major exercises like the Cyber Storm series, and heightened concern about cyber in military professional literature and budgetary investment in training and capabilities all provide signals of American commitment to the defense of its warfighting networks. It is not necessary for all intrusions to be prevented, moreover. In the case of Chinese APTs, the public disclosure of Chinese tradecraft and PRC government responsibility, heightens the suspicion future Chinese intruders must face when trying to hide or deny their involvement in espionage. Indeed, Chinese APT activity paused and reconfigured following the February 2013 exposé by cybersecurity firm Mandiant, which used a variety of lapses in Chinese tradecraft (e.g., operator reuse of names and emails, command and control servers located in China, operators checking personal Facebook accounts from their attack infrastructure, etc.) to identify a specific Chinese military unit in Shanghai involved in the exploitation of 141 English-speaking targets worldwide.[40]

The distinction between punishment and denial breaks down for forms of cyber defense that involve counterattack and deception against the intruder. U.S. officials have begun to announce that the cyber attribution problem is not as daunting as once believed and that attackers will be met with a decisive response in cyberspace or elsewhere.[41] A reputation for skill at cyber operations is useful not only for the deterrence of cyber attacks (because of the risk of detection

---

[40] Mandiant, *APT1: Exposing One of China's Cyber Espionage Units*, White Paper, (February 2013), http://intelreport.mandiant.com/Mandiant_APT1_Report.pdf; *M-Trends: Beyond the Breach, 2014 Threat Report* (Mandiant, April 2014).

[41] Elisabeth Bumiller and Thom Shanker, "Panetta Warns of Dire Threat of Cyberattack on U.S.," *The New York Times*, October 11, 2012. See also testimony from Admiral Rogers, cited above n36: "I believe there can be effective levels of deterrence despite the challenges of attribution"

and punishment or of denial) but also for more general coercion, if they augment the effectiveness of other threats or undermine the target's confidence in defending against them. So far the U.S. has demonstrated the greatest capacity for and willingness to use cyber operations through its Olympic Games program which allegedly produced, among other mischief, the Stuxnet attack on Iran. Unfortunately, this same case highlights the deterrent limitations of cyber punishment, as Iran continued to enrich uranium and even accelerated the modernization and relocation of its program after 2010 while also pursuing its own offensive cyber program.[42]

Deception is an underappreciated strategy that is particularly promising for network protection. Ruses, honeypots, digital bait, data obfuscation, and more aggressive counterintelligence techniques have already been employed by security engineers. Deception can confuse, delay, misdirect, or even harm the attacker, for instance by enabling the exfiltration of harmful malware to infect the attacker's home networks. Whereas pure deterrent strategies punish intrusion and pure denial strategies impede it, deception actually encourages intrusion, but then turns it against the intruder. For this reason, deception is rightly considered a distinct protective strategy, even though in practice it operates to reinforce defense or deterrence by punishment or denial. While deception has always been available and has been practiced in the past, the growth in complexity of information technology from the telegraph to the internet has made deception more possible and useful than ever before. The supposed offense dominance of cyberspace is actually a reflection of its more fundamental potential for deception. However, defenders can also use deception to enhance punishment and denial. An attacker in a house of mirrors, or who suspects a defensive gambit, cannot take its advantage for granted. There are

[42] Christopher Bronk and Eneken Tikk-Ringas, "The Cyber Attack on Saudi Aramco," *Survival* 55, no. 2 (April 3, 2013): 81–96; Lindsay, "Stuxnet and the Limits of Cyber Warfare."

clearly some operational and legal challenges associated with cyber deception, and all forms of active defense; defensive deception is a complex, if potentially rewarding, strategy for protection. As with control adjuncts generally, the advantage in deception thus goes to the strong.[43]

There is no reason for deterrent threats to be limited to cyber actions. In fact, the most important bounds on the severity of cyber aggression observed thus far probably have nothing to do with the technology of cyberspace. Victims of aggression can and likely will look to responses not only in kind but also through whatever other means they possess, ranging from conventional military retaliation, irregular warfare and covert subversion, trade and financial sanctions, and so forth. The problem of attribution is often thought to preclude the ability to retaliate, but as mentioned above, this problem is probably overstated for any serious attack. The question for an aggressor contemplating serious cyber harm thus becomes not whether but how the victim will retaliate. The only exception would be for the attacker to launch a cyber attack so devastating that it effectively paralyzes all ability for the target to retaliate, but as we have argued above, such paralysis is a myth for operational reasons alone—a splendid first cyber strike is simply too complex and full of uncertainties to reliably deliver its results—let alone the risks of cross-domain retaliation. If the victim retains significant options in other domains where the attacker's ability to resist is slight, then the attacker has strong incentives to avoid provoking a response in those domains. Even if the target is asymmetrically more dependent on the internet, making disruptive cyber attack seem like an attractive possibility, advanced industrial countries are for the same reason more likely to have other advanced military and economic options

[43] The strategy of deception is described in more depth by Gartzke and Lindsay, "Weaving Tangled Webs."

available where the asymmetries do not favor the attacker. Serious attacks invite serious responses, which need not be in kind.

Perhaps the simplest form of cross domain response to cyber threats is to forgo the use of the cyber domain altogether. While it is hard if not impossible to limit exposure to nuclear weapons and even a determined conventional assault, the risk of cyber attack can be completely eliminated by disconnection from digital networks. The internet is an artificial environment and connection to it is voluntary. Individuals, organizations, and states retain the ability to unplug completely, limit their online transactions, or erect various barriers to connection. Obviously disconnection is not very feasible commercially, socially, and militarily today, but this is more of an indicator of how positive the benefits of interconnection are compared to the perceived risks. If the risks were perceived as extreme, then firms and states could go back to making a living as they did before 1991 (when WWW went public). This is a cross-domain threat because it entails exiting the cyber domain altogether to leverage more traditional economic and military transactions. The threat of disconnection follows from the more general logic of international organizations, where contracts must be self-enforcing.[44] On the internet as in institutions, ties among egoistic actors under anarchy must be mutually beneficial. If the internet is a bad deal for actors, they can throw up boundaries or exit cyberspace altogether. If repeated exposure to adversarial exploitation causes states to lose more than they gain from being online, then they can undermine the attacker's very means for accessing the victim. The threat of voluntary disconnection is especially relevant for repeated interactions, or repeated exploitation, rather than a one-shot "bolt from the blue" cyber attack (which is better countered with cross-domain retaliation). The threat of disconnection is implicit in the voluntary nature of connection to the

[44] Inter alia, Krasner 1983, Snidal 1985, Oye 1986

internet, and the potential loss of the ability to make future attacks exercises a deterrent effect on attacks in the present. An aggressor who does not want to lose the cyber adjuncts for espionage and disruption it has invested so much in developing will show restraint in their employment. This does not mean that coercion cannot take place online, but it is bounded by excess value. One implication is that the countries that can be most coerced on the internet will be those that have the most to lose by leaving it.

To sum up this section, we have reviewed five strategies for deterring cyber attacks, three relying on actions within the cyber domain and two relying on "cross domain" actions beyond it. Detection—leveraging cyberspace as a panopticon—removes the cloak of anonymity cyber attackers depend on and facilitates retaliation. Denial—a reputation for effective defense—counters attackers who believe they can, nonetheless, maintain their tactical covertness. Deception—using attackers' strengths in stealth against them—reinforces both punishments and denial. Importantly, all three of these strategies are useful for the entire range of cyber attack exploitation, adjuncts as well as irritants, although for many irritants the effort at deterrence may not be worthwhile. By contrast, the other two cross domain strategies—retaliating by any means necessary or disconnecting from the threat altogether—are aimed more at deterring high impact adjunct aggression or keeping aggression confined more toward the irritant end of the spectrum. As we will discuss in the conclusion, this gives rise to a cyber analogue of the stability-instability paradox.

3.2 Cyber Compellence

Schelling and others have argued that compellence is harder than deterrence. Deterrence dissuades while compellence persuades. Deterrence stops something in motion while compellence starts something at rest. Deterrence need only signal, "Don't cross this line," while

compellence must signal, “Move across this line (not that one), and only so far.” If the U.S. threatens or conducts a cruise missile attack against a state believed to be harboring a suspected leader of Al Qaeda (i.e., using the base state coercion strategy described by Keren Fraiman in this volume), the goal is to get the target to extradite the Al Qaeda leader. Would a cyber attack or threat of attack be able do the same thing, in this or any other coercive scenario? When would coercion work because of cyber means but not because of something else? Is there any target that cyber tools have particular power to compel? If cyber deterrence is thought to be difficult, surely cyber compellence is more difficult still. While limited, however, it is not impossible.

The biggest obstacle to cyber coercion is the difficulty of credibly signaling about potential harm that depends on secrecy to be harmful. Advertised cyber threats that are specific enough to be coercively useful (i.e., for immediate rather than general coercion) can be readily neutralized through patching or reconfiguration. Sacrifice of the anonymity on which offensive deception depends exposes the cyber attacker to retaliation. Coercive cyber threats thus tend to be more generalized, which undercuts their effectiveness in targeted or crisis situations. We discussed deterrence by detection above, the threat of using internet exploitation to attribute identity and improve the credibility of punishment threats. The effectiveness of the cyber panopticon depends more on generalized concerns that an intelligence service might be reading email or snooping around networks than any immediate threat of detection. This form of deterrence is especially effective when employed by a strong state with a sophisticated digital surveillance system. The compellent analogue is the generalized, and generally latent, threat of escalation from cyber exploitation to disruptive attack.

By heightening the intensity of intrusion for intelligence purposes, an actor can also signal that it might be possible to convert an intrusion into something more dangerous. Detection

of intelligence surveillance might reveal preparation for (or at least enhance paranoia about) something more disruptive.[45] The potential for escalation from exploitation to attack is latent most of the time in reality, and the distribution of actual harms inflicted through the internet reflects a lot of a little harm and very little of a lot. Yet the inherent potential to use exploitation for disruption encourages a visceral fear of helplessness in cyber discourse: what if we lost the connection on which our prosperity and strength depends? What if defense against societal paralysis is as fruitless as defense against cybercrime? What if all it takes is a change of mind by the adversary to convert a lot of a little into a lot of a lot? Technological, intelligence, and operational constraints in fact render the fungibility of technique from exploitation to attack less of a realistic concern than oft feared (because engineering destruction on command requires significant additional expertise and testing), or restrict it only to large actors with resources and experience to make disruptive adjuncts work. Nonetheless, the prospect of ongoing technological innovation and falling barriers of entry to cyber attack tempt many observers to take this dormant potential seriously. It is important to emphasize that escalation latency is not a specific threat of harm, but more of a generalized paranoia that "our networks are already so penetrated that resistance is useless!" Since this fear rests on a misperception—escalation is operationally nontrivial, victims of network disruption often find alternatives, and both sides will learn both of these points as the crisis drags on—cyber latency is not very reliable for compellence on its own. But anything that creates fear can be at least a little bit useful for a broader gambit—again cyber appears useful as a complement to not a substitute for other means of coercion.

[45] This is a common ambiguity in intelligence operations, not unique to cybersecurity. The same organizations often conduct collection and covert action, because similar tradecraft is useful for infiltrating agents whether they plan to steal or to kill, and for protecting command and control with them as they do so. This ambiguity is particularly pervasive in cybersecurity, which is one reason why the same officer commands both the NSA and U.S. Cyber Command.

Specific and credible coercive signaling is more possible in the realm of irritant cybercrime that takes place beneath the threshold of state-sanctioned retaliation. There exist extortion scams that use malware ("ransomware") to disable a computer or lock out access to data unless the victim pays a ransom into an account which the anonymous attacker can access.[46] This strategy uses the threat of disconnection in a compellent rather than a deterrent role. Cyber blackmail is only useful at small scales where the ransom is less than the reconstitution cost of the embargoed system or the cost of contacting law enforcement. It thus can be most effective against victims who themselves might be on the wrong side of the law, say a threat to embargo a bookie's books or computers that support an illegal racetrack.

Authoritarian states can also use threats of disconnection from the internet to cow media outlets and muzzle dissidents in order to shape the information environment. Historical attempts to cut off a population from the internet altogether, for example in Egypt during the Arab Spring or in China during the Xinjiang unrest of 2009, fall more into the category of control rather than coercion, although expectations that a "kill switch" might be thrown could have a slight persuasive effect on dissidents (who would surely worry much more about riot police if it came to that). It is also possible, with a lot of imagination, to conjure up scenarios where a cyber attack causes costly malfunctions but for some reason audience costs prevent the victim from publicly admitting to the damage or the mechanism. For instance, an attacking state could offer to halt the attack or repair the damage if elites in the victim state refrain from opposing the attacker's whims. Joel Brenner describes one such improbable scenario in "June 2017" where China blackmails a US President by knocking out large sections of the US power grid run that only China can repair (through its monopoly on a particular type of generator), but only if the US

[46] Ian Urbina, "Hackers Find New Ways to Breach Computer Security," *The New York Times*, June 21, 2014.

recalls a carrier strike group en route to Taiwan.[47] Importantly, all these examples of cyber extortion can only work where the victims are weak or lack alternatives or recourse for some reason. Brenner's scenario would almost certainly backfire on the Chinese.

We discussed above how deception can enhance deterrence by punishing intruders who steal baited data or attack honeypot systems. Deception can also enhance compellence by making the desired action appear more attractive to the target than it actually is. The strategic essence of deception is persuading an adversary to voluntarily take action that is not, ultimately, in its interest. Compellent deception can be described as seduction. It is the bread and butter of commonplace internet fraud and one of the charms of the Nigerian princess who needs your help to recover lottery winnings held by an alien hovering over London. Cyber "seduction" might also be useful for sending bogus orders to enemy troops in the field instructing them to retreat or stow their weapons or to lure enemy commanders into an ambush. Importantly, as with other cyber adjuncts, the window of vulnerability created through such a ruse is only useful if the seducer plans to exploit it in the terrestrial domain.

More dramatic forms of "cross domain" coercion can involve cyber attack (and exploitation) in conjunction with other forms of military force—e.g., disabling early warning or command and control networks in support of an air campaign—which might make a threatened intervention seem more potent, if an actor is inclined to threaten military action. That is, a hostile adversary armed with cyber weapons in addition to tanks, fighters, and missiles might appear more able to overcome defenses and thus be better able to issue compellent threats. Alternatively, a preparatory cyber campaign to destabilize communications or carry out deception operations could signal a willingness to escalate to more meaningful forms of

[47] Brenner, *America the Vulnerable*, chap. 7.

aggression. Certainly the presence of aggression online helps to cultivate an air of crisis which could be useful for strategies of risk or brinksmanship. Something like this seems to have happened in Ukraine in 2014 where cyber attacks accompanied the incursion of Russian commandos to take Crimea without resistance, while mechanized Russian forces mobilized and exercised on the border.[48] Another route to coercion involves outbursts of DDoS attacks and website defacement by civilians to cultivate an air of crisis and provide elites with a diplomatic argument that their hands are tied by popular nationalism. The hands-tying argument loses credibility if the state has demonstrated an ability in tamp down online outbursts at will, as China has. Importantly, all of these mechanisms depend on the coercer having the ability to escalate beyond the cyber domain, leveraging the fear of substantive military assault. Cyber aggression is used to signal the risk or potency of cross-domain escalation.

Cyber defense can likewise enhance the credibility of coercive military threats. If a coercer has and is known to have robust cyber defenses, then the target will have less faith in preempting or defending against the threatened punishment. Conversely, a coercer who lacks adequate cyber defenses is like the proverbial stone-thrower in a glass house. Threats will lack credibility if exercising them means assured retaliation, in this case a cyber counterattack on military C4ISR that delays or degrades the ability to carry out the promised punishment. But if cyber defense can be assured, then it is easier—or at least less hard—to project military power in the service of coercive diplomacy. The protection of computer networks is thus an important complement to the issuance of cross-domain threats that depend on them.

[48] David E. Sanger and Steven Erlanger, "Suspicion Falls on Russia as 'Snake' Cyberattacks Target Ukraine's Government," *The New York Times*, March 8, 2014; Kim Zetter, "Russian 'Sandworm' Hack Has Been Spying on Foreign Governments for Years," *Wired Threat Level*, October 14, 2014, http://www.wired.com/2014/10/russian-sandworm-hack-isight/.

In this section we have reviewed five compellent strategies involving the use of cyber instruments analogous to the five deterrence strategies, three exclusively cyber and two cross domain in nature. Latency exploits the ambiguity between cyber exploitation and attack to create fears of harmful punishment if the target doesn't comply with demands. Extortion denies the target's use of cyber resources for blackmail purposes. Seduction uses deceptive techniques to lure the target into a position where compellent punishment and denial will be more effective. Escalation uses cyber aggression to signal a risk of more punishing consequences in other domains, while the protection of cyber assets improves the credibility of military threats in other domains (i.e., through assuring that coercive command and control remains connected). This whole discussion shows how closely related deterrence and compellence are. Deterrence can facilitate aggression by blocking counterattack. This may be especially the case for cyber operations, particularly for strong states like the U.S. that can use the deterrent threat of military retaliation to cover the employment of offensive cyber tools. As in all forms of coercion, cyber compellence appears to be more complicated than deterrence, even as they share many constraints and signals. For both deterrence and compellence, cyber coercion is most effective when employed as a complementary adjunct with capabilities to hurt in other domains. The capacity of cyber means to operate as a substitute is highly constrained and effective mostly for aggression of the irritant class.

3.3 Misperception

No discussion of coercion would be complete without some attention to the psychological dimension. After all, coercion seeks to influence the mind of the opponent by generating credible signals of future costs and benefits associated with taking (avoiding) certain behavior. These signals can be missed, garbled, or misinterpreted. If, as we argue, cyberspace is a domain that is

especially conducive to deception, then misperception is liable to be a major problem.[49] Cyber operations are characterized by secrecy and ambiguity. They often avoid the clarity associated with, say, a military invasion of territory. Strategic interaction in cyberspace should be considered in terms of intelligence, counterintelligence, and subversion—the language of force-on-force conflict or nuclear diplomacy is misleading and inappropriate. This secrecy and duplicity inherent in cyber operations makes it easy to misinterpret signals, if indeed it is possible to signal at all. The literature on the coercive utility of intelligence and subversion remains an undeveloped area in strategic studies in general, as Austin Long points out in this volume.

The most pessimistic result of misinterpretation would be inadvertent escalation. This would be a crisis in which the use of cyber means and responses to them lead actors to inflict and accept greater costs than they would have been willing to pay at the start of the crisis. For example, the use of cyber attacks to degrade an opponent's command and control systems, either as part of a coercive risk strategy or to facilitate limited conventional strikes, could put the opponent into a "use it or lose it" situation. Fear of the even partial loss of command and control could lead to premature retaliation. This scenario would be most dangerous in a case where a target fears that its nuclear deterrent is imperiled by strikes on command and control intended to facilitate limited conventional operations. If cyber makes a coercer more confident in conventional offensives by both protecting friendly C4ISR while degrading the adversary's, then there may be a lower threshold for such operations. Unfortunately, a lower threshold for

[49] Rose McDermott, "Decision Making Under Uncertainty," in *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy*, ed. National Research Council (Washington, D.C.: National Academies Press, 2010), 227–42.

conventional attacks on C4ISR also makes nuclear escalation more risky.[50] The dangers of inadvertent escalation through the cyber domain are thought to be particularly salient in a US-China conflict scenario due to dangerous combination of US preferences for aggressive C4ISR counterforce operations, Chinese convictions that information warfare must be used preemptively and decisively, and nationalist pressures in both states to retaliate for costly losses.[51]

Nevertheless, the advent of cyberspace as a domain of contestation is also cause for optimism. There is an emerging consensus in political science that uncertainty is a major—if not the major—cause of war. If the costs and outcomes of wars were knowable in advance, than belligerents would be better off negotiating a settlement.[52] Uncertainty in the current and future balance of power or resolve, however, can make fighting more attractive. If better intelligence can improve knowledge of the actual balance of power and interests, then war should be less attractive. Cyber operations are most useful in an intelligence role, and they can potentially convey information about interests (including the willingness to escalate) without actual fighting. Deliberate collection and pervasive leaks all enhance transparency, which should make conflict less likely. Furthermore, cyberspace is a manmade construct of commonly embraced and mutually beneficial protocols—interoperability is the condition for the possibility of cyber

[50] Barry R. Posen, *Inadvertent Escalation: Conventional War and Nuclear Risks*, 1 edition (Ithaca, N.Y.: Cornell University Press, 1991).
[51] David C. Gompert and Martin Libicki, "Cyber Warfare and Sino-American Crisis Instability," *Survival* 56, no. 4 (2014): 7–22; Avery Goldstein, "First Things First: The Pressing Danger of Crisis Instability in U.S.-China Relations," *International Security* 37, no. 4 (2013): 49–89.
[52] James D. Fearon, "Rationalist Explanations for War," *International Organization* 49, no. 3 (1995): 379–414; Erik Gartzke, "War Is in the Error Term," *International Organization* 53, no. 03 (1999): 567–87; R. Harrison Wagner, "Bargaining and War," *American Journal of Political Science* 44, no. 3 (2000): 469–84.

operations—so states collaborate in making internet infrastructure more stable and reliable.[53]At the same time, the secrecy of cyber operations, their complexity, and the asymmetry of access to high quality intelligence may only exacerbate problems of uncertainty. The appearance of transparency in the information age may just be an artifact of the overall increase in the amount of secrets kept by states and firms. This is a question for future research.

4 The Stability-Instability Paradox in Cyberspace

In a half-century of internet history we have seen far more cyber espionage and crime than disruption and warfare—and more irritant than adjunct activity. The attacks that do occur are minor and reversible, like website defacement and service denial, rather than serious and destructive, like attacks on industrial control systems. The distribution of actual harms inflicted through cyber operations reflects a lot of minor aggressions and very little of anything major.

[53] Jon R. Lindsay, "Cybersecurity and International Relations: Evaluating the Threat from China," *International Security*, Forthcoming 2015; Daniel W. Drezner, "The Global Governance of the Internet: Bringing the State Back In," *Political Science Quarterly* 119, no. 3 (2004): 477–98.

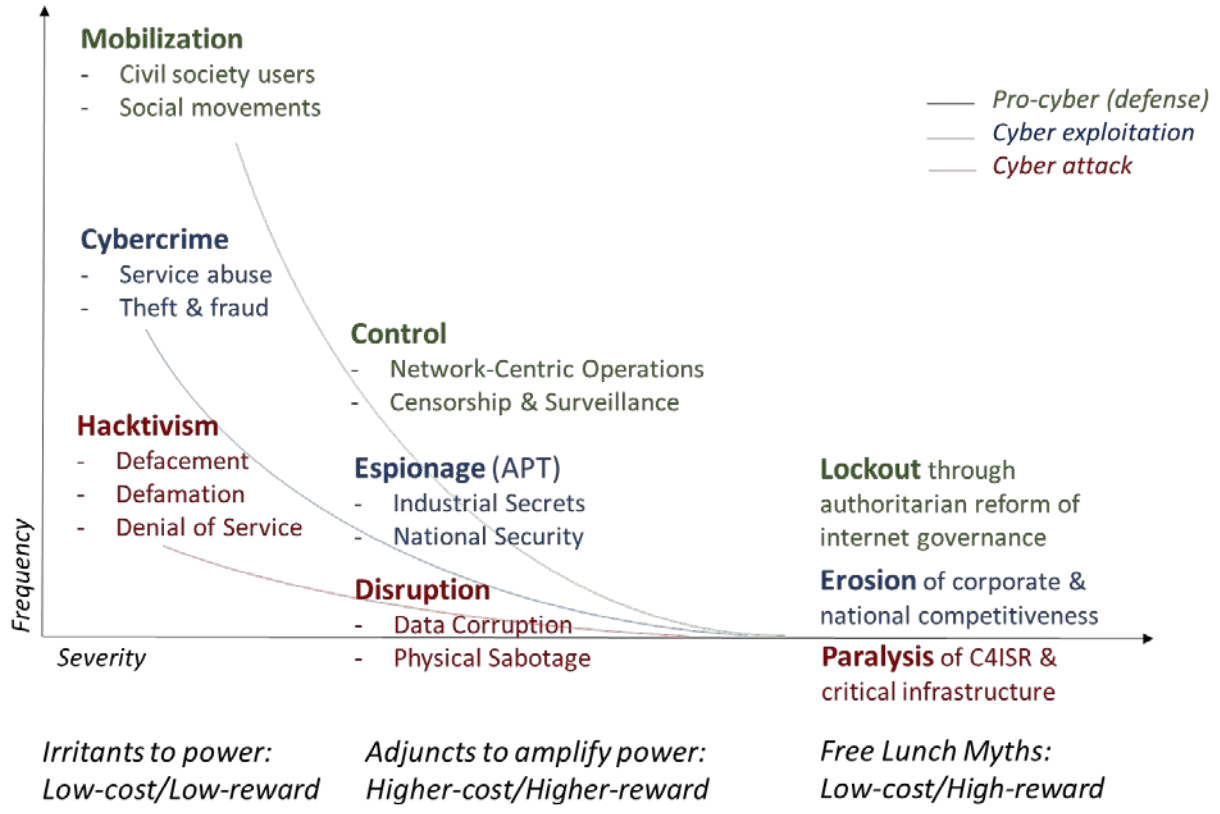

Figure 1: The distribution of cyber threats

This trend can partly be described by the barriers to entry to these different categories of harms. We distinguished irritant from adjunct cyber harms based on the low costs (and low rewards) of the former and the higher resource and effort requirements (and potentially higher rewards) of the latter. Anyone can get into cybercrime. It is a highly differentiated market to facilitate easy entry and exchange. By the same token, it's hard for most cybercriminals to make much money. High-end espionage is more complicated because it focuses on particular targets with heterogeneous networks rather than indiscriminate predation on homogenous assets. Even successful intrusions do not translate into successful absorption and application (thus fears of "erosion" of competitive advantage are largely mythical). Cyberwarfare is more complicated still and, for now, is mainly only a nation state competency. It is possible to imagine sophisticated non-state groups overcoming the technical, organizational, and intelligence support hurdles to

seriously disruptive attacks (even if "paralysis" remains out of reach for all), but they cannot take anonymity for granted if the offense is serious; moreover, digitally-savvy groups are more likely to find the internet useful in an adjunct supporting role than as a vector for major attack.[54] Apart from the inherent complexity of attack planning, moreover, defenders also tend to invest more effort protecting high-reward targets than they do against low-reward targets, which further exacerbates the differential barriers to entry.[55]

Operational costs are only half the story, however. Deterrence also helps to explain the distribution of cyber aggression. Restraint is built into strategic interaction in cyberspace, under most conditions. Cyber exploitation as a direct harm (i.e., "brute force" to change the balance of power) is likely to stay beneath the threshold at which the harm inflicted relative to the benefits of connection is great enough to trigger disconnection. The exception is certain situations where pervasive surveillance is used to compel miscreants to avoid using the internet altogether, thus realizing some benefit from disconnection. Usually, however, political economic actors want to continue to be able to use the internet productively even as they cheat at the margins of mutual agreement. Cyber attack as a direct harm will likewise be contained to situations where the disruption of computation is minor enough so as not to trigger cross domain retaliation (e.g., serious loss of life or incapacitation of critical infrastructure), or in situations where cyber disruption provides a tactical window of opportunity for a broader military operation. Cyber options are attractive as substitutes for force only when they are calibrated to enable the attacker to realize some benefit without the exposure to risks which the use of force usually involves

---

[54] David C. Benson, "Why the Internet Is Not Increasing Terrorism," *Security Studies* 23, no. 2 (2014): 293–328.

[55] Thus irritant attackers can use deception for offense with wild abandon (where defenses are poor) while actors with seeking to maintain adjunct control can use deception for defense (to enhance protection). To the degree that cyberspace is offense-dominant at all, it is only at the lower level of attack intensity. We develop this argument in Gartzke and Lindsay, "Weaving Tangled Webs."

(e.g., using Stuxnet to degrade Iranian enrichment rather than an airstrike which would have surely foreclosed diplomacy and encouraged retaliation by terrorist proxies). Limited cyber attacks are most appealing (to those who have the capacity to conduct them) when aggressors are deterred from using more violent measures.

The combination of cross domain deterrence and voluntary connection to the internet gives rise to a variant of the classic stability-instability paradox. In Glenn Snyder's original articulation of the paradox, mutually assured destruction could deter nuclear war. However, MAD was not credible for, and might even encourage, limited conventional war.[56] Or as Robert Jervis puts it, "To the extent that the military balance is stable at the level of all-out nuclear war, it will become less stable at lower levels of violence."[57] Cyber capacity is a poor substitute for nuclear weapons, myths of paralysis notwithstanding, yet there is a similar logic constraining the distributions of harms which are possible via information technology. To extend this logic to the cyber domain, there are a variety of deterrent mechanisms contain the most disruptive types of cyber attacks yet fail to contain, and even enable, a wide variety of online espionage, subversion, symbolic protest, and criminal predation. In cyberspace we observe a rather stable damage contest (i.e., no "paralysis" and limited "disruption") but a very unstable intelligence-counterintelligence contest (lots of "espionage" and "fraud" vying with efforts at "control" and "mobilization"). Thus the actors that have the ability to carry out highly destructive cyber attacks (mainly state actors for now) lack the motivation to attack; by contrast, these same actors as well as many different actors have both the ability and motivation to inflict irritant aggression with little fear of suffering consequences By and large, cyber options fill out the lower end of the

[56] Glenn H Snyder, *Deterrence and Defense: Toward a Theory of National Security* (Princeton, N.J.: Princeton University Press, 1961).
[57] Robert Jervis, *The Illogic Of American Nuclear Strategy* - (Ithaca, NY: Cornell University Press, 1984), 31.

conflict spectrum where deterrence is not as credible or reliable. The very few cases of physically disruptive cyber attack we do observe—mainly powerful states conducting covert action or battlefield support operations against militarily weaker opponents—have notably involved stronger actors who not only have the capacity to plan and conduct a sophisticated attack but also have the ability to deter retaliation against their use of cyber attack.

This cyber variant of the stability-instability paradox has a slightly different logic, however. In the nuclear realm, actors cannot disconnect from the threatened harm, and this is what makes the threatened destruction both mutual and assured. When there are many missiles with many warheads, the chance of intercepting them on the ground through a disarming counterforce strike or in the air through ballistic missile defense with any confidence becomes vanishingly small. Not so in cyberspace, where connection to the internet or acceptance of connections through it is voluntary. There is "no forced entry in cyberspace" in Libicki's phrase, and so the hundredth cyber attack against a closed vulnerability is as ineffective as the first. Attackers thus rely on deception to exploit vulnerabilities and ensure that they stay open. However, offensive deception can fail in the "fog of cyberwar" and defenders can be deceptive as well, both of which are more likely against high-reward targets (where cross domain deterrence also more credible). The need to preserve internet connections to facilitate ongoing and future deception as well as the need to preserve stealth to avoid the consequences of getting caught imposes discipline on attackers.

Actors cannot enjoy the substantial benefits of interconnection without accepting some risk of exploitation (hacking to spy) and attack (hacking to disrupt). Thus the successful "lockout" of the internet, with advantage accruing exclusively to one political group or another, is not realistic. Moreover, because these harms share similar techniques, the observed abundance

of the former represents a latent potential for the later. The latent escalatory potential of even minor irritants leads to rampant fears of unrestrained catastrophe, to be sure. Yet this latent potential is difficult to harness for targeted coercion because the threat is self-effacing. Declared cyber threats that highlight the vulnerability to be exploited are readily mitigated. Instead, the ineradicable threat of cyber catastrophe (ineradicable as long as the internet continues to be useful) creates a general if diffuse deterrent effect among all parties who value their connection to the internet. No one who wants to make money on the internet really wants to have a cyberwar, and this includes states as well as criminals.

Which types of actors are most able to benefit through internet coercion and which are most vulnerable to coercion? Large powers like the U.S. are highly dependent on the internet but also highly skilled at inflicting harm, both through cyber and traditional military force. Poor powers across the digital divide may have little vulnerability at all, while medium powers may have vulnerability but lack a range of forces to deter attacks. This might imply a "curvature" to the utility of cyber coercion. Big-capable countries are vulnerable to cyber harm but can deter through other military instruments. Poor states are not vulnerable. It may be the prosperous small or digitally developing who are in trouble, since they cannot credibly deter and have high dependence on the internet. The information revolution is often thought to be a boon to non-state actors, and indeed it is, but mainly in the irritant class of cyber operations. Moreover, the increasing ubiquity and sophistication of information technologies can be expected to have something of a democratizing effect on intelligence and counterintelligence techniques whereby firms and citizens will have access to and be concerned about the types of things that were historically the purview of obscure state intelligence agencies. However, it would be a mistake to use the increasing ferment of low-intensity information contests to infer the shape of higher

intensity activity. On the contrary, the traditional logic of war will continue to dominate the expression of cyber aggression.

Because threatened internet harms depend on voluntary connections in the first place, and as many actors have alternative means to inflict (cross domain) harm in retaliation, the coercive utility of cyberspace is actually somewhat limited. At the same time an ever increasing variety irritants and more temperamental adjuncts becomes available for global political interaction. The "net" result is that opponents have strong incentives to impose costs via the internet but also to keep those costs low enough to preserve interconnection and avoid retaliation. Therefore, contests in damage will remain relatively stable while contests in intelligence will be increasingly unstable. The human-built world is becoming more complex, to be sure, but it is not necessarily more dangerous. As long as it is desirable to connect to the internet tomorrow, there will be only limited harm via the internet today.