The University of Chicago

Cyber Pearl Harbor or Cyber Hiroshima

Coercion and Escalation in Cyberspace

By Jonathan Katz

August 2023

A Paper Submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts Program in the Committee on International Relations

Faculty Advisor: Paul Poast

Preceptor: Burcu Pinar Alakoc

Abstract

Technological advancement in the arena of cyber weaponry has opened the door for new strategies of coercion to be utilized by states. The theory presented in this thesis attempts to understand the future practice of coercion in cyberspace. It argues that destabilizing offensive cyber operations advances the ability of states to achieve coercive success without the necessity of employing conventional kinetic weaponry. A rebuttal to the notion that a destabilizing offensive cyber operation will resemble the attacks on Pearl Harbor is presented. We argue that such an attack would more so resemble the bombing of Hiroshima. This is because such an attack would be chosen in attempt to minimize the loss of civilian life and would simultaneously use all four major strategies of coercion. This conclusion calls into question arguments supporting the "Cyber Pearl Harbor" narrative as well as previous studies which have dismissed the strategic and tactical utility of cyber weaponry. Illustrative case studies concerning the effectiveness of economic sanctions and the economic damage of large-scale attacks, as well as the bombing of Hiroshima and the attack on Pearl Harbor, provide initial validation of the theoretical arguments made here.

Contents

"The most destructive scenarios involve cyber actors launching several attacks on our critical infrastructure at one time, in combination with a physical attack on our country. Attackers could also seek to disable or degrade critical military systems and communication networks. The collective result of these kinds of attacks could be a "cyber Pearl Harbor:" an attack that would cause physical destruction and the loss of life. In fact, it would paralyze and shock the nation and create a new, profound sense of vulnerability".[1]-Secretary of Defense Leon Panetta

1: Introduction

Since the dawn of the information age the internet has been seen as the frontier of future warfighting. New ways to conduct espionage, and target: civilian, government, and military infrastructure were foreseen, and quickly realized. One of the earliest such examples was seen during the first Gulf War, when hackers using a bug developed two years prior hacked into DoD and DoE servers and accessed information on troop movements, and weapons systems which they then attempted to sell to Saddam Hussein.[2] At the end of the decade another strategy was seen, distributed denial of service attacks (DDoS). Launched as a response to US and NATO action in Kosovo, pro-Serbian hackers rendered the NATO public affairs website inoperable, defaced the White House's website, and destroyed information on UK databases.[3] This attack represented the first cyber operation by a warring party against an opponent in the midst of conflict. Yet the greatest fear of analysts and government/military officials was not attacks such as these which they had experienced, rather it was the potential for a surprise offensive cyber

[1] Leon E. Panetta, "Defending the Nation from Cyber Attack" (speech, New York, New York, October 11, 2012) Business Executives for National Security, https://nsarchive.gwu.edu/document/21479-document-78.
[2] The Evolution of US Cyber Power. https://www.afcea.org/committees/cyber/documents/TheEvolutionofUSCyberpower.pdf.
[3] Ibid

operation (OCO) by a rival power aimed at disrupting, destroying, and/or controlling civilian and military infrastructure, often dubbed as a "cyber Pearl Harbor" or "cyber 9-11".

Cyberspace is often seen as an offensively dominant domain, in which the attacker will always have the advantage because defense is aimed at preventing the use of offensive strategies that have been used previously, rather than the potential strategies attackers can develop in the future. Essentially, the defender can only ever be a step behind the attacker. Furthermore, the defense is the only side with anything to lose in the dynamic, there is no loss of life and little loss of money when compared to a conventional attack. Even when an attack is thwarted, the attackers gain information on how to make a future attack successful.

The element of surprise is seen as a natural benefit of offense in cyberspace because of the lack of ability to gain intelligence on the development of adversaries' cyber capabilities, and their plans for future OCOs. While adversaries' conventional military developments can be monitored through many methods of intelligence collection, human intelligence (HUMINT) is the only possible method to gain intelligence on opponents' cyber capabilities. This is extremely difficult due to the small size of teams working on OCOs within their respective nation's intelligence and defense agencies. Furthermore, attackers can leave bugs and viruses in defenders systems for long periods of time, sitting dormant until an attack is launched, as was seen in the Stuxnet attacks.

Stuxnet was a virus developed by the United States and Israel beginning in 2005 and executed against Iran in 2010, aimed at slowing Iranian centrifuges. It has ultimately been painted as a failure due to its ability to only destroy a fifth of the centrifuges, only disrupting the Iranian nuclear program for about 6 months.[4] This has led some to argue that the experience with

[4] Slayton, Rebecca. "What Is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment." *International Security* 41, no. 3 (2017): 72–109.

Stuxnet has shown that OCOs are an ineffective form of disruption.[5] But, we should not assume that the attack was launched to completely knock the Iranian nuclear program out forever, bugs and viruses are always patched, and Stuxnet naturally made its way to infecting aspects of the public internet leading to a need for the best private cybersecurity companies to defeat it, which they did quickly. Rather, we should see OCOs as part of a coercive strategy aimed at demonstrating resolve, which Stuxnet successfully achieved. Just 3 years later, Iran began talks to end its nuclear program, which had been non-negotiable prior to the Stuxnet attacks.

OCOs will continue to be a major aspect of cyber-capable states coercive arsenal, eventually reaching a point where such operations are used to induce destabilization of an adversary's society. But what will a destabilizing OCO look like? Will destabilizing OCOs present states with an easier ability to achieve coercive success? How will an exchange in cyberspace lead to the use of a destabilizing OCO? Will destabilizing OCOs succeed where past strategies of coercion have failed? This paper seeks to answer such questions and provide a research agenda for the future.

1.1: Assumptions and Definitions

Prior to the discussion of this paper's theory, it is important to lay out a few basic assumptions and engage in some definitional discussion. Starting off, this theory assumes that states are the primary actors. While recent history has shown that there have been significant cases of cyber operations carried out by foreign non-state actors, actions such as these rarely have coercion or even political goals in mind, rather they are aimed at profiting off of holding large corporations' networks or information for ransom. There are some cases of non-state actors who in the past launched ransomware attacks and later carried out cyber operations without

[5] Ibid

demanding ransom, we can assume that in these instances they were acting on behalf of a state. We have seen this notably with groups located in Russia who have attacked American or Ukrainian government agencies, but the nature of these attacks and the lack of ransom being demanded should lead us to assume that these groups are operating under orders, and likely being funded by their government. Otherwise, they would have no incentive to forgo strategies that earn them significant profit merely for their state's well-being. Criminal organizations such as these are motivated by one thing, profit. Even in cases where these organizations do demand ransom, such as the Colonial Pipeline hacks, we should be wary of jumping to the conclusion that their home government had no involvement. We should hold cases such as these in a similar light to cases of conventional terrorism where the groups are sponsored by foreign governments who have incentives to attempt to impact the state in which the group is operating or the states which the group targets. This does vary across states, in states with strong cyber capabilities it may be less likely that a non-state actor is acting on behalf of the state, but there should always be skepticism regarding whether there was involvement of an adversaries government in an attack.

This is a neorealist theory of coercion that assumes states act rationally the overwhelming majority of the time. Rationality here is not measured through traditional means of expected utility maximization, rather it follows the Mearsheimer-Rosato model of rationality. The Mearsheimer-Rosato model attributes rationality to states following credible theories, and deciding on actions through a deliberative process.[6] Importantly when discussing coercion, despite significant debate over the effectiveness of strategies of coercion we still cannot deem any of them to be non-credible. Here irrationality would likely be seen when the theory of why

[6] Mearsheimer, John J. and Sebastian Rosato, How States Think: The Rationality of Foreign Policy (forthcoming from Yale University Press)

coercion should be used is non-credible. For example, if a communist state emerged and a state attempted to coerce them due to fears stemming from the domino theory, that would be irrational, as the domino theory has been widely discredited. In addition to this, it could be possible that the decision to attempt coercion would be made without a deliberative process. This is the most likely scenario where irrationality would emerge, as a coercive strategy may need to be implemented quickly in order to have the highest chance of success, leaving no time for fruitful deliberation. Despite this, history has shown very few cases of states acting irrationally, and some of the most significant cases of coercion such as the attacks on Pearl Harbor, and the USSR's stationing of nuclear weapons in Cuba have been highly rational and fulfilled the aforementioned criteria to a tee. Lastly, rationality is not based on outcomes, coercion may fail but its failure cannot singlehandedly deem the attempt to have been irrational.

A key distinction between this theory, and other theories of coercion, most notably Robert Pape's, is the temporal focus. This theory is solely focused on coercion prior to the outbreak of war, while those such as Pape look at coercion both before and during war, viewing any cessation of conflict prior to decisive victory as successful coercion. That is not to dispute the findings of those such as Pape, in fact, a major aspect of this theory revolves around the importance of changing a state's probability of benefits through strategies of denial, a notable claim of his.[7] Rather, this theory assumes that states place a premium on avoidance of direct conventional conflict and will engage in significant attempts at coercion through non-conventional means most notably OCOs.

The final and most important assumption is that states place a premium on survival, and will do anything possible to ensure their survival. This is a direct link to the prior assumption as

[7] Pape, Robert A. *Bombing to Win: Air power and Coercion in War*. Cornell University Press, 1996.

an adversary's actions may threaten the survival of a state, incentivizing them to attempt to coerce them in order to change their behavior. But, it may also be the case that conventional war also threatens the state's survival providing the state with further incentive to attempt non-conventional forms of coercion, minimizing the likelihood of an adversary launching a conventional attack. It is possible that this logic could guide a weaker state in a strategy of coercing a conventionally stronger state, as well as a non-nuclear state attempting to coerce a nuclear state.

Due to the recent emergence of cyber conflict, significant definitional work is vital to the formation of theories aimed at explaining and predicting coercion in cyberspace. Coercion is any attempt to alter the behavior or decisions of another state, if this is achieved coercion may be deemed successful. As previously mentioned an OCO is an operation utilizing cyber weaponry to attack the digital infrastructure of another state. Importantly, the sole focus here is OCOs which attempt to impact the critical infrastructure of another state or government agencies. While some OCOs are launched in order to gain access to an adversary's sensitive information this has little coercive value except in extreme cases. State secrets get leaked often, especially in the digital age, and there is little evidence to suggest that a state has succumbed to coercion to prevent information from being leaked.

Important to this study is which OCOs would be considered to be destabilizing. A destabilizing OCO is an attack that is aimed at wreaking havoc on a state through attacks on critical infrastructure, affecting not only civilians, but also the government, with the goal of disrupting them for a sustained amount of time. For example, an attack on an adversary’s power grid while also attacking command, control, and communication (C3) systems would be a destabilizing OCO. Even an attack on just a power grid may be destabilizing so long as the

disruption is sustained for a period of time which greatly impacts civilian life and government response. Put simply, destabilization is attributed to two things, target, and duration. This brings the question of when an OCO is not destabilizing. Attacks on private companies or banks would likely not be destabilizing unless significant damage was done to the economy with little likelihood of a speedy recovery. In addition to this, the Stuxnet attacks were not destabilizing even if the goal of the attacks had been achieved, as they targeted nuclear centrifuges at one nuclear facility causing little to no effect on civilians, and not reducing the ability of the Iranian government to launch a conventional response. In a sense a destabilizing OCO would attempt to paralyze all levels of the target state's society.

Key to this theory is the dichotomy between a “Cyber Pearl Harbor” and what I have termed “Cyber Hiroshima ". As previously mentioned, “Cyber Pearl Harbor” is an often-mentioned possibility of what might emerge as the future of warfighting, this thesis began with a description of what Former Secretary of Defense Leon Panetta viewed such an attack to look like, it is essentially a metaphor for a destabilizing OCO that would be coupled with a conventional attack, causing significant physical damage and loss of life. It is very possible that such an attack would not be coercion at all rather it could just be a method of initiating a conflict. If anything this metaphor may be a little misleading as the Japanese did not decide to launch the attacks on Pearl Harbor to instigate conflict, rather they launched the attacks to attempt to coerce the United States into ending its embargo on Japanese oil which was threatening Japanese survival. Even more misleading is the lack of clarity regarding who will launch such an attack, Panetta seemingly alludes to the possibility that non-state actors may utilize such a strategy, a view shared by some scholars who have coopted the phrase referring to such an attack as a

"Cyber 9/11".[8] But what is most important to us is timing, those who view such an attack as a possibility seemingly view it as the first step, the key difference from a "Cyber Hiroshima".

A "Cyber Hiroshima" is a last-resort attempt at coercion aimed at using a destabilizing OCO to coerce an enemy through not only infrastructural damage but also shock and awe. While obviously this metaphor is not perfect because we are looking at coercion prior to war, and Hiroshima occurred while war was ongoing, Hiroshima was an act of last resort aimed at avoiding conventional means of coercion (invasion and occupation). It is important to note that a "Cyber Hiroshima" may cause loss of life or physical damage but that is not the main goal of the coercer. Such instances would be incidental, for example, an OCO which disrupts a power grid could very well lead to deaths say from traffic lights malfunctioning causing car accidents, but the attacker is not relying on such instances to bring about successful coercion. Rather, the attacker would view the damage done to critical infrastructure, digital infrastructure and the demonstration of unmatched cyber capabilities as bringing about successful coercion. The attacker would view such an attack as likely leading to successful coercion because the target state would have to face the possibility of future similar attacks if they do not comply with the attackers' demands. In addition, the target state may be unwilling to respond with conventional attacks or launch a war against the attacker for fear of continued and intensified attacks which would only decrease the likelihood of success in war.

1.2: Literature Review

The idea that cyberspace is an offensively dominant domain has been argued since early in the conception of cyberwarfare. Joseph Nye has argued that because of the internet's design

[8]Magee, Clifford S. "Awaiting Cyber 9/11." *Joint Force Quarterly* 70 (2013): 76-82; Leon E. Panetta, "Defending the Nation from Cyber Attack" (speech, New York, New York, October 11, 2012) Business Executives for National Security

focus being on ease-of-use the offense inherently dominates in cyberspace.[9] Liebarthal and Singer have pointed to the internet's role of enabling the easy flow of information as reason for offensive dominance in cyberspace, the internet simply was not designed to restrict the flow of information.[10] Some have even argued that there should be a reduction of effort in defensive cyber activities and more focus should be given to teaching better internet security practices such as two-factor authentication because the arbitrary complexity of defensive systems favors the attacker.[11] Slayton gives a scenario in which the attacker and defender are in a race to find a vulnerability, the defender needs to find and patch every vulnerability in the system, while the attacker needs to just find one it can exploit, therefore the attacker has a natural advantage.[12] Others such as Libicki argue that because the attackers have anonymity, deterrence against cyber-attacks is impossible, giving offense the advantage.[13] It is important to note that some in the offense dominance camp view deterrence as possible in cyberspace because of the risk of facing a strong conventional or even nuclear response, arguing that because of this the overall distribution of power is unlikely to change.[14]

Yet some argue cyberspace is defensively dominant because of the difficulty in pulling off complex cyber-attacks on civilian infrastructure. Effective coordination is required, any mistake will alert the defense of the attacker's presence, ruining the attacker's plans. Furthermore, their weapon can only be used one time because of the "use-and-lose" quality of

---

[9] Nye, Joseph S. *Cyber power*. Cambridge: Harvard Kennedy School, Belfer Center for Science and International Affairs, 2010.

[10] Lieberthal, Kenneth, and Peter Warren Singer. *Cybersecurity and US-China relations*. Brookings, 2012.
[11] Slayton, Rebecca. "What Is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment." *International Security* 41, no. 3 (2017): 72–109.
[12] Ibid
[13] Libicki, Martin C. *Cyberdeterrence and Cyberwar*. RAND Corporation, 2009.
[14] Kugler, Richard L. "Deterrence of Cyber Attacks." *Cyberpower and National Security* 320 (2009): 309-340.

cyber weapons.[15] Rebecca Slayton argues that there is not an inherent advantage in cyberspace, and the advantage is given to the side which has better organization and technical skill.[16] Slayton argues that while defense may require more manpower and thus more organization and better coordination, this alone does not translate to an offensive advantage.[17]

Many proponents of the "Cyber Pearl Harbor" metaphor view such an attack as a possibility due to the ability cyberwarfare provides weaker states to effectively target stronger states.[18] This is because cyberwarfare allows the weaker state to damage a stronger state with a lower risk of the target striking back.[19] Goldman and Warner argue that this is likely, because in the buildup to conflict a weaker state will be incentivized to delay the stronger state by preemptively attacking their military, and cyber means provide this ability without risking public outrage.[20] Wirtz argues that the surprise inherent to a "Cyber Pearl Harbor" is possible despite the fact that we can imagine it, because decision-makers have fallen victim to surprise attacks in the past due to intelligence failures.[21]

Opponents of the "Cyber Pearl Harbor" metaphor view such an attack as unlikely because companies and governments have invested so heavily in cyber security that an exploit large enough to allow for such an attack would likely have been found by now.[22] In an early article refuting the possibility of an "Electronic Pearl Harbor" George Smith argues that hacking into a system you do not understand undermines the attacker's ability to achieve much success at

---

[15] Rid, Thomas. "Cyber War Will Not Take Place." *Journal of Strategic Studies* 35, no. 1 (2012): 5-32.
[16] Slayton, Rebecca. "What Is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment." *International Security* 41, no. 3 (2017): 72–109.
[17] Ibid
[18] Wirtz, James J. "The Cyber Pearl Harbor." *Intelligence and National Security* 32, no. 6 (2017): 758-767.
[19] Rattray, Gregory J. *Strategic Warfare in Cyberspace*. MIT press, 2001.
[20] Goldman, Emily O., and Michael Warner. "Why a Digital Pearl Harbor Makes Sense... and Is Possible." *George Perkovich and Ariel E. Levite Understanding Cyber Conflict* 14 (2017): 147-157.
[21] Wirtz, James J. "The Cyber Pearl Harbor Redux: Helpful Analogy or Cyber Hype?" *Intelligence and National Security* 33, no. 5 (2018): 771-773.
[22] Kallberg, Jan. "Bye bye, Cyber Pearl Harbor." (2021).

all.[23] This logic has been used to argue against the perceived ease for attackers to successfully gain control of military and civilian infrastructure systems.[24] Others have argued that fears related to a "Cyber Pearl Harbor" have distracted from actually protecting against real threats in cyberspace such as election interference.[25] Many constructivist scholars have viewed the rhetorical usage of the metaphor through the lens of securitization theory which holds that security threats are not predetermined, rather they are constructed through political discourse by political leaders for political purposes.[26]

There is disagreement over whether the public would support an escalatory cyber-attack in response to a cyber-attack by an adversary. Hedgecock and Sukin found that survey respondents had similar support for a military response to a cyber-attack as they did for a military response to a kinetic attack.[27] Because of this, they argue that it is necessary to maintain covertness in cyberspace.[28] On the other hand, Leal and Musgrave have found that survey respondents had stronger support for a cyber response to a cyber-attack than a military response.[29] Furthermore, they found that respondents supported a more drastic response when the attacker was an individual or terrorist organization than a state.[30] They also found that

---

[23] Smith, George. "An Electronic Pearl Harbor? Not Likely." *Issues in Science and Technology* 15, no. 1 (1998): 68-73.
[24] Slayton, Rebecca. "What Is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment." *International Security* 41, no. 3 (2017): 72–109.
[25] Nye, Joseph S. "*The Kremlin and the US Election*." Project Syndicate, August 30, 2017.
[26] Lawson, Sean, and Michael K. Middleton. "Cyber Pearl Harbor: Analogy, Fear, and the Framing of Cyber Security threats in the United States, 1991-2016." *First Monday* (2019).
[27] Hedgecock, Kathryn, and Lauren Sukin. "Responding to Uncertainty: The Importance of Covertness in Support for Retaliation to Cyber and Kinetic Attacks." *Journal of Conflict Resolution* (2022)
[28] Ibid
[29] Leal, Marcelo M., and Paul Musgrave. "Hitting Back or Holding Back in Cyberspace: Experimental Evidence Regarding Americans' Responses to Cyberattacks." *Conflict Management and Peace Science* 40, no. 1 (2023): 42-64.
[30] Ibid

respondents' response preferences have a positive correlation to the damage and deaths caused by a cyber-attack.[31]

This past work is crucial to understanding where the arguments of this theory are rooted. This theory inherently falls on the offensive dominance side of the offense-defense debate, while also taking significant consideration of the work done by scholars such as Slayton who take a defensive dominance or an agnostic stance in the debate. Factors such as the "use-and-lose" quality of cyber weapons play a pivotal role in the process which will be laid out in this thesis. Furthermore, this theory will reject the notion that surprise alone is an offensive advantage, and whether a destabilizing OCO will even be a surprise at all, rebutting arguments made by supporters of the "Cyber Pearl Harbor metaphor" and many within the offensive dominance camp. This will play a role not only in the process which will be laid out in the following section, but also in the argument for why a destabilizing OCO will resemble a "Cyber Hiroshima" more than a "Cyber Pearl Harbor". Lastly, civilian response preference will play a significant role in understanding how a destabilizing OCO may lead to coercive success. Leal and Musgrave found a positive correlation between the damage and deaths, and the response preferences of citizens. While not rebutting their findings, this theory will question how the lack of physicality seen in the damages, and the lack of direct deaths may allow for the attacker to achieve coercive success.

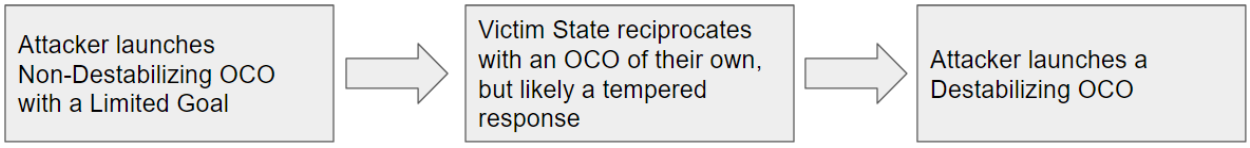
2: Theory

I argue that an OCO aimed at the destabilization of an adversary's society will resemble a "Cyber Hiroshima" not a "Cyber Pearl Harbor" as has often been argued. This is because such an action is most effectively used for last-resort coercive purposes in an ongoing exchange short of conflict. Put simply, a destabilizing OCO will be the final stage of a cyber exchange, rather than

[31] Ibid

the introductory stage. Unlike the Stuxnet attacks, a destabilizing OCO will be characterized by the simultaneous disruption of multiple layers of civilian and/or military infrastructure such as power grids, internet infrastructure, communication systems, transportation systems, and natural gas pipelines. This will likely be the result of an ongoing reciprocal exchange between adversaries which would begin with lower level OCOs such as DDoS attacks, aimed at demonstrating resolve in the pursuit of a limited goal. The theoretical process I have proposed is illustrated in Figure 1 below.

Figure 1: Escalation Process



Using the Stuxnet attacks as an example I will demonstrate the manner in which this dynamic will play out. The United States began the development of Stuxnet at the request of the Israelis who were attempting to gain President Bush's permission to launch a conventional strike against Iranian nuclear facilities. Seeing the dangerous potential outcomes of such an attack the Bush administration ordered the development of an alternative method of disrupting the Iranian nuclear program, opting to go the route of cyber disruption in order to demonstrate American and Israeli resolve in preventing the continued development of the Iranian nuclear program without the potential for escalation that military action would bring. Initially, the coercive success was not seen, and Iran launched a reciprocal OCO, as predicted by my theory. But due to the recent development of an Iranian military cyber unit, these were rudimentary DDoS attacks on American banks. Ultimately the American demonstration of resolve was successful and three years later the Iranian nuclear talks began and were a success much to the dismay of the Israelis. But what if the attacks had failed, and Iranian nuclear weapons development continued?

This exchange mirrors the likely scenario predicted by my theory for a few key reasons: the initial attack was an attempt at coercion aimed at achieving a limited goal of the attacker (ending Iranian nuclear development), the non-destabilizing nature of the initial attack, and a retaliatory attack by the victim state that was somewhat tempered due to the lack of destabilization caused by the initial attack. This exchange fulfilled the first two steps of the coercive process predicted by my theory, but the third, a destabilizing OCO was not reached because the United States ultimately succeeded in its attempt to coerce Iran.

If Stuxnet had failed at coercing the Iranians we likely would have seen another OCO but this time one aimed at complete disruption of Iranian civilian infrastructure in order to disrupt military infrastructure. The Iranian power grid likely would have been targeted in order to disrupt Iranian military communication and command and control systems (C3). This attack could have been executed in order to support an Israeli offensive, or in an attempt to coerce Iran to end its nuclear program without having to participate in conflict against the United States. While both motivations may have been possible, the latter is more likely. This is because the United States had begun its cyber operation in order to avoid the potential for conventional conflict between Israel and Iran which the United States likely would have become involved in. This may sound like a hypothetical scenario, but recently revealed information has shown that the United States had such a plan dubbed Nitro Zeus in the works, and in active development had Iran not agreed to begin negotiations on their nuclear program.[32]

Three factors make such an exchange and destabilizing attack a likely outcome of continued offensive cyber development. First, there are massive difficulties in attributing attacks in cyberspace to specific actors. Because of this, states will be incentivized to follow President

[32] Sanger, David E., and Mark Mazzetti. "U.S. Had Cyberattack Plan If Iran Nuclear Dispute Led to Conflict." The New York Times, February 16, 2016.

Bush's lead and initially attempt an OCO for a limited goal, which they may or may not fail to achieve. Second, OCOs have the potential to successfully coerce through punishment, denial, manipulation of risk, and leadership decapitation, all four major strategies of coercion. Third, OCOs have the ability to destabilize an adversary's society through attacks on internet infrastructure, civilian infrastructure, and military infrastructure without the need to risk loss of the attackers' life or resources, such as pilots, planes, or drones, or the lives of civilians within the target state. No other domain of warfighting provides all of these benefits.

2.1: Cyber Coercion through the Lens of Aerial Coercion Strategies

Naval warfighting is the only other domain that possesses the first factor, attribution difficulties, through difficulties in tracking submarines and the ability for ships to falsify Identify Friend or Foe (IFF) systems. There are some attribution difficulties at sea, yet these are short lived and defenders can overcome them quickly through simple visual checks or the use of codes. It can be argued that naval forces can successfully coerce through punishment and denial, but that is not true in all cases. Say for example the target state is non-coastal, naval artillery may not be able to reach their target from sea. Furthermore, the use of conventional kinetic munitions may be entirely ineffective in coercion. As Robert Pape has argued, air strikes aimed at coercion through punishment have failed because they lead to an increase in nationalism and civilian support of the government.[33] While aerial power certainly can coerce through denial and possesses destabilizing capabilities, it is almost certain that the sheer loss of life that such strikes would cause guarantees an escalation to conflict especially among major powers.

[33] Pape, Robert A. *Bombing to Win: Air power and Coercion in War*. Cornell University Press, 1996.

When engaging in comparison between the prominent domains which have coercive capability, and the strategies which may be used, it is important to lay out Robert Pape's formula which explains when a state will likely achieve a coercive victory. This formula is laid out as follows: R = B p (B) – C p(C); where R= the value of resistance; B= the value of the benefits of resistance; p (B) = the probability of gaining the benefits of resistance; C=the value of the potential costs resistance; p(C) = the probability of suffering the costs of resistance.[34] Pape posits that coercion succeeds when it is costlier to continue resistance rather than surrender. Importantly he argues that this is most effectively done through attempts to alter the probability of benefits, which is best done through a strategy of denial. He argues that attempts to alter cost, often done through engaging in a strategy of punishment, entailing significant damage to civilians and civilian infrastructure, is ineffective and leads to significant backlash.

Interestingly, cyberwarfare pits not just the strategies of denial and punishment against each other, but also the strategy of leadership decapitation. John Warden in his original conception of the strategy was heavily focused on using the strategy to target and paralyze the leadership apparatus of adversarial states, believing that it would lead to complete paralysis of the state. Additionally, the strategy of leadership decapitation is most often thought of as targeting and killing the leader, or those high up in the hierarchy of leadership, but execution is not a necessary aspect of the strategy. Warden argues that in order to achieve coercive success the most important element must be targeted first, leadership, which impacts the four other major subsystems (in order of importance): production, infrastructure, population, and fielded forces.[35] Any damage to one of those five subsystems (rings) will cause negative effects across the

[34] Ibid
[35] Warden, Col. "John A. III, USAF.“The Enemy as a System.”." *Airpower Journal* (1995).

system.[36] Although execution is heavily effective at this, anything which cuts off leadership from the other four subsystems will have widespread ramifications. This is because in order for the four subsystems to function effectively they need to communicate with the leadership and gain guidance from the leadership.

A destabilizing OCO, much like strategic bombing, can alter the probability of benefits by targeting military infrastructure. But, unlike strategic bombing campaigns, such an attack would likely target C3 systems cutting off communications between aerial units who would likely respond to the attack conventionally from their command. In addition to this, it's possible that the ability to launch a nuclear response could be impacted, again incentivizing a state to attempt coercion through cyber means. This would mean that any attempt at using an OCO for denial would also likely invoke the logic of leadership decapitation, this convergence would not only alter the probability of benefits but damage could trickle down throughout the rest of the system. For example, in a destabilizing OCO as envisioned by this theory, power grids would likely be targeted as well as communication systems such as television, radio, and internet, and military communication systems such as C3 systems. If this were to happen the leadership is effectively cut off from communicating with their citizens as well as military units. Even if leadership could meet, and military plans could be drawn up, it would take a significant amount of time to distribute the plans and execution would be extremely difficult, all while a crisis of uncertainty is unfolding within the populace. While this military response is being attempted, every other aspect of society: production, infrastructure, civilian communication, etc. will need to have recovery plans drawn up and distributed across the country, possibly without the ability to use rail, air, or sea transportation.

---

[36] Ibid

At the same time, a destabilizing OCO would alter the probability of costs as the attacker has demonstrated their ability to launch significant OCOs against the target state, something which could potentially be done again if the target state continues to resist. This is similar to what Thomas Schelling described as the manipulation of risk, which in a conventional war is illustrated as continuously bombing closer and closer to the target state's cities.[37] Lastly, a destabilizing OCO could also alter costs, by causing economic damage by shutting down power grids and other systems crucial for the functioning of industry, as well as inducing infrastructure paralysis which would have widespread effects on the target state's economy and civilian welfare, without directly inducing loss of life. Bringing about a change in the most problematic aspect of coercion through punishment, the targeting of civilians. All of these strategies can be achieved in other domains, but cyber possesses unique capabilities to use these strategies effectively for two reasons: a single destabilizing OCO can achieve the objectives of all four of these strategies at the same time, and it can be done without directly causing the loss of civilian life.

2.2: Methodology

While formal testing of this theory remains impossible due to the lack of uses of a destabilizing OCO and the classified nature of relevant information, it is still possible to engage in a test of the theory which has been presented. Plausibility probes in the form of brief, illustrative case studies allow us to test the initial validity of a theory which is unable to undergo more significant testing.[38] The following two sections use this approach in order to bolster two key arguments: the effectiveness of destabilizing OCOs for economic coercion, and how a

[37]Schelling, Thomas C. Arms and Influence. Yale University Press, 2020.
[38] Levy, Jack S. "Case studies: Types, designs, and logics of inference." Conflict management and peace science 25, no. 1 (2008): 1-18.

destabilizing OCO would resemble the bombing of Hiroshima, more so than the attack on Pearl Harbor.

Through a brief case study and analysis of past methods and instances of economic coercion we can see exactly how a destabilizing OCO would be more successful at coercion. Furthermore, we are able to see how the economic damage such an attack would inflict is far more severe than any similar attack or use of sanctions in the past. Importantly, we see how the shortcomings which cause sanctions to fail would be avoided through the use of a destabilizing OCO. The following case studies on the bombing of Hiroshima and the attack on Pearl Harbor illustrate how the use of a destabilizing OCO would resemble the former more so than the latter. Two key points of comparison are established in this case study: simultaneous employment of all four major strategies of coercion, as well as attempts to minimize civilian lives loss.

2.3: Cyber Coercion versus Economic and Nuclear Coercion

Avoiding directly causing the loss of civilian life will be a primary reason states will employ a coercive strategy using cyber weaponry. As previously mentioned, coercion through punishment is often ineffective through conventional means. The striking of a single or few cities at a time outrages citizens throughout the rest of the state, increasing support against the attacker, making the possibility that a government will bow to the commands of the attacker highly improbable because of the severe loss of citizen support for the government it would entail. While nuclear deterrence revolves around the strategy of punishment because of the scale of damage and significant loss of life caused by nuclear weapons, this logic does not hold true for conventional weapons. This is because when State A launches a nuclear strike on another nuclear power it is essentially making the decision to incur high civilian costs of its own when State B

responds. In a sense, State A is inducing its own punishment when it chooses to initiate a nuclear exchange.

Using this logic to understand why a state would choose to launch a destabilizing OCO, we see the opposite scenario emerge. State A would launch a destabilizing OCO, while State B would be left with four options: respond with conventional weapons, respond with nuclear weapons (if they are a nuclear power), do nothing, or follow State A's demands. If State B chooses the first option this will likely be ineffective unless the damage caused by the attack outweighs State A's resolve. If they choose the second option they will just incur significantly higher losses as State A will be forced to launch a nuclear response of its own. If they do nothing then the state opens itself up to continued attacks until the costs outweigh the value of resistance. Their best option will be to cede to State A's demand unless the benefits of continued resistance are extremely high such as when the dispute concerns territory. In launching a destabilizing OCO, State A is demonstrating that their resolve in achieving the limited goal that is at the root of the exchange is so high that very little could be worth it for State B to continue to resist. Importantly, without experiencing significant loss of life, civilians will begin to realize this too, pressuring their government to cede and avoiding the backlash effect seen when conventional means are used for strategies of punishment.

Much like nuclear warfare, a destabilizing OCO would cause state-wide damage at a massive scale, but unlike nuclear warfare, there would not be a massive loss of civilian lives. That is not to say that no civilians would die, as medical and emergency services would likely be rendered temporarily ineffective, and water filtration systems may be damaged, but this would be an indirect effect of the attacker's actions. Furthermore, unlike in the case of strategic bombing, it is likely that the civilian outrage will be aimed at the target government, rather than the attacking

state due to the fact that they failed to effectively protect critical infrastructure and the power grid. Take for example the Colonial Pipeline hack in 2021, in which a group with Russian origins targeted the billing system of the largest oil pipeline in the US, causing mass fuel shortages, and increasing the price of gas to the highest it had been in six years. The outrage of many citizens and politicians was aimed at the Biden Administration, and oil companies rather than the Russian government which had been harboring the attackers.[39] While this is not a direct comparison to the attack I have described up until this point, it allows us to see how protection of critical infrastructure is seen as a responsibility of the government, and when citizens are affected by government failures in protecting such infrastructure they often blame the government even if a foreign actor was the cause. Importantly, unlike in the case of infrastructure destroyed in a strategic bombing campaign, the citizens have valid and rational reasons to believe the state could have effectively protected such infrastructure. The lack of physical proof of destruction will only decrease the likelihood of a backlash effect commonly seen when punishment is employed.

A valuable point of comparison here is the effectiveness of sanctions or economic means of punishment. There has been significant literature focused on the ineffectiveness of sanctions or strategies such as targeting crucial nodes of production during times of war. It is fair to argue that modern states have complex economies which allow them to be resilient despite such trying economic times. They could rely more heavily on trading partners, or rebuild the industry that had been targeted. But if cyber weapons are used to target a state's economies it may be the case that economic punishment could become far more effective. Depending on the duration of the

[39] Lemon, Jason. "Graham Calls out Biden on Colonial Pipeline Hack Response: 'Weak.'" Newsweek, May 16, 2021.

attacks and the duration that the damage remains, it may be highly difficult to coordinate imports and exports easily. Ports have long been targets of cyber-attacks and the speed of imports and exports have been heavily affected, in a situation where such an attack is being launched by a state actor who will not be appeased by payouts and is using more complex technology, this could be damaging for long periods of time.[40]

I have previously stated that multiple layers of civilian infrastructure will be targeted in a destabilizing OCO, which will create an unparalleled amount of economic damage. For example, the attacks on 9/11 reduced American GDP growth by around .5%, even though there were only a few industries that the attacks had a direct effect on namely air travel, tourism, and the insurance industry.[41] A destabilizing OCO could have direct effects on many industries such as shipping, air travel, energy, and tourism while having an indirect effect on many others. In the case of Stuxnet, it took six months for an effective fix be developed, and this was with significant help from private cybersecurity companies who needed to respond because the bug spread and affected systems worldwide. It would be reasonable to assume that an attack on the scale of a destabilizing OCO would take significantly longer to mitigate, as there would likely be multiple methods and technologies used, it would not just be one bug.

Again questions regarding backlash against the attacker who is inducing such severe economic conditions on civilians remain. But, civilian reaction to sanctions is a mixed bag where in some cases nationalism remains high and surrender is unlikely. There are however cases such as Iran where we have seen significant civilian protests in response to the hyperinflation and high

[40] Alejandro Mayorkas, "Remarks at the Maritime and Control Systems Cybersecurity Conference" (speech Ft. Lauderdale, Fl, March 21, 2022), Hack The Port 22.

[41] Roberts, Bryan W. "The Macroeconomic Impacts of the 9/11 attack: Evidence From Real-time Forecasting." Peace Economics, Peace Science and Public Policy 15, no. 2 (2009): 341-367.

unemployment caused by U.S. sanctions.[42] This ultimately boils down to the value a state places on the benefits of resistance (B). While the attacker cannot alter B, it can demonstrate that its resolve is so high that the target state will not ever see the benefits of resistance. In the case of Iran, the U.S. has rarely indicated an end to sanctions since it left the JCPOA, this has demonstrated such high resolve that civilians have questioned whether it is worth the economic damage just to possess nuclear weapons. In other cases such as Cuba, we see the benefits of resistance (B) being so high that no amount of American resolve will cause Cuba to surrender to American demands. This is because the U.S. has essentially been trying to convince Cuba to cede its political sovereignty and depose the Communist government which has ruled for 64 years with significant popular support. Coercion will never succeed when political or territorial sovereignty is at hand.

2.4: Cyber Pearl Harbor vs Cyber Hiroshima

From the outset of this thesis I have argued that the major threat of inter-state cyber coercion is not a "Cyber Pearl Harbor" but rather a "Cyber Hiroshima". While this is not the perfect metaphor, and may not even seem dissimilar from what Secretary Panetta and others have warned of, I have chosen this name for two major reasons: the simultaneous employment of multiple strategies of coercion, and the attempt at minimizing the loss of civilian lives. Both of these factors were seen in the bombing of Hiroshima. Hiroshima was chosen as a target for a few major reasons, its role as an industrial center as well as its importance to the Japanese military. The Second General Army consisting of 400,000 men responsible for the defense of Southern Japan was headquartered in the city, in addition to this multiple mobile divisions, anti-aircraft

[42] Yee, Vivian, and Farnaz Fassihi. "'Out-of-Reach Dreams' in a Sickly Economy Provoke the Rage in Iran." The New York Times, October 2, 2022.

divisions, and battalions were stationed there as well.[43] Lastly, the city was crucial for communication, as previously mentioned the Second General Army headquarters were located within the city, which was responsible for communication to and from the troops who were stationed correctly where an American invasion was to take place.[44] In addition to this, the city served as an assembly order to incoming troops should an invasion have occurred.[45]

Here we see an incredible reflection of what a destabilizing OCO would look like in the situation if one was employed. As I have argued, a destabilizing OCO sees all four major strategies of coercion be levied at once in a single attack, the bombing of Hiroshima saw just this. Denial was seen through targeting and destroying industry responsible for the war effort. Punishment was seen in the high civilian losses and infrastructural damage. Manipulation of risk was seen as the U.S. had hoped that Japan would surrender after the first attack due to fears based on the risk that a second bomb would be dropped. Lastly, leadership decapitation was seen as the command of the general army responsible for defense against a possible invasion was cut off from communicating with the fielded units. While deterrence theory is reliant on just the punishment a state would incur if they used nuclear weapons. In a time before mutual deterrence was an option, Hiroshima was a coercive success, because of the use of every major strategy across the board. This is crucial in understanding the future landscape of coercion which will utilize emerging technologies few states may have. Merely relying on punishment alone will never see success as states will be willing to incur high costs in many situations and they do not know the full strength of your newly realized cyber arsenal, which is not necessarily advantageous to the attacker. A state must attempt to alter multiple factors of the formula laid out

[43] Giangreco, Dennis M. *Hell to Pay: Operation Downfall and the Invasion of Japan, 1945–1947*. Naval Institute Press, 2017.
[44] Jones, Vincent C. *Manhattan, the Army and the Atomic Bomb*. Vol. 8. US Government Printing Office, 1985.
[45] Ibid

by Pape, in order to leave their target feeling as if they have no option but to surrender. By paralyzing the target's military, leaving them in fear of future attacks, and cutting off lines of communication between the leadership and citizenry, few options are left for the target except surrender.

I have argued, a state would be incentivized to launch a destabilizing OCO as a form of last-resort coercion in part due to the minimal amount of civilian lives within the target state that would be lost. While the decision to use a nuclear weapon obviously would lead to high civilian losses, there was an element of minimizing civilian losses at hand. The American bombing campaign on the Japanese homeland incurred a massive loss of life for civilians. Most estimates claim that the firebombing of Tokyo on March 9th, 1945 was the deadliest bombing in all of World War 2, even deadlier than both Hiroshima and Nagasaki. The entire bombing campaign against Japan claimed the lives of 300,000 to 400,000 people, about double to triple the lives lost in Hiroshima and Nagasaki combined. If an invasion of the Japanese mainland had taken place, the civilian lives lost would have been extremely high as bombing campaigns would have continued and civilians would have attempted to resist the invading American troops.

Pearl Harbor on the other hand reflects few of the coercive aspects that led to the decisions to launch the attack that we would see in the use of a destabilizing OCO. Pearl Harbor was a rational preventive strike purely rooted in the strategy of denial.[46] Japan was attempting to coerce the United States into lifting the embargo on Japanese oil by altering the probability of benefits by crippling America's Pacific fleet. Loss of civilian life and damage to civilian infrastructure, as well as the killing of leaders such as admirals, was of no concern in the planning of the attack. In addition to this there is little to show that they were attempting to

[46] Mearsheimer, John J. and Sebastian Rosato, How States Think: The Rationality of Foreign Policy (forthcoming from Yale University Press)

manipulate risk, while there was some concern that there would be Japanese air raids on western cities, the goal of the attack was not to inspire such fear. Lastly, there was no form of calculus concerning civilian lives, while some civilians did die in the attack, they were not the target or a factor. Simply put, the attack was not meant to impact civilians, it was purely aimed at minimizing America's ability to act in East Asia. This may seem minimal, but a destabilizing OCO would have widespread effects on civilians and civilian lives with that being a fundamental goal of the operation. Lastly, as I argue, unlike the attacks on Pearl Harbor, a destabilizing OCO would not be a surprise in terms of occurrence but rather a surprise in terms of the strength and nature of the attack. This again is more like Hiroshima, where the Japanese knew an American attack was incoming but they were surprised that it was a nuclear attack.

2.5: Neo-Realism, Rationality, and Cyber Coercion

Previously, I mentioned that this was a neorealist theory of coercion, and uses a unique definition of rationality proposed by John Mearsheimer and Sebastian Rosato. This is a necessary aspect of this theory, as the three-stage process at the center of this theory is based on the idea that the actors involved in the exchange are acting rationally. Furthermore, they are acting with the primary concern being ensuring and maximizing their survival. Additionally, it explains why, even if coercion fails, or even if State B fails to correctly attribute who launched the initial attack, how participating in the process can and likely will be rational. Remember, there are two sole criteria that must be fulfilled, a state must be acting based on a credible theory, and the decision must be made through a deliberative process.[47] Using Stuxnet as a point of comparison we can understand the first criteria clearly. The United States believed that its survival would be threatened if Iran gained nuclear weapons, they also believed allowing Israel to attack an Iranian

[47] Ibid

power plant could lead to a conflict which the U.S. would be dragged into, which could also threaten the survival of the U.S. This was credible because, despite the logic of mutually assured destruction, Iranian rhetoric had been so anti-American that such an occurrence could not be ruled out. In addition to this, the nuclear program had become a point of pride for the country, and if Israel, a bitter rival, destroyed the program, conflict could easily have emerged.

If this was the theory being operated on, then they fulfilled the first criterion, and while we do not have the information to know if the second was fulfilled it is likely that multiple members of the Obama administration deliberated on the merits of launching the attacks, and Obama himself was the ultimate decider. Given the amount of deliberation seen throughout the process of writing and agreeing on the JCPOA, the fact that it was a joint Israeli-US operation and Israel had initially wanted to launch a conventional strike, and the use of new technology, one would assume that there were significant discussions across both the Bush and Obama administrations.

The focal point being survival should demonstrate why this process may rarely be seen, the initial limited goal with which the first OCO is launched must be an attempt by State A to ensure its survival, meaning State B is currently involved in an endeavor or is proposing a policy which threatens the survival of State A. Furthermore, State A must have sufficient strength, wealth, and technological advancement to have sufficient cyber weaponry which would allow them to achieve their limited goal without the need to threaten or use conventional weaponry. This is because in launching the initial OCO, they know that should State B call their bluff and launch a retaliatory attack, they will need to launch a destabilizing OCO. Few states likely have such ability, and those states are the Great Powers: The United States, China, and Russia. Furthermore, these three states specifically avoid threatening the survival of the other two in

order to avoid such a dangerous situation from developing. Because of this, we have seen comparatively few OCOs launched by a great power against another great power in an attempt to coerce the other. Rather, we have seen attempts at causing economic damage to one another, or more commonly espionage. These do not threaten the survival of the target state and are not launched because State A feels as if its survival is threatened, rather they likely are attempts to demonstrate resolve, or in the case of espionage in order to increase the likelihood that a Great Power can continue to be a Great Power. Attacks such as these are a non-factor in this theory, they are simply not attempts at coercion. For the process laid out in Figure 1 to emerge, State A's survival must be threatened.

Up to this point, I have laid out how this theory follows two of the key tenets of neorealism: survival is the primary goal of states, and states are rational actors. But there is one more neorealist assumption heavily at play within this theory, states can never be certain about the intentions of other states. This plays a crucial role in understanding why the retaliatory response in stage 2 is necessary. The target state can never assume that the initial attack was the end goal of the state. They do not know if the goal achieved through the initial attack was the only goal or a part of a larger multifaceted campaign of coercion. Furthermore, they cannot know if the attacker is even able to correctly gauge how effective the initial attack. Depending on the target chosen, it may take a long time for a state to even be able to gauge the tactical effectiveness of an OCO. In addition to this, the attribution difficulties are only exacerbated because of this uncertainty of intentions, any number of adversaries can be responsible, or perhaps even states that one is friendly with. This is why even if a retaliatory attack is launched against a state other than the attacker, it may be rational. If they have a credible theory backing

their reasoning for making an assumption about the identity of the attacker, even if it is wrong, it can still be rational.

2.6: Implications

While cyber defense should be key in future military development, there is little that can be done to effectively protect ourselves from cyber-attacks. The first stage seen in Figure 1 is impossible to protect against, states will attempt to use coercion in order to achieve limited goals, and OCOs provide an effective means of doing just that. Furthermore, such operations are effective at demonstrating resolve through maintenance of the technological edge causing a state to fear the possibility that the scenario could reach stage 3, increasing the odds of coercive success early into the process. Again, this was seen after the Stuxnet attacks where despite the lack of tactical effectiveness the strategic goal was eventually realized due to the benefits of displaying resolve. Because of this, states should continue the development of such lower-level cyber weaponry, while also retaining the capabilities to pull off a destabilizing OCO if need be.

In addition to this, states should expect to face a stage 1 OCO and possess cyber weaponry that would demonstrate high technological capability to use in stage 2 of the process seen in Figure 1. This would force the attacker state to have to reckon with the possibility that should they decide to go to stage 3 and launch a destabilizing OCO, then they might be the target of such an attack in response. Much like with nuclear deterrence where a state in making the decision to initiate a nuclear exchange is making the decision to face a nuclear attack, this scenario incentivizes the target state to force the attacker into having to make such a decision. If this could be done then the process will never reach stage 3 and diplomacy could be used to resolve the situation. Essentially states should maintain a diverse cyber arsenal allowing them to respond to cyber-attacks with reciprocal attacks similar in nature and effectiveness.

But this is an imperfect solution as the attribution issues may cause the target state to respond incorrectly and launch a reciprocal attack on a state different from the state which attacked them in the first place, proving a technological gap exists, and causing the attacker to believe that there is little threat of a retaliatory destabilizing OCO if they launch such an attack. In order to overcome this issue, significant resources should be placed into overcoming the attribution difficulties inherent to cyber warfare. Baliga, Bueno de Mesquita, and Wolitzky, have echoed this using game theory to argue that it is possible to overcome the attribution difficulties in cyberspace and this is necessary to improve deterrence in cyberspace.[48]

One might argue that if a state has such strong offensive capability in cyberspace they will first use a destabilizing OCO in order to attempt to reach the limited goal they are seeking out. But states will always begin with a more minor OCO because of the "use-and-lose" quality of cyber weaponry. If a state uses the strongest weapon in its cyber arsenal it can only use it one time, and these weapons take years to build, and significant money and resources. After the weapon is used, states and private cybersecurity companies gain an understanding of how to protect against such an attack rendering it ineffective in the future. Furthermore, the difficulty in attacker attribution seen in cyberwarfare allows for an attacker to believe it can launch an attack with a limited goal without having to face a retaliatory response. Blame could be given to non-state actors or another state altogether, making a reciprocal conventional attack risky for the victim of a cyber-attack. This is why the retaliatory response will likely be a cyber-attack, it is far less risky and in case the target is chosen incorrectly the attacker has the benefits of attributional uncertainty. Yet, by the third stage, when a destabilizing cyber-attack is launched the defender will likely have deduced who the attacker is, making the attack less a surprise in

[48] Baliga, Sandeep, Ethan Bueno De Mesquita, and Alexander Wolitzky. "Deterrence with imperfect attribution." American Political Science Review 114, no. 4 (2020): 1155-1178.

terms of occurrence and more a surprise in terms of the nature of the attack. The attributional difficulties may have been overcome because of successful identification in stage 2, or simply deducing which adversaries would have an incentive to launch an initial attack with such an aim. In addition to this, it is possible that states, even those who have an adversarial relationship with the target state and a friendly relationship with the attacking state, may attempt to clear their name in order to avoid facing the reciprocal attack in stage 2.

Outside of the technological realm, there is a major improvement every state can make to improve their defense against facing a destabilizing OCO, improving doctrinal clarity. Few if any states have clear doctrine guiding their response to OCOs. While the United States has claimed such an attack would be viewed as an act of war, there is little to make an attacker believe the U.S. would respond as such. Rather, there have been a variety of responses from decision-makers concerning what a response would look like. Some have claimed that a reciprocal OCO would be launched mirroring what I have argued. While others have claimed that a conventional attack would be launched, conflicting with what I have argued, but also seemingly ignoring the dangers of misidentification which may incentivize states to launch OCOs against the U.S. If the U.S. were to launch conventional attacks in response, it may incentivize a state to attempt a "bait-and-bleed" strategy in which they bait the U.S. into conflict with a third state through the use of an OCO, hoping that the U.S. misidentifies the attacker.[49] The safest strategy would be to have a doctrine stating a reciprocal OCO will be launched in an attempt to institute a form of cyber deterrence. Regardless of how a state wants to ensure deterrence in cyberspace, massive improvements to attacker attribution must be made.

[49] For discussion on the strategy of bait-and-bleed see, Mearsheimer, John J. *The Tragedy of Great Power Politics*. WW Norton & Company, 2001.

3: Conclusion

This thesis has outlined a new theory of coercion in cyberspace aimed at understanding the ultimate threat, a "Cyber Hiroshima". Technological advancement has shaped and will continue to shape interstate conflict, and emerging technologies will often be turned to for last-resort purposes. Coercion and emerging technology inherently go hand-in-hand as the goal of coercion is to reduce the value an opponent places on continued resistance. New technology drastically reduces these benefits because of the overwhelming amount of difficulty it takes to overcome a threat that has never been seen or experienced before. A destabilizing OCO may be thought of as a silver bullet much like the atomic bombs were, and even if this is not the case decision-makers will still opt to experiment with such an attack when the only alternative is to engage in conventional conflict.

This theory is based heavily on work done by air power strategists and theorists who sought out effective tactics to achieve successful coercion with the least amount of risk to the attacker. As I have argued, the best coercive strategy involves not just one of these strategies but use of all four major strategies: denial, punishment, risk, and leadership decapitation. Using the end of World War 2, we see how in cases of last resort not just one strategy can solely lead to a coercive victory. Inherently, by placing the attacker in a position of last resort the target state has demonstrated that it has such high resolve that the attacker will need to impact the value of resistance heavily. As Robert Pape has demonstrated, each strategy only affects one factor in his equation which determines the value of resistance. But when multiple factors are affected, the value of resistance will fall dramatically no matter how high the benefits of continued resistance may be. This is a major departure from the air power school of coercion, as air power alone may not be feasible for an all-encompassing strategy of coercion. Furthermore, air power strategists in theorizing the best strategies for coercion have attempted to find which strategy is most effective

and efficient at coercing an opponent. While an all-encompassing strategy may be effective it would likely not be efficient with air power alone, as the time, cost of planes and munitions, as well as the risk, could all be significantly higher than a destabilizing OCO.

The effect of high civilian casualties as well as physical damage is also a major factor in this theory. Civilian backlash is often key in a state's decision to continue resistance as decision-makers themselves are essentially factoring in the value they place on continued power in governance to the equation. The lack of physicality is key to the success that a destabilizing OCO may have at coercing an opponent. As we have seen in the war in Ukraine, seeing physical destruction, and knowing those who have lost their lives are powerful forces contributing to increased nationalism, which is the biggest obstacle to successful coercion. One would assume that seeing critical infrastructure physically undamaged would not evoke the same effect, as civilians will be left questioning how the government failed to protect them from such an attack. Civilians know that little can be done to protect against massive airstrikes, rockets, and other conventional munitions, but the question of whether they will understand that the same could apply to cyber weaponry remains.

I have introduced a model in this thesis which lays out the process that will lead to the use of a destabilizing OCO. This model is a three-step process which illustrates a scenario involving two actors beginning with an actor using an OCO for a limited goal and ending with that actor using a destabilizing OCO. Importantly, this model does not aim to understand how every exchange in cyberspace will go, diplomacy can and should prevent such a process from emerging, but the focus of this theory is when a state will launch a destabilizing OCO. Furthermore, sometimes an OCO launched with an independent goal will achieve the goal and it would not be worth it for State B to respond. This model is highly simplified and international

relations is filled with nuance and intervening variables such as power status, whether a state has nuclear weapons, who a state is allied with, etc. All of these would likely have an effect on whether such a process occurs, but this model does explain why states would be incentivized to go down such a dangerous path.

Engaging in topics related to emerging technologies, especially cyber weaponry is inherently difficult due to the lack of information available. The major shortcoming of this theory is the question of whether a destabilizing OCO as I have envisioned it would be possible. Heavy pushback exists regarding these questions, due to the nature of how critical infrastructure is guarded. Critical infrastructure systems are often offline, and many argue that they will continue to remain so, in order to avoid the potential dangers posed by cyber weaponry. But anecdotal information regarding plans for continued cyber-attacks by the United States against Iran has hinted that states such as the U.S. may be able to attack systems. Furthermore, tactical failures of Stuxnet have been leveled against claims of the likelihood that we will see the use of such technologies as a staple of future warfighting. Two points are often repeated: Stuxnet failed to destroy the majority of the centrifuges targeted, and Stuxnet was fixed within 6 months and the centrifuges were repaired quickly. These arguments fail to address questions regarding the effect that Stuxnet had on Iran eventually agreeing to the JCPOA and overlook how effective a lengthy shutdown of industry or critical infrastructure may be at coercing an opponent. Furthermore, Stuxnet was dealt with relatively quickly for one reason, it spread to the open web necessitating a response from private cyber-security companies. One would assume, states have learned from this and will attempt to ensure that cyber weaponry cannot spread to the open internet to avoid such a scenario in the future.

Cyberwarfare is just one of many possible components of militaries of the future and we should continue to attempt to theorize the impact that emerging technologies have. AI, space-based weapons, and anti-satellite weapons all may play an extremely active role in the coercive arsenal of states in the near future. Through theorizing how and why states may use such technologies we can gain valuable information regarding how to best avoid worst-case scenarios from emerging. Despite progress made in nuclear de-proliferation since the fall of the Soviet Union, we are now closer to doomsday than ever. As discussions regarding topics such as cross-domain deterrence continue its imperative to understand the possible effects emerging technologies could have on the prospect of nuclear war.[50] As I mentioned previously, the scenario envisioned in this thesis is less likely to be initiated by a non-nuclear state against a nuclear state, but questions remain regarding the likelihood of such a scenario occurring between nuclear powers. Work on cross-domain deterrence should continue to be done in an attempt to evaluate the potential such a scenario would occur and whether a state would launch a nuclear response if the attacker is also a nuclear power. Furthermore, work concerning the strength of defense doctrine that establishes responses to OCOs should continue to be done, investigating the effectiveness of doctrine on the reduction of attacks and resulting increased diplomacy. In addition to this work aimed at reducing the attribution difficulties inherent to cyberspace are inherent in mitigating the incentives of bad actors to operate through the domain.

While not directly engaging in the cyber offense-defense debate, this theory reveals just how difficult it is for the defense to overcome the advantage of the offense in cyberspace. While organization and skill may be just as important as the technology being used, this does not root out the possibility that a destabilizing OCO may be used. States develop cyber arsenals and

[50] Gartzke, Erik, Jon Lindsay, and Michael Nacht. "Cross-Domain Deterrence: Strategy in an Era of Complexity." In *International Studies Association Annual Meeting, Toronto*, vol. 9. 2014.

launch OCOs because they think they will succeed, and they often do. Even when a state fails it gains valuable information which may allow for future offensive success. All the offense needs is one or a few vulnerabilities to exploit and they can do massive damage to another state. Furthermore, they are incentivized to launch OCOs because of the possibility they may not even be identified as the attacker, and thus do not have to face the response, or even that another adversary is believed to be the attacker and faces a response. In addition to this, a cyber-attack may be a cheap and effective way to demonstrate resolve in an ongoing situation that conventional strategies would not be appropriate for. Demonstrating a technological advantage against an opponent is highly effective not just for the mitigation of the situation at hand but also because the state is demonstrating technological prowess system wide. This will allow for far more effective coercion in the future, as other states may be coerced merely by the threat of facing their cyber arsenal.

The theory presented in this thesis has opened the door for significant future research aimed at understanding what a destabilizing OCO would look like. Furthermore, as this theory predicts a destabilizing OCO will likely be used, research should seek out the best possible solutions for the protection of critical infrastructure from such an attack. As I have argued, defense is extremely difficult, but research aimed at efficient recovery from such an attack is crucial. Defense alone is likely impossible, but the response of the target state is critical to the effectiveness of such an attack. Alternative lines of communication that are not bound by internet infrastructure should be researched, as communication between government and civilians will be critical in the wake of a destabilizing OCO. The economic side of the situation should also generate compelling research, questions such as: How much economic damage would a state endure? How long will it take for the target's economy to stabilize? Which industries will most

likely be targeted? Are all crucial to understanding how successful coercion through a destabilizing OCO will likely be. Lastly, more attention to the doctrinal response of the target state should be given. In cases where a state is bound by doctrine to respond to a cyber-attack with a conventional attack, questions emerge concerning whether the path to a destabilizing OCO is slowed down or sped up.

Works Cited

AFCEA. The Evolution of US Cyber Power. https://www.afcea.org/committees/cyber/documents/TheEvolutionofUSCyberpower.pdf.

Alejandro Mayorkas, "Remarks at the Maritime and Control Systems Cybersecurity Conference" (speech Ft. Lauderdale, Fl, March 21, 2022), Hack The Port 22,

Gartzke, Erik, Jon Lindsay, and Michael Nacht. "*Cross-Domain Deterrence: Strategy in an Era of Complexity*." In International Studies Association Annual Meeting, Toronto, vol. 9. 2014.

Giangreco, Dennis M. Hell to Pay: *Operation Downfall and the Invasion of Japan, 1945–1947*. Naval Institute Press, 2017.

Goldman, Emily O., and Michael Warner. "Why a Digital Pearl Harbor Makes Sense... and Is Possible." *George Perkovich and Ariel E. Levite Understanding Cyber Conflict* 14 (2017): 147-157.

Hedgecock, Kathryn, and Lauren Sukin. "Responding to Uncertainty: The Importance of Covertness in Support for Retaliation to Cyber and Kinetic Attacks." *Journal of Conflict Resolution* (2022)

Jones, Vincent C. *Manhattan, the Army and the Atomic Bomb*. Vol. 8. US Government Printing Office, 1985.

Kallberg, Jan. "*Bye, Cyber Pearl Harbor*." (2021).

Kugler, Richard L. "Deterrence of Cyber Attacks." *Cyberpower and National Security* 320 (2009): 309-340.

Lawson, Sean, and Michael K. Middleton. "Cyber Pearl Harbor: Analogy, Fear, and the Framing of Cyber Security threats in the United States, 1991-2016." *First Monday* (2019).

Leal, Marcelo M., and Paul Musgrave. "Hitting Back or Holding Back in Cyberspace: Experimental Evidence Regarding Americans' Responses to Cyberattacks." *Conflict Management and Peace Science* 40, no. 1 (2023): 42-64.

Lemon, Jason. "*Graham Calls out Biden on Colonial Pipeline Hack Response: 'Weak*.'" Newsweek, May 16, 2021.

Leon E. Panetta, "Defending the Nation from Cyber Attack" (speech, New York, New York, October 11, 2012) Business Executives for National Security

Levy, Jack S. "Case studies: Types, designs, and logics of inference." Conflict management and peace science 25, no. 1 (2008): 1-18.

Libicki, Martin C. *Cyberdeterrence and Cyberwar*. RAND Corporation, 2009.

Lieberthal, Kenneth, and Peter Warren Singer. *Cybersecurity and US-China relations*. Brookings, 2012.

Magee, Clifford S. "Awaiting Cyber 9/11." *Joint Force Quarterly* 70 (2013)

Mearsheimer, John J. *The Tragedy of Great Power Politics*. WW Norton & Company, 2001.

Mearsheimer, John J. and Sebastian Rosato, *How States Think: The Rationality of Foreign Policy* (forthcoming from Yale University Press)

Nye, Joseph S. Cyber power. Cambridge: Harvard Kennedy School, Belfer Center for Science and International Affairs, 2010.

Nye, Joseph S. "*The Kremlin and the US Election*." Project Syndicate, August 30, 2017.

Pape, Robert A. *Bombing to Win: Air power and Coercion in War*. Cornell University Press, 1996.

Rattray, Gregory J. *Strategic Warfare in Cyberspace*. MIT press, 2001.

Rid, Thomas. "Cyber War Will Not Take Place." Journal of Strategic Studies 35, no. 1 (2012): 5-32.

Roberts, Bryan W. "The Macroeconomic Impacts of the 9/11 Attack: Evidence From Real-time Forecasting." *Peace Economics, Peace Science and Public Policy* 15, no. 2 (2009): 341-367.

Sanger, David E., and Mark Mazzetti. "*U.S. Had Cyberattack Plan If Iran Nuclear Dispute Led to Conflict*." The New York Times, February 16, 2016.

Schelling, Thomas C. *Arms and Influence*. Yale University Press, 2020.

Slayton, Rebecca. "What Is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment." *International Security* 41, no. 3 (2017): 72–109.

Smith, George. "An Electronic Pearl Harbor? Not Likely." *Issues in Science and Technology* 15, no. 1 (1998): 68-73.

Warden, Col. "John A. III, USAF. "The Enemy as a System."." *Airpower Journal* (1995).

Wirtz, James J. "The Cyber Pearl Harbor Redux: Helpful Analogy or Cyber Hype?." *Intelligence and National Security* 33, no. 5 (2018): 771-773.

Wirtz, James J. "The Cyber Pearl Harbor." *Intelligence and National Security* 32, no. 6 (2017): 758-767.

Yee, Vivian, and Farnaz Fassihi. "'*Out-of-Reach Dreams' in a Sickly Economy Provoke the Rage in Iran*." The New York Times, October 2, 2022.