Emerging Anti-satellite Capabilities and the Recourse to Deterrence Theory in Practice

Emerging Anti-satellite Capabilities and the Recourse to Deterrence Theory in Practice

Kaili Ayers

A Thesis in the Field of International Relations

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University
May 2023

Abstract

The expansion of space-based capabilities has ballooned over the past six decades in response to the rapid growth of technological advancements and simultaneous lowering of costs for their production. Satellites are now being relied upon to provide support for essential military operations in addition to global society's daily activities. However, the largely inferential nature of satellite capabilities suggest that deterrence as a theoretical and strategic construct may not effectively function in the outer space environment. Juxtaposed against nuclear deterrence which relies on the physical exhibition of military power, satellite capabilities such as non-kinetic physical, electronic, and cyber are not readily visible, thus diminishing the levels of perceptibility and signaling strength necessary for successful deterrence.

The purpose of this thesis is twofold: 1) to provide a theoretical analysis of deterrence in outer space, and 2) to empirically test the utility of inferential anti-satellite capabilities. The first half will examine literature on the theoretical revision of deterrence, focusing on the principle of credibility. The second half will test the efficacy of the use of inferential deterrence strategies by applying it to a revised expected utility model. The investigation of both theoretical and empirical dimensions of deterrence theory can provide the groundwork for expanded research in the field; especially in a continually evolving international system.

Acknowledgments

Writing this thesis was an endeavor I had long been equally wary and exited to embark on, yet more rewarding than I could have ever imagined. None of this would have been possible without the generous support and guidance from a number of people who have helped make this research possible, and of whom I am immeasurably appreciative.

I am eternally grateful for Dr. Nicholas Gvosdev, Professor of National Security Affairs at the Naval War College and lecturer at the Harvard Extension School, who was kind enough to take me on as a thesis candidate, and who devoted much of his time towards directing this research. I learned immensely from his course: Evolution of Deterrence Theory, where he encouraged me to delve deeper into the implications of nuclear deterrence, and which prompted various, decisive points of analysis in this work.

I also feel very fortunate to have been advised and taught by Dr. Doug Bond, Lecturer and Research Advisor at the Harvard Extension School's Master of Liberal Arts (ALM) Program. Under his careful guidance, I was able to produce a comprehensive thesis proposal which allowed me to further refine my ideas as I progressed through the thesis program.

To the staff of the ALM Program, thank you for all of your guidance; particularly Trudi Goldberg Pires, Executive Assistant to the Dean of Academic Programs, and Gail Dourian, Thesis Coordinator for the Office of ALM Advising and Program Administration, thank you for providing me with much needed direction on how to

navigate both the thesis program and the horrors of Microsoft Word formatting difficulties.

Lastly, I would like to thank my friends and family for all their kind words of support. Particularly that of my mother and significant other, who also endured my insufferable requests to assure me of my semi-colon placement, as well as encouraged me to pursue my dreams.

Table of Contents

List of Tables

List of Figures

Chapter I.

Why Satellites Matter

Earthbound societies are happily ensconced in digital living and remain dependent on an increasing inventory of space technologies and services to conduct even its most basic functions. Such technological capabilities span a diverse range of sectors including communications, crisis management, national security, and transport. GPS satellites for example, could be applied to secure financial transactions in order to keep data parcels in sync across the internet, as well as determine whether a transaction aligns with the location of the appropriate individual. Without these types of space-based capabilities, Internet speeds could come to a halt, access to the cloud could degrade, and systems that control energy grids, water, and transportation could begin to fail (Black, 2018, para. 4). Satellites are also necessary components of verification regimes by nuclear-armed states. If a situation arises where a state can no longer adequately verify the status of an opponent; whether due to malfunctions in a surveillance satellite or other impairment, it could be pressured to escalate tensions by launching a preemptive attack (Black, 2018, para. 6). Satellites have thus become so profoundly integrated into our societies that even a slight disruption could have far-reaching consequences that extend beyond our national borders.

This chapter examines why and in what way satellites matter to the discussion of deterrence theory in space by looking at some of the primary drivers behind their proliferation, some criticisms of satellites, as well as the various types and respective purposes of satellites. This chapter concludes with satellite vulnerabilities before transitioning to a discussion on deterrence theory.

Enabling Factors

The heavy dependence on, and mass production of satellites has been driven primarily by two enabling factors: 1) technological advancement and 2) the commercial space industry, which includes space-based economic drivers and large-scale financial investments. Firstly, new technologies and resultant capabilities are driving the research and development of satellites. Some terrestrially-based economic drivers, which include autonomous vehicles, precision agriculture, expanded use of the Internet, growth of the global middle class, and infrastructure and environmental management, are considered to be inherently disruptive (Eftimiades, 2022, p. 5). This is in part because they present challenges to global governance in terms of defense and security. Autonomous vehicles for instance, could be used to increase security in ungoverned or contested territories, but at the same time be exploited by revisionist states or non-state actors in order to upend the status quo (Hughes, 2016, p. 42). These technologies necessitate satellite proliferation as their operation is greater enhanced and fueled by commercial demand.

Secondly, space-based economic drivers are the decreasing cost of launch, the mass production of satellites (i.e. a result of miniaturization technology), on-orbit processing power, and artificial intelligence. These drivers are significantly changing the cost-benefit analysis for commercial and government to employ satellites. Production time in particular has been reduced such that most technologically advanced, commercial satellites have the capacity to launch into orbit every 3-4 months (Eftimiades, 2022, p. 7). Comparatively, it takes a government-sponsored satellite program on average 7.5 years to develop and launch a first vehicle, and just over 3 years for a subsequent vehicle (Davis & Filip, 2015, p. i). The considerable difference in timelines between commercial and government programs can be attributed partly to the degree of technology development at the beginning stages of the program. While the former performs technology development prior to making a product available to customers, the later incorporates it into its overall process (Davis & Filip, 2015, p. i). However, considering that comparable aspects of both programs are similar (e.g.. start of development, assembly to launch, and overall program duration), the government potentially has the capacity for exponential growth similar to what commercial programs are experiencing currently, if the government satellite acquisition process is streamlined (Davis & Filip, 2015, p. 1).

Large Scale Investments

According to Merrill Lynch's 2017 report, the space market is expected to grow from $339 billion in 2016 to $2.7 trillion USD by 2045. Morgan Stanley's projections offer a similar outlook, estimating that upwards of $1 trillion in revenue will be generated from the space industry by 2040, with satellite broadband estimated to represent 50%-70% of total projected growth, see Figure 1 (Space, 2020, para. 10). Focusing on broadband will help to drive down the cost of data as the demand for it increases exponentially. Moreover, demand from services such as autonomous vehicles, the Internet of Things, AI, and virtual reality will further necessitate the expansion of satellites with broadband capabilities. Thus, high-capacity data transmission capabilities will prove to be an important driving factor in the growing commercial space industry presently, and going forward.

Figure 1: The Global Space Economy

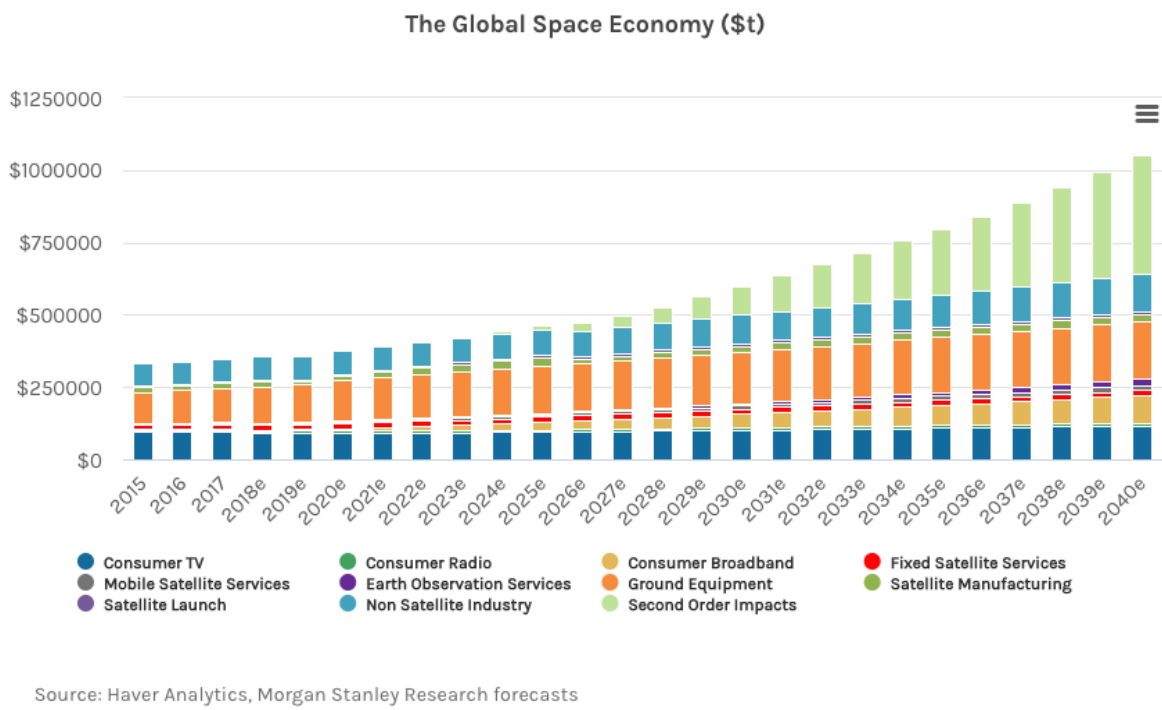

Figure 1illustrates the revenue of the global space economy by sector of investment (Space, 2020).

In 2022, Bryce Tech's Start-Up Space report identified three important overarching investment trends: 1) the continued large-scale investment in space ventures with an increasing number of investors funding recipients with larger than average deal sizes, 2) the emergence of public markets as a major source of funding for companies across all space categories (i.e. launch and in-space services), and 3) 2021's record number of mergers and acquisitions (M&A) for start-up space companies (BryceTech, 2022, p. 2). The last trend suggests that companies are pursuing acquisitions in order to vertically integrate or improve short-term revenues; processes which require streamlining operations in order to strengthen supply chains, reduce production costs, and accessing new distribution channels (Tarver, 2022, para.6).[1] Indeed, the pace and financial volume of the space industry has also encouraged an environment of zealous innovation. Space start-ups alone have attracted over $15 billion in total financing in 2021, breaking the 2020 record of $7.7 billion (BryceTech, 2022, p. 2). Additionally, the launching cost of a kilogram of mass to LEO (low earth orbit) has decreased by a whopping 90% (Poponak & Siewert, 2017, 03:42). Part of this decrease can be attributed to SpaceX's development of an integrated vehicle production line which enabled the company to lower costs to about a third of their competitor's. Furthermore given these lower costs, the barriers to entry for the space industry as a whole has dramatically fallen (Matthews, 2018, para. 2). In effect, space is no longer limited to the largest multinational companies, and its 'democratization' is working to significantly alter the commercial landscape. The entrance of smaller companies such as Rocket Lab and Vector Launch serve to further drive down the overall cost by saturating the market. Take for example, United Launch Alliance's decision in 2017 to drop the price of its Atlas 5 rocket by roughly one-third in response to rising competition from SpaceX and others. Such an action represents a clear display of the effects of competitive pricing (Matthews, 2018, para. 5). With large-scale investment as a powerful enabling-factor of satellite proliferation, it comes as no surprise that outer space itself is becoming dangerously crowded.

[1] Vertical integration involves streamlining operations in order to strengthen supply chains, reduce production costs, gain upstream profits, or access to new distribution channels. It occurs when one company acquires a related company within the same industry; this could include a producer, vendor, supplier, or distributor (Tarver, 2022, para. 6).

Satellites by the Numbers

The number of satellites currently in orbit supports the projected trends in investment. The Union of Concerned Scientists (UCS) which maintains an open-source record of operational satellites, reported that as of Jan. 1st, 2022, there are an estimated 4,852 satellites currently orbiting Earth; 2,944 of which belong to the United States (*UCS Satellite Database*, 2022). Furthermore, SpaceX owns 1,655 of those satellites, accounting for more than half of the total U.S. satellite volume (*Every Satellite Orbiting Earth and Who Owns Them*, 2022). In 1976 the United Nations Index of Objects Launched into Outer Space went into force, which was the product of Resolution 3235 (XXIX) Convention on Registration of Objects Launched into Outer Space. It has thus far recorded 12,874 objects[2] in space, 6,198 of which are registered under the United States. In fact, between the years 2019-2020, the number of launched spacecraft jumped from approximately 500 to 1,300, and again between 2020-2021 from 1,300 to just over 1,800, see Figure 2 (*Smallsats by the Numbers 2022*, 2022, slide 6). Such a quick rise points to the larger trend that small satellites, also referred to as smallsats, are gaining momentum as the preferred vehicle to meet the commercial and security demands of various governments and commercial enterprises.

[2] The discrepancy between UCS and UN data lies in the definition of 'space object' as outlined in Article I(b) of Resolution 3235 (XXIX). A 'space object' it states, "…includes component parts of a space object as well as its launch vehicle and parts thereof…", (United Nations, 1976). UCS in contrast, only records operational satellites in orbit excluding launch vehicle and miscellaneous parts.

Figure 2: Spacecraft Launched 2012-2021, by Mass Class

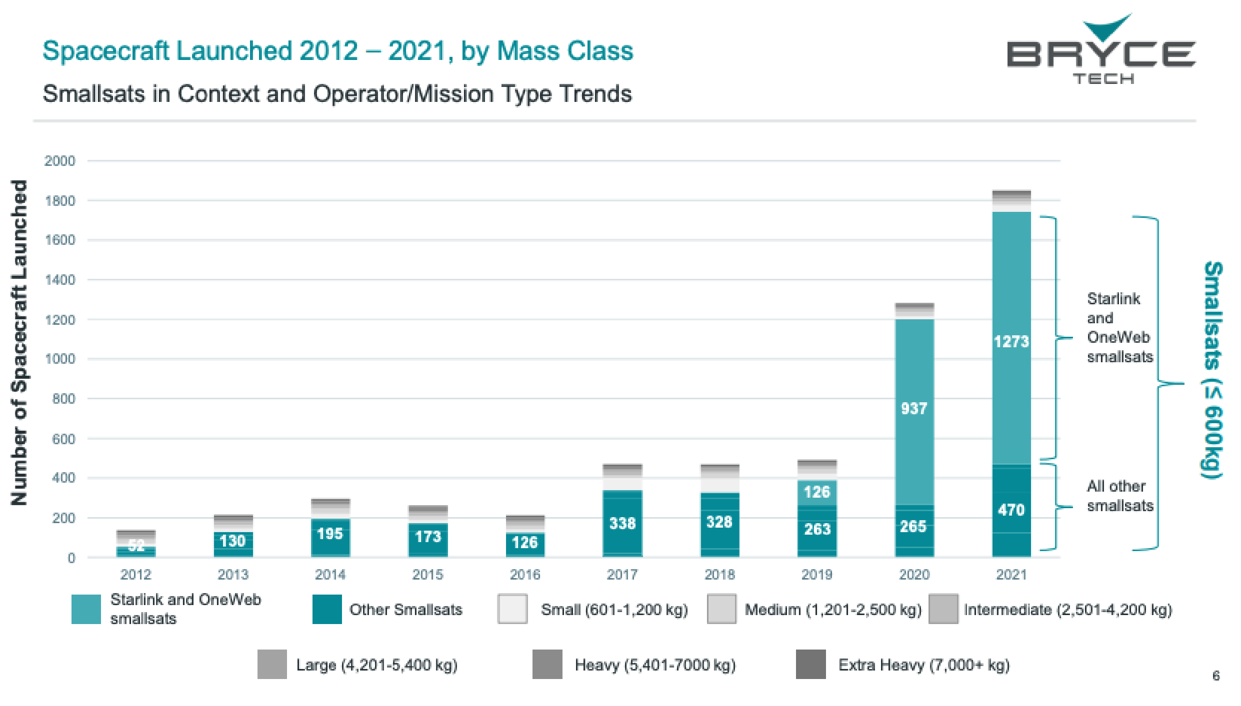

Figure 2 demonstrates the sharp increase in smallsat launches from 2020 onwards (BryceTech, 2021).

Criticisms

While outer space pundits have enthusiastically welcomed the growth of satellite programs, astronomers view them as a source of alarm. This is because the expansion of artificial objects in space obstructs many important astronomical observations around the globe. More importantly, interference could inhibit not only routine scientific endeavors, but lead to possible missed scientific discoveries (Witze, 2022, para. 4). Companies such as SpaceX have attempted to mitigate such problems by testing methods of dimming

Starlinks[3], but how effective these solutions are have yet to be determined. Sunshades for example, reduce the brightness emitted by satellites, and other materials that could adhere to a satellite such as stickers could reduce reflectivity towards Earth. However, sunshades appear to hamper Starlink communications signals and have consequently been left out of newer Starlink satellites (Witze, 2022, para. 12-13). Other negative impacts include affecting wildlife that relies on celestial navigation, as well as inhibiting human knowledge systems (e.g. Indigenous knowledge systems) that rely on information from the night sky to observe important events throughout the year (Witze, 2022, para. 10). In spite of these criticisms, the commercial space industry continues to attract large-scale investments, creating a highly innovative technology climate.

Satellite Types and Capabilities

There are four broad categories of satellites, each designed to perform a specific job and categorized according to what responsibilities it must fulfill. The first category is communications (SATCOM). They enable Internet and other communications to span remote areas without the need for ground based infrastructure. An estimated 80% of spacecraft launched in 2021 were dedicated to communications operations (*Smallsats by the Numbers 2022*, 2022, slide 2). The next category is ISR (intelligence or reconnaissance) and Earth observation satellites, which gather reams of data via remote sensing to support military intelligence activities such as the tracking and monitoring of military forces as well as the continuous observation of key events and locations.

[3] Starlink is a satellite network (an individual Starlink is thus one satellite unit) developed and launched by SpaceX in order to provide internet services at a low cost to remote locations. Also dubbed a 'mega-constellation', the company aims to have as many as 42,000 satellites in orbit (Mann & Pultarova, 2022, para. 1).

Navigation-related satellites are as its name implies, used for positioning, navigation, and timing (PNT) services by transmitting timing signals for land, air, sea, and space navigation, tracking, and precision weapons guidance (DIA, 2022, p.1-2). Lastly, satellites in the category of science technology are specialized to in gathering data on astronomical objects and phenomena including but not limited to: magnetic fields, space radiation, Earth and its atmosphere, the Sun or other stars, as well as planets and their moons (Grey, n.d., para 7).

Satellite Vulnerabilities

Considering that satellite applications are seemingly limitless and that there is yet a robust governing set of international rules governing their use, the increasing number of satellite systems will present challenges to U.S. national security. The most visible threat to satellites are collisions via debris and 'kinetic kill' strategies. There are currently more than 27,000 pieces of orbital debris that are being tracked by the Department of Defense's global Space Surveillance Network (SSN) sensors (Garcia, 2021, para. 6-7). These objects travel at speeds up to 17,500 mph, which is fast enough to damage a satellite or spacecraft. According to NASA, even tiny particles have the capacity to damage spacecraft when travelling at these velocities; several space shuttle windows needed to be replaced due to damaged caused by material that was later shown to be flecks of paint. In fact, the highest risks to spacecraft in LEO are millimeter-sized orbital debris. For context, the kinetic energy released during a 10 km/s collision with a particle of just a few grams in mass is proportionate to that of a hand-held grenade and can very well destroy a spacecraft (Lafleur, 2011, p. 1). Whereas debris are largely the result of unintentional collisions, kinetic-kill is a type of counterspace concept where missiles designed to destroy satellites (i.e. an anti-satellite weapon, abbreviated to ASAT) are intentionally launched by a weapon system that is not placed in orbit (DIA, 2019, p. 10). These tactics also produce orbital debris, increasing the likelihood of accidental or unavoidable collision. China's 2007 ASAT test for example, significantly and instantly increased the amount of on-orbit debris by 25% (Lafleur, 2011, p. 1).

Other types of counterspace activities open vulnerabilities for active satellites. Aside from kinetic physical-type attacks, ASAT capabilities include non-kinetic physical, electronic, and cyber. The first type (kinetic physical) houses conventional weapons capabilities that represent the pinnacle of traditional arms: direct-ascent ASAT, co-orbital ASAT, and ground station attacks, among others. These types of attacks tend to cause permanent damage to compromised systems while simultaneously demonstrating the strength of a state's military capabilities. While kinetic physical attacks have the highest probability of the loss of human life, no state has ever conducted a kinetic physical attack against another state's satellite. However despite this, the U.S., Russia, China, and India all have successfully tested direct-ascent ASAT weapons (Harrison et al., 2020, p. 3).

Non-kinetic physical on the other hand, can be labeled as the inverse of the first as it groups together capabilities that are not visibly physical, including high power radio frequency, high energy-powered lasers (also referred to as dazzling), neutral particle beams, EMPs (electromagnetic pulse), and HPM (high powered microwave) (Harrison et al., 2020, p.3-5). Such capabilities do not make physical contact with targets. High-powered microwaves for example, can cause irreversible damage to a satellite's circuitry, as well as disrupt critical electronic hardware. More notably, non-kinetic attacks can be less visible to third-party observers and more difficult to attribute; which makes assigning responsibility to a perpetrator all the more difficult (Harrison et al., 2020, p.3). These two categories respectively (kinetic and non-kinetic), are what Krepon characterizes as demonstrable and inferred capabilities- so named for their inherent physical or non-physical attributes (Krepon, 2013, p.15).

Electronic capabilities typically refer to uplink and downlink jamming and spoofing. Such weapons capabilities target the electromagnetic spectrum through which space systems transmit and receive data. Similar to non-kinetic physical attacks, electronic threats are difficult to attribute. This is due in part to the difficulty in distinguishing them from accidental interference. Some forms of attack such as jamming or spoofing are reversible because once turned off signals can return to normal (Harrison et al., 2020, p. 4).

Lastly, cyber capabilities refer to data intercept or monitoring, data corruption and seizure of control of a spacial system. Cyberattacks target the data itself as well as the systems related to the flow of data. They could result in the loss and degradation of data or services which a satellite provides. The barrier to entry is also relatively low; cyberattacks can thus be contracted out. This means that even if an adversary lacks internal cyber capabilities, it could still very much pose a threat (Harrison et al., 2020, p. 4-5).

The last broad category of satellite vulnerabilities to be discussed are dual-use spacecraft in unmanned proximity operations. These types of operations involve maneuvering a spacecraft close enough to another object in order to physically affect it in some way (Chow, 2018, p. 108). Active debris removal (ADR) or on-orbit servicing (OOS) are two examples of peaceful uses of unmanned proximity operations. Despite its peaceful intent, ADR and OOS could just as easily be commanded to operate as space weapons. The dual-use capacity is critical because it implies that an object launched into space with the purpose of ADR or OOS (in other words, not specifically manufactured for military purposes), has the potential for weaponization, could nevertheless be permitted under international treaties such as the draft PPWT[4] (Su, 2010, p. 86).

As illustrated above, there exists many threats to satellite systems in outer space, made more potent by enabling factors such as technology drivers and large-scale investment. Eftimiades states that "…one could argue that the nature of the emerging commercial space sector is disruptive and, therefore, threatens US hegemony in space activity and technology (Eftimiades, 2022, p. 4)." Astorino-Courtois and Bragg (2018) define disruptors as "…actors whose behaviors and innovations trigger broad change in a system (Astorino-Courtois & Bragg, 2018, para. 4)." Put alternatively, a commercial disruptor such as SpaceX is a company that significantly alters- for good or for bad, the ability of the United States to fulfill its national security space objectives.[5] While Astorino-Courtois and Bragg argue that a disruption can change the nature of the relationship between the U.S. government and commercial space actors, a disruption can also take the form of serious impairments to national security strategy. Specifically, that the changing technological environment, or the disruptors, necessitate a revision of

[4] PPWT is the abbreviated form of Treaty on the Prevention of the Placement of Weapons in Outer Space and of the Threat or Use of Force against Outer Space Objects. It is a joint draft treaty proposed by Russia and China in 2008 which seeks to resolve some of the ambiguities in the Outer space Treaty especially with regards to the use of weapons in space (Jaramillo, 2009, para. 1).

[5] Astorino-Courtois and Bragg's 2018 report: Commercial Space Actors: Disruptors or Solid Partners for National Security concludes that commercial actors' organizational advantages in terms of innovation will make it likely that they will become the dominant actors in outer space in the medium to long-term. Secondly, they state that if the U.S. government does not embrace these commercial actors, the government will lose influence and potentially degrade national security. Furthermore, controlling or limiting these actors too tightly may effectively mean losing their business to a foreign country (Astorino-Courtois & Bragg, 2018, p. 8).

underlying theories that guide national security strategies, such as that of deterrence theory.

This means that states now must consider new challenges to the space governance regime with regards to the proliferation of anti-satellite capabilities. According to the principles of deterrence theory, a state may effectively deter an aggressor by demonstrating that the cost of attacking is more than the benefit expected to be gained from such an attack. Traditionally, deterrence theory operates in a demonstrable weapons capability environment, wherein the ability to visibly detect or identify the negative consequences of attack is crucial in deterring adversaries. After all, if a potential attacker is unable to recognize the signs or repercussions of an attack, how can they otherwise be deterred? More importantly, the proliferation of satellite capabilities significantly alters this cost-benefit calculus due to inferential satellite capabilities (i.e. non-kinetic physical, electronic, and cyber). The lack of ability to perceive aggressions is a substantial shortcoming in classical deterrence theory. Thus, since threat perception is one of the fundamental points of examination in classical deterrence, and because conventional weapons capabilities have evolved to include virtually undetectable forms of attack, it is not unreasonable to conclude that successful deterrence will be difficult to achieve in the domain of outer space. How then, can a state reliably deter potential adversaries when the traditional method of signaling the intent to use force is no longer readily visible? The discussion of these new elements to the deterrence regime is thus a timely, and highly relevant undertaking.

Chapter II.

Deterrence: Inception to Cold War

This chapter begins with a general overview of deterrence theory, then proceeds with a discussion on nuclear deterrence as well as its potential applications in outer space. This chapter then analyzes deterrence's core tenet of credibility, focusing on the mechanisms of signaling and perception. Next, criticisms of classical deterrence are discussed followed by an examination of the principles of credibility in the context of outer space and inferred anti-satellite capabilities.

Since an authoritative source on deterrence theory does not currently exist, its principles must be constructed from a collection of various sources. Its inception lay in the argument put forth by Bernard Brodie's *The Absolute Weapon (1946)* , which illustrates the context out of which classical deterrence theory emerged. Brodie asserts that the international environment pre and post-World Wars were fundamentally different from each other and as a consequence, a new theory would need to replace the conventionally held theoretical paradigm of the past (Zagare & Kilgour, 2000, p. 3). The pre-World War international system was multipolar, consisting of a mix of states of varying power that competed for authority and influence. The defeat of the Axis powers in 1945 yielded a post-war system predicated on the balance between two hegemonic states and its unshakable bipolar stability. It was this redistribution of powers that created the conditions that would result in the emergence of deterrence theory as an intellectual tradition.

Although nuclear deterrence is the main strand of theory this discussion looks at, there are several other theoretic strands within the deterrence literature including direct and extended deterrence, complex deterrence, deterrence by punishment and denial, intermediate and general deterrence (including sub-strains put forward by Paul Huth- direct-immediate, direct-general, extended-immediate, and extended general deterrence), tailored deterrence, and the most recent addition to the theoretical strand, perfect deterrence theory (Quackenbush, 2010, p. 61). Each type varies on how one deters an aggressor, how long of a duration deterrence measures are undertaken, as well as who is the target of the deterrence strategy.[6] While there is still debate as to whether each variant constitutes its own theory, Frank Zagare (1996) argues that much of the deterrence literature can be housed under the single doctrine of classical deterrence theory (Quackenbush, 2011, p. 5).[7]

Deterrence as a concept generally refers to "…the use of threats by one party to convince another party to refrain from initiating some course of action (Huth, 1999, p. 26)." It is effectively, the practice of discouraging an act through the effectuation of negative consequences in order to generate compliance (J. Black-Branch, 2020, p. 325). While there exists no clear-cut set of rules that define classical deterrence- the first iteration of deterrence theory, its main principles can generally be summarized as:

1. The international system is inherently anarchic (Mearsheimer, 1990, p. 44).
2. Actors are motivated either by their nature to maximize power, or by their environment to maximize security, reflecting its intellectual origins of

[6] For further discussion on deterrence types refer to Mazarr (2018) and Arie (2016).

[7] While Zagare indeed argues that much of deterrence literature can be housed within a single theory: classical deterrence theory, he also allows for the division of classical theory into two subgroups: structural deterrence theory and decision-theoretic deterrence theory (Quackenbush, 2011, p. 6).

realpolitick, “political realism”, or “power politics” (Zagare & Kilgour, 2000, p. 7).

3. A bipolar system contains only one power dyad across which war may break out and is thus more stable than a multipolar system which has many potential conflict situations (Mearsheimer, 1990, p. 14).
4. Successful deterrence requires examining deterrence’s core tenets:
   - Cost-benefit calculus: The process of weighing the relative costs versus benefits of a given deterrent action in order to determine whether or not it is reasonable to carry out said action. Generally, deterrence’s core tenets are factors that influence the cost-benefit calculus of a given situation.
   - Credibility: A threat may be considered credible if the defender possesses the military capabilities to administer considerable damage (costs) on the attacker, and if the attacker believes that the defender is resolved to employ its force (Huth, 1999, p. 29). In other words, it is the effective communication to an adversary through the means of a deterrence posture.
   - Motivation and Interests: Motivation and Interests: Huth refers to this concept as *interests at stake* and is the justification of risks a state leader maintains in order to legitimize the act of entering into a military conflict (1999, p.34). In other words, when heads of state have critical interests at stake, they will be more disposed to use force and absorb military losses for the purpose of securing those interests.
   - Rationality: Rationality: within deterrence theory, rationality refers to the use of rational judgement; whereby both state actors, who are subject to a

hostile relationship, calculate the costs and benefits of an attack option. If the costs outweigh the benefits then the state may be considered rational only if the decision to attack is abandoned (Koichi, 2016, p.23).

  - (Costly) Signaling: costly signals communicate the credibility of a defenders resolve and refers to actions and statements that clearly increase the risk of a military conflict and the costs of reneging from any given deterrent threat, thus exposing the level of commitment of a state to defend against an attack (Huth, 1999, p.31)
  - Stability: within deterrence theory, stability is represented as precautionary measures taken to prevent one's retaliatory deterrence posture from being undermined as a result of an adversary's belief of one's deteriorating posture; resulting in the adversary's decision to perform a preemptive strike (Koichi, 2016, p.23).

Classical deterrence theory is generally divided into two subgroups: structural deterrence theory and decision-theoretic deterrence theory. Structural deterrence is closely aligned with realism and argues that the balance of power is stabilizing, thus engenders peace. If two states are equal in power, each will be deterred from taking any preemptive actions since neither will be able to confer an advantage over the other. Decision-theoretic deterrence theory on the other hand, implements game theory and expected utility to construct models of deterrence in order to calculate the distinct point of threshold wherein a state is either willing to initiate or defer (i.e. be deterred) from conflict (Quackenbush, 2011, p. 6).

Scholars that heavily contributed to the theory's early development and refinement include Herman Kahn, Thomas Schelling, Albert Wolstetter, Oskar Morgenstern, William Kaufmann, and Glenn Snyder; who as Zagare and Kilgour remark, produced a theoretical doctrine that "…came to be seen as the Rosetta Stone of the nuclear age (2000, p. 4)." So impactful was deterrence theory, that it not only provided a means of explaining why there was no subsequent buildup of a U.S.- Soviet war after 1945, but could also be employed as an effective informative strategy to navigate possible conflicts between superpowers. Zagare and Kilgour remark further that "…if ever the was a theory that enshrined the status quo, this was it (2000, p. 5)."

Deterrence has always held its place as a relevant strategy for national security, and its policies have more often than not, been the pervasive framework with which conflicts have been explained and rationalized. Its use as a strategy to explain the actions taken by nation states has been and continues to be widespread. For example, in a study examining extended-immediate deterrence between 1885-1984, there were 48/58 cases of attempted deterrence (83%) involving major powers as defenders (Huth, 1999, p.27). Other examples of the use of deterrence strategies include the crises that led to both of the World Wars (in which great Britain failed to deter Germany), and the United States establishment of an alliance and military presence in South Korea following the end of the Korean war to deter against the threat of potential invasions from North Korea (Huth, 1999, p.27). While traditionally thought of as a policy for superpowers, deterrence has also been used by smaller states in territorial conflicts where major powers defer intervening directly (i.e. proxy wars). Accordingly, deterrence is a theory that has broad applications and is most potent in conflicts that demonstrate a state's military capabilities.

What reinforced the idea of maintaining deterrence strategy within the conventional military toolkit was the advent of nuclear power, where the build-up of military capacity- particularly nuclear arsenals by both the U.S. and the Soviet Union, buttressed the principles of bipolarity and thus came to be acknowledged as necessary conditions for peace and stability. Mearsheimer (1990) ominously concluded that the post-war multipolar system would increase the likelihood of major crises and armed conflict in Europe, and attributes three factors for the absence of war: "…the bipolar distribution of military power on the Continent; the rough military equality between the two states comprising the two poles in Europe the United States and the Soviet Union, and the fact that each superpower was armed with a large nuclear arsenal (Mearsheimer, p. 6-7)." So deeply entrenched was the bipolar approach to political thought, that classical deterrence theorists like that of Mearsheimer contemplated whether or not the West was paradoxically interested in maintaining a powerful Soviet Union (Mearsheimer, 1990, p. 52).

Indeed, the conventionally held belief that the Cold War era was inherently stable due to its bipolarity does not logically follow after the collapse of the Soviet Union, where the international system gave way to multipolarity. If according to classical deterrence theory, the Cold War was the apotheosis of bipolar stability, then how could one explain the relative peace and prosperity the inherently instable, post-war multipolar system brought to bear? Such is the fundamental inquiry of classical deterrence theorists, who attempted to rationalize multipolarity at the conclusion of WWII.

Some classicalists have called the era of Pax Americana, or Long Peace brought on by this multipolar phenomenon an aberration, others reasoned that it was simply the result of several factors including economic liberalism, peace-loving democracies, bipolarity of the distribution of power, parity in military power, and nuclear weapons- which significantly enhanced the cost of war, in turn making deterrence increasingly effective (Mearsheimer, 1990, p. 11). Another prospective explanation came from the obsolescence of war theory, which posited that modern warfare has advanced such that it can no longer be considered as a sensible means to achieve national goals.

However, all explanations fail to account for the 45 years prior to the Cold War. During this period, there was little economic exchange between the Soviet Union and the West, thus the economic liberalist theory did not adequately explain the era of Long Peace. The peace-loving democracies approach also fell short because it overlooked the fact that the Soviet Union and its allies had not been democratic during the same period. Lastly, obsolescence of war is not entirely convincing because it assumes that all conventional wars are drawn-out, bloody wars of attrition. It is possible for instance, to achieve a quick and decisive victory without inciting the costs that usually accompany a protracted war (Mearsheimer, 1990, p. 30). A state could resolve to use nuclear weapons in a precise, limited capacity thus ending a conflict within a short period of time and without incalculable damage.

In fact, it was on the heels of the Korean War in the 1950s that the idea of limited war emerged as a viable option in the repository of nuclear deterrence strategies. Limited war was a way to escape the automatic escalation of massive retaliation, which was brought on by the U.S.'s infamous policy of 'striking at a time and place of our choosing.' 'Limited' within this context primarily referred to the range of destructive power a tactical weapon held, which also meant that it had the capability of containing destruction to a point that did not spell out certain doom for humanity.[8] Thus, the fear of automatically escalating conflict to its highest levels was, as proponents of limited war suggested, not guaranteed (Freedman & Michaels, 2019, pg. 135-153). In the modern era, instead of limiting war in terms of geographic and escalatory factors, we are beginning to recognize the utility of hypersonic capabilities as one that circumvents conventional deterrence. Hypersonic capabilities, or assets that can travel at speeds of Mach 5 or greater, have been touted as revolutionary since there are currently no systems that can effectively mitigate its effects- they are too fast to be detected, and if equipped with nuclear warheads, too destructive to simply overlook (Atherton, 2021, para 1-3). Taken holistically, the hypersonic conundrum is not new; rather it is a reiteration of traditional deterrence in a new form. Thus, we are still left with several inconsistencies of deterrence that require redress, including the unaccounted era of Long Peace and relevant conditions that existed prior to the start of the Cold War. However, these are relatively minor infractions in comparison to the issue of deterrence theory's inapplicability to anything other than nuclear weapons capabilities.

Most criticisms of classical deterrence stem from its lack of applicability outside of a nuclear weapons-reliant international environment. Fischoff (1987) denounces deterrence theory as "…a parody of a scientific theory. Its fundamental behavioral assumptions are wrong. Its basic terms are ill-defined. It is used in inconsistent and contradictory ways. Commonly cited examples of effective deterrence are often based on flawed readings of history, sometimes reflecting ignorance, sometimes deliberate misrepresentation (Harvey, 1998, p. 675)." Many other scholars such as George and

Smoke (1989), Jervis (1989) and Lebow and Stein (1989), have provided evidence that rational deterrence is unrealistic (Harvey, 1998, p. 675). What appears to be ubiquitous is that each tenant of deterrence is in themselves concepts which are interpreted in differing ways, and thus it is difficult to come to a consensus about how these parts interact with each other, as well as what results as a consequence of each interaction. This in effect, leads to what Zagare and Kilgour call the *paradox of mutual deterrence*, or the contradiction that arises when, in spite of the direct relationship between the status quo and the cost of conflict, the status quo fails to be a rational choice due to both states preferring capitulation (Woo, 2020, p.10).[9] Zagare and Harvey among others, have attempted emphasize the need for continuing research in this area in order to manage the international system in the future (Zagare, 2004, p.382).

In fact, Zagare lists several methods other scholars have tried that could potentially provide recourse to resolution; one includes developing a theory that requires actors to be simultaneously rational and irrational. In another, one could construct a new theory based on a so-called divergent theory of rationality (breaking away from the classical definition), and yet another could attempt to show that deterrence theory does not necessarily have to be paradoxical in nature (Zagare, 2004, p. 381). However, Zagare posits that any reconstruction or resolution should account for the absence of a war (i.e. the obsolescence of war theory) between superpowers post- Cold War, in addition to the empirical relationship between power parity and major power conflict (Zagare, 2004, p. 382).

---

[9] Another way to think about the paradox of mutual deterrence is in the following terms: by conceding, the defender is able to acquire a more favorable outcome than if she chose to defy and thus escalate the conflict. However, if the challenger knows that the defender will concede, the challenger will always attack and thus deterrence will always fail.

However, remedying the persistent and great theoretical issues of deterrence theory is not the purpose of this discussion. Rather, it is in recognition of these issues in the face of a significantly altered international environment, and the acknowledgement that deterrence theory is increasingly ineffective, that deterrence cannot be expected to effectively function in any other environment except that from which it originated without altering its core tenants in some way, or affixing a supplemental mechanism to account for its inconsistencies.

Chapter III.

Core Tenets, Assumptions, and Prerequisites

Most of the shortcomings that afflict classical deterrence theory are products of the theoretical construct within which it operates. Since the theory relies heavily on the presumption of a nuclear-armed environment, its shortcomings have risen out of the need to explain deterrence strategies outside of the nuclear context. In outer space where the nature of conflict has evolved substantially, the theoretical failings of deterrence have become increasingly more evident and are aggravated by the proliferation of satellite capabilities. For these reasons, it is important to go over each principle in brief before beginning the discussion on credibility; of which is an essential factor in formulating an effective theoretical framework for deterrence in space. As noted in the previous chapter, deterrence calculus is based on the tenets of cost-benefit calculus, (costly) signaling, stability, credibility, rationality, and motivation.

To begin, the basic assumptions of deterrence are as follows: 1) the primary unit of analysis is the international system, which is in itself anarchic, 2) the system is comprised of secondary units of analysis called states, which are assumed to be rational decision-making entities (Zagare, 2004, p. 109).[10] An instrumentally rational actor will for example, be able to arrange a set of choices in ascending order of preference based on

[10] The concept of rationality has been one mired by consistent misinterpretation and confusion within the social sciences, attributable to divergent theoretical traditions, which are generally taken to be incompatible: procedural rationality, which uses a psychological approach, and instrumental rationality, which is favored by rational choice theorists. See Zagare and Kilgour (2000) pages 38-44.

completeness and transitivity.[11] Lastly, 3) the notion that international stability is dependent on the distribution of power and the absolute costs of war. This last point is perhaps of the most consequence because it is what theorists have used to reach the conclusion that bipolar systems are inherently more stable than multipolar systems, and that anything which increases the costs of war, will only serve to reinforce the stability of bipolar relationships (Zagare, 2004, p.110). [12] Classical theorists further assert that the era of peace that began after the Cold War endured due to the belief in the assumption that war becomes irrational after having achieved parity in terms of power.[13] It follows theoretically that power disparities encourage war by increasing the potential for aggression, and that the possibility of armed conflict is conversely mitigated when they are minimal (Mearsheimer, 1990, p. 18). These core assumptions largely guide the theoretical matrix of deterrence, but whether or not a strategy is successful is dependent on its meeting or exceeding a set of prerequisites.

Literature within the doctrine often varies with regard to the precise definition of each prerequisite, but the underlying principles remain consistent. The four basic prerequisites for successful deterrence, also referred to as requirements or conditions, are as follows:

---

[11] Completeness as defined by Zagare and Kilgour (2000) is the capability of an actor to discern a relative preference given a set of choices. On the other hand, transitivity is the logic that "…for any alternatives *a*, *b*, and *c*, if an actor prefers *a* to *b*, and *b* to *c*, then the actor must also prefer *a* to *c*. (p. 40)."

[12] For further discussion on why bipolarity is inherently more stable than multipolarity, see Zagare and Kilgour (2000) pages 8-9, and Mearsheimer (1990) pages 13-19.

[13] Mearsheimer (1990) also contends that in addition to bipolarity, the stability of the postwar era endured because of the approximately equal military strength between the U.S. and the Soviet Union, as well as the development of nuclear weapons which increased the gravity of conflict, thus making deterrence even more relevant (1990, p. 11).

A. Clearly defining what constitutes unacceptable behavior

B. Communicating a commitment to punish any violations to challengers

C. Possessing the means (capability) to defend the commitment[14]

D. Demonstrating resolve to carry out retaliatory responses in the event of noncompliance (Harvey, 1998, p.676).

These prerequisites are guidelines that that exemplify the theory's core tenants and are representative of those most cited by theorists when analyzing strategies of deterrence. Most if not all of the scholarship on the conditions required for successful deterrence reference these four basic prerequisites in some shape or form. For example, George and Smoke's (1974) two major conditions (within this context, prerequisites) require the knowledge of:

> (1) the initiator's view of the exact nature of the defender's commitment, if any, on behalf of its ally or friend; and (2) the initiator's judgement of whether the risks of a particular option open to him can be calculated and/or controlled so as to make that option an acceptable risk (George & Smoke, p. 523).

In this case, (1) and (2) point to the general principle of costly signaling and can be interpreted to mean that the extent to which an initiator's (aggressor) perception of the defender's commitment is necessary, but not a sufficient condition for successful deterrence. In other words, by incorporating both prerequisites A and B more broadly, it is necessary not only for the defender to effectively communicate resolve to carry out

[14] Harvey further explains that a nation state may defend a commitment by punishing adversaries who challenge it, or by denying the challenger the attainment of certain objectives. Further, if a defender wishes to be seen credible, they must have the capacity to defend their commitments since the lack of ability to defend encourages the challenger to doubt the defender's resolve (1998, p. 676)

punishment, but for the aggressor to have the capacity to perceive such a intent in order to have an outcome result in successful deterrence.[15]

George and Smoke also list six other minor conditions in addition to the first two major conditions: (3) the aggressor's perception of the military capacity of the defender, (4) the aggressor's perception of the defender's motivation, (5) the aggressor's perception that only force or the threat of force can achieve the desired change (i.e. in the status quo), (6) the aggressor's willingness to accept an alternate form of compensation, (7) the extent to which the aggressor is motivated to change the status quo, and (8) the extent to which time pressures the aggressor for achieving desired change (George & Smoke, 1974, p. 530-532).[16] What is particularly notable, is the fact that these conditions deal in perception; be it of the aggressor, defender, or of the use of force. While other variations of the prerequisites incorporate other tenets such as rationality or motivation, the larger portion of literature focuses on perception, credibility, and rationality.

Within the debate of nuclear deterrence for example, Sagan and Waltz state that its successful implementation requires that a state's nuclear arsenal appear to have the capacity to survive an attack, be able to launch a retaliatory strike, and that a state must

---

[15] This is consistent with Thomas Schelling's 3 C's of capability, commitment, and communication. To Schelling, war was a process not wholly controllable, thus deterrence needed to be premised on incalculable risk and bargaining power that lies in the capacity to 'hurt'. In demonstrating that you have the capabilities to carry out an attack, one must already have had the means to effectively communicate it to an adversary, and ultimately be willing to execute one's attack as initially communicated For more on Schelling's 3 C's, see Arms and Influence (1996) chapters 5-7.

[16] Whether or not a condition is necessary or sufficient is the foundational logic behind the necessary and sufficient condition hypothesis as explained by Most and Starr (Most & Starr, 2015) . They take the study of logic and apply it to analysis in international relations. Such a methodology has been used to empirically test and model deterrence theory. It also provides an approach that reduces the tendency on the part of international relations analysts to work backwards from the result of events (e.g. war) in order to isolate their sufficient antecedents (causes) which leads to overlooking cases where the causes occurred even though it was not preceded by the result. That is, working backwards does not allow the analyst to determine whether or not a cause was sufficient for the result. For a detailed discussion, see Most and Starr (1989) pages 47-67.

not respond to what may be false alarms while maintaining control of their arsenal (1995, p. 20). In different analysis on extended deterrence, the RAND Corporation identified 16 variables that influence deterrence outcomes (Mazarr et al., 2018, p. xii). These factors were further grouped according to the following questions: how intensely motivated is the aggressor? Is the defender clear and explicit regarding what it seeks to prevent and what actions it will take in response? And, does the potential aggressor view the defender's threats as credible and intimidating (Mazarr et al., 2018, p. 35)? As demonstrated above, it can be reasonably stated that throughout the literature on deterrence, the core tenets such as motivation, signaling, perception (of the aggressor and defender), as well as credibility, are all consistently addressed. Although, one must ask whether or not the common themes exhibited in these prerequisites are the only indication of uniformity. In an introductory commentary on the strategy of conflict and deterrence, Schelling asserts:

> What is impressive is not how complicated the idea of deterrence has become, and how carefully it has been refined and developed, but how slow the process has been, how vague the concepts still are, and how inelegant the current theory of deterrence is…the literature on deterrence and related concepts has been mainly preoccupied with solving immediate problems rather than with a methodology for dealing with problems (Schelling, 1981, p. 7).

As a result of the processes by which deterrence theory emerged, there exists an abundance of interpretations of what defines its prerequisites, in which situations it be considered a success or failure, as well as which factors necessarily elicit successful use of its strategies. While the various results of these processes has led to the great expansion of the theory, as evident by the many types of deterrence, the formulation of a theoretical framework which can broadly be applied to many different types of strategic situations is yet to be seen. After all, deterrence is in effect a theory of the nonuse of military force and requires a broader methodological basis than what currently is available (Schelling, 1981, p. 9). Thus, it is not surprising that scholars have been attempting to affix new theoretical components and empirical frameworks to deterrence theory in order to address many of its shortcomings in the modern age.

Credibility

The tenet most impacted by the proliferation of inferential anti-satellite capabilities is that of credibility, which will be the object of careful examination in this section. Inferential anti-satellite capabilities are unique in their ability to remain largely undetected by the physical eye; only through its effects can their impact be observed, and even this can be limited in some cases. There are currently two different approaches to examining whether or not a threat can be considered credible: the first is a theoretical examination, using the largely lexical style of argumentation that deterrence theorists have thus far used to explain the causes of war, and the second is a game-theoretic method, which explains credibility in terms of sub-game perfect equilibria. The latter method will be expanded upon in Chapter 5.

As a general rule, credibility can be characterized as the effective communication to an adversary through deterrence posture, so as to compel the adversary to believe the utility of the planned attack. A threat may be considered credible if the defender possesses the military capacity to exact enough costs on the attacker that the later believes the former is resolved to use its capabilities (Huth, 1999, p. 30). Deterrence theorists have routinely argued that deterrence success is more probable if a defender's deterrent threat is perceived as credible to an attacker (Huth, 1999, p. 29). This is made more apparent when the attacker can demonstrably (i.e. physically) see a defender's commitment to retaliate.

However, it is difficult to formulate a broad principle of credibility since its success largely depends on a wide scope of factors including commitment, capability, communication (i.e. Schelling's 3 C's), as well as frequency of use and attribution. The main difference between nuclear and inferential anti-satellite capabilities is the extent of visibility each maintains. Although both can be physically seen; that is, both a satellite and a nuclear warhead can be spotted in their respective domains (e.g. in outer space or in a silo), only the later reserves the capacity for a purely visible attack. While being physically present allows for a certain degree of deterrent power, there remain issues of whether or not the given capability will in fact be employed.[17] One can argue that it is the accumulation of an actor's past behaviors that render a threat credible. In this way, the frequency of use becomes a relevant measure of credibility for anti-satellite capabilities. Let's propose that state A frequently implements inferential anti-satellite capabilities to interfere with state B's radio broadcasting frequencies in order to repress the transmission of liberal ideas within state B's borders. Does state A's consistent behavior in its use of inferential anti-satellite tactics lend it credibility in the instance it threatens to escalate its interference by targeting physical radio infrastructure in order to further its ideological objectives? Not necessarily. The dilemma with credibility is that it does not operate consistently across all levels of aggression unless said aggression is already

[17] The presence of a weapon, including the display or demonstration of military strength either formally within the context of a war or informally through a military parade, are tactics that can elicit a deterrent effect. This is due to the fact that such a display permits the demonstrator to explicitly communicate the extent of its military capability, which is then perceived by its target state who in turn can perceive the demonstration as behavioral basis on which to judge the credibility of any future threats made by the aggressor; if the demonstrator issues a threat based on the military capabilities it had previously demonstrated, then the earlier demonstration had lent it some level of credibility.

established as an expected behavior. Establishing this behavioral consistency is a component of the means within which credibility functions, thus consistent behavior must be one of the factors to be considered when evaluating whether or not a threat is credible. For example, North Korea has launched and increasing number of missile tests, starting with four in 2020, to eight in 2021, and more than 90 by the end of 2022 (Dotto et al., 2022, para. 1-2). It can be argued that simply by increasing the frequency of launch, that North Korea is escalating. However, let's instead focus on the idea of North Korea deciding to escalate by employing a more destructive type of missile for its tests. This allows us to theoretically examine whether or not North Korea's threat to use more destructive missiles can be considered credible after having used less destructive missiles consistently since 2020. Setting aside any challenges to this notion based on the rationality of its state leader, Kim Jong-un, we cannot conclude that a threat to escalate based solely on the move towards a more destructive missile is credible, since no prior, consistent actions have proven North Korea to commit to employing a missile of increasing destructive power.

In the context of nuclear deterrence, the success of credibility rested primarily on the visibility of a nuclear stockpile and was strengthened by highly visual displays of the amassment of arms (Krepon, 2013, p. 15). Although kinetic-physical type ASAT are similar to the traditional concept of weapons, non-kinetic physical, electronic, and cyber-type attacks are largely inferential, suggesting that the comparatively low visibility of such capabilities may significantly reduce the level of an opponent's ability to perceive threats. Inferential ASAT capabilities simultaneously increase the perception threshold a state or nonstate actor must meet or exceed in order for its opponent to register any given signal as existent. Thus, in order for a threat to be credible, it must be signaled (communicated) by the aggressor, and perceived (acknowledged) by the defender. Since the value of signaling lies in the opponent's perception, deterrence failures in this regard mainly arise from the misinterpretation or imperception of such signals (Mazarr et al., 2018, p. 11). Perception may be further impaired due to limitations in space situational awareness (SSA).[18] For example, not all satellites are monitored at all times, and even then diagnostics and environmental monitoring capabilities are limited. This not only makes it difficult to monitor and diagnose satellite failure, it prevents states from being able to detect adversary signals or signs of preemptive aggressions (Morgan, 2010, p. 15). Being unable to determine the source of hostile behavior inevitably means that the ability to respond in kind is degraded. Ultimately, this indicates an erosion of deterrence since issuing out a retaliatory response is a vital component of establishing credibility for deterrence strategies. Thus, one of the priorities for deterring adversaries in outer space should focus on developing a robust attribution process. One way to accomplish this is by taking inspiration from existing nuclear attribution processes; creating a triad comprised of SSA and space forensics, the intelligence community (IC), and allies and commercial partners (Klein, 2020, para. 4). While SSA can detect many variations of kinetic ground and space-based capabilities, it is unable to readily to recognize non-kinetic (inferential) aggressions such as attacks on networks, jamming, and lasing. Space forensics, which refers to the process of surveying and determining the source and pathway of aggressions on space architectures, is a method that can simultaneously bolster SSA as well as

[18] SSA has also been referred to as space domain awareness (SDA), as coined by the United States Air Force. This is due to the fact that the U.S. military now considers outer space as a domain of warfare, much like those of the land, sea, and air (Erwin, 2019, para. 4).

address the proliferation of inferential ASAT capabilities (Klein, 2020, para. 7). Relying on a multi-component attribution process can serve to increase the deterrent strength of spacefaring nations by lending credibility to retaliatory threats, thereby mitigating to an extent, the issue of attribution posed by inferential ASAT.

Arce and Sandler (2009) state that "…in the current era of asymmetric warfare, the principles of credibility and proportionality assume increased importance when investigating deterrence (Arce & Sandler, p. 406)." With the emergence of technological advancements, the focus of deterrence theory must also logically shift to account for non-traditional threat capabilities and actors at all levels of conflict. Such a shift can be reflected in Lupovici's observation that an empirical shift frames the classical realist paradigm in the context of contemporary threats such as terrorism, rouge states, and ethnic conflicts (Lupovici, 2010, p.705). This is important to note because it hints at the refocusing of the theory from a bipolar environment to one that is multipolar; an environment inhabited not only by nation-states, but by non-state actors who also have access to the same types of satellite capabilities that could, in the acquisition of such capabilities, act to deter or threaten nation-states. Needless to say, the current multipolar environment in conjunction with ballooning satellite technologies poses a threat to states traditionally accustomed to dealing with internationally recognized entities as opposed to terrorist groups or third party actors who do not necessarily adhere to the standards of rationality.

In fact, it is relatively easy for a non-state actors to obtain the components necessary to manufacture an anti-satellite weapon. In 2009 Brazilian satellite hackers used a medley of homemade tech to hijack ultrahigh frequencies (UHF) used by U.S.

Navy satellite transponders (*The Great Brazilian Sat-Hack Crackdown*, 2009, para. 2). Other criminal groups frequently exploit satellites for criminal activities. Rouge loggers in the Amazon for example, use them to disseminate coded warnings when authorities close in, and drug dealers employ them as a tool to coordinate underground operations (*The Great Brazilian Sat-Hack Crackdown*, 2009, para. 6). Not unexpectedly, the use of satellites to achieve national, or criminal objectives are gaining traction. In a show of defiance to a small radio station in Britain for broadcasting issues of human rights and freedom of speech against Libya, the Libyan government illegally jammed two international satellites, subsequently disrupting dozens of TV and radio stations servicing Britain and Europe; even disrupting U.S. diplomatic, military, and FBI communications (Hencke & Gibson, 2005, para. 1-5). These events serve to emphasize that satellites, specifically satellite jamming, are increasingly being implemented to carry out activities that threaten the operations of both civilian and military operations.

More than 80% of satellite jamming incidents have been triggered by diplomatic and political differences between nations. These tactics are now being used to manipulate, deny, and corrupt information necessary to carry out strategic, economic, and military agendas (Mountin, 2014, p. 105). This suggests that the relationship between state and nonstate actors of varying military strength has a high probability of resulting in asymmetric conflict, which will continue to increase as non-state actors acquire satellite capabilities. This begs the question of how a stronger actor should respond to a weaker actor such that, in violation of existing norms or laws, a security challenge results to the disadvantage of the former? The answer requires examining the relationship between credibility and proportionality.

“Proportionality has come to mean the amount of force used to address a security challenge, relative to the precipitating exigency (Arce & Sandler, 2009, p. 385).” Traditionally, credibility has been studied in situations of pre-emptive threats, where their credibility and escalatory behaviors are examined. However, Arce and Sandler have put forth scholarship that examines credible retaliatory threats and responses that obviate the need to escalate conflict. Using game models, they show that both credibility and proportionality of deterrence are dependent on adversary technical capabilities. This suggests that such capabilities could be sufficient indicators of the potential magnitude of the preemptive threat or attack. The results of their modeling show that Chicken and Prisoner’s Dilemma (PD) are at opposite ends of the proportionality- credibility spectrum, and that within the game parameters of Chicken, deterrence strategies have the potential to succeed even when a less-than-proportional response exists (Arce & Sandler, 2009, p. 405).[19] More on Chicken and the Prisoner’s Dilemma will follow in Chapter 5. In asymmetric conflicts when participating actors differ substantially in their military strength, lower levels of proportionate responses tend to explain phenomena such as the proliferation of nuclear capabilities in Southeast Asia, North Korea, and Iran, as well as serve to establish logic-based modeling framework with which to analyze other asymmetric conflicts similar to those between states and terrorist organizations. Thus, how strong of a retaliatory measure a stronger actor should execute on a weaker actor is governed by the type of deterrence strategy being employed, as well as an adversary’s technical capabilities.

[19] Arce and Sandler’s approach is a comparative analysis of Chicken and the Prisoner’s Dilemma (PD), demonstrating that retaliatory responses in Chicken are less proportional but do not always require credibility and that retaliatory responses in PD are credible and could be proportional. On this, Arce and Sandler state that the issue still stands (2009, p. 405).

The application of these findings to deterrence in asymmetric, outer space conflict suggests that the nature of such conflict is more akin to salami tactics. The small incremental degradation of satellite infrastructure through strategies that incur an increasing amount of resources to remedy- such as bumping a satellite out of orbit in the hopes of forcing an operator to burn through its fuel source through repeated maneuvers, rendering it unusable; can be well-taken with actors who maintain an inferior position within an asymmetric relationship (Cole, 2014, p. 3). While falling short of what can be considered an outright hostile move, salami tactics in outer space can be debilitating from an operational perspective. A relatively proportionate response on part of the actor who maintains a superior position could be to bolster its defenses by increasing SSA or SDA, imposing effective space traffic management (STM), hardening satellite infrastructure against radiation, and disaggregating capabilities (BAE Systems, n.d., para 3-6).[20] The lower barrier to entry, combined with the proliferation of actors and capabilities in the outer space domain indicate that the use of salami tactics will also rise in kind. Considering that proportionality and credibility are intimately linked, a less than proportional response could be sufficient in successfully deterring a threat, in accordance with the results of Arce and Sandler's examination of Chicken.

[20] SDA and SDA, as discussed earlier, are useful in presenting a full picture of the outer space environment, including detecting potential kinetic and inferential threats. Effective space traffic management on the other hand, refers to the strategy of addressing the proliferation of objects in space, and how to efficiently navigate such an environment (BAE Systems, n.d., para. 3-6). The nascent field of STM focuses on developing international and national authorities to track, regulate, and oversee object movement (standards of conduct), as well as allocate responsibilities within the space domain (Sadat & Siegel, 2022, para. 5). Other defensive measures for satellite systems, hardening and disaggregation, provide the means with which to increase the sustainability and resiliency of satellite systems. The former is a strategy to physically protect against radiation, while the later increases the resiliency of satellites by employing larger volumes of satellites for the same mission, thus simultaneously increasing the number of targets and decreasing the efficacy of potential aggressions.

Principles of Credibility

The idea that deterrence may succeed when a less-than proportional response exists, goes against the traditionally held view of rational deterrence theorists who argue that successful deterrence is more probable if the defender's retaliatory threat is perceived as credible or stronger to an attacker. This is even more so when the defender possesses the military capability to exact considerable damage to the attacker, and if the attacker perceives that the defender is resolved to use its military strength. While Huth (1999) states that there is no universally accepted set of factors that when implemented, result in a definitive and credible deterrent to threats, he also identifies four convergent factors, if whose thresholds are met, are representative of credible deterrence (Huth, p. 30). These factors are military capabilities, signaling and bargaining behavior, and reputations.[21] Note that other components to credibility not identified by Huth such as those previously discussed: frequency of use and attribution must also be taken into account when examining the credibility of threats.

Strength of Military Capabilities

Military balance or the extent of a state's military capabilities, are a factor in the process of an aggressor determining whether or not an attack can be carried out successfully at a low cost, and is consequently an indicator of how credible an attacker

[21] Huth also lists interests at stake as being a factor that other scholars have considered in their analysis of credibility. Its main focus is to evaluate and determine what specific 'interests at stake' would motivate the defender to risk military conflict. Proponents have claimed that its role is vital in determining deterrence outcomes since the balancing of interests by a political leader inevitably guides the pathways within conflict. Yet the ambiguity involved in ascertaining the ranking (preference) of interests, as well as relying on "interpersonal comparisons of utilities or questionable claims about the attacker's recognition of the legitimacy of the status quo," indicate that the methodology of measuring what constitutes 'interests at stake' is too inconsistent to produce a reliable framework for analysis (Huth, 1999, p.32).

perceives their own military capability. Furthermore, a defending state needs the strength of its military to respond quickly and proportionately to a range of strategic possibilities in order to deny the attacker its military objectives from the outset of conflict (Huth, 1999, p. 30). Military strength also plays into the concept of demonstrability as it reflects the importance for a defending state to possess an adequate amount of military force. In other words, it must demonstrate that it has the military power as well as resolve- that is, capability and commitment, to defend against an attack. Conversely, the absence of military strength would raise the aggressor's probability of a quick victory, in turn increasing the confidence of both political and military leaders of the aggressor state that the benefits of the attack would outweigh its costs.

The demonstrability of military force would substantially weaken in the context of inferential anti-satellite capabilities since they result in little if any, visible display of power in terms of carrying out an adverse action on their target. It could similarly be argued that simply the presence of these satellites in outer space, particularly within a close proximity to other actor's space infrastructure, signals an aggressive posture and can consequently be interpreted to be an action of escalation. Take for example, a technological development that has the capability of rendezvousing with other satellites, then electronically links and assimilates it, then proceeds to perform the same process on other satellites until a collective of compromised satellites forms, producing a swarm of potentially dangerous machines under the command of an adversary (Chris Flaherty, 2023). In a similar vein, the use of ADR or OOS could also pose a danger to space infrastructure. While many have already been placed in outer space for the purpose of debris removal, they can just as easily be re-maneuvered to operate as space weapons (Chow, 2018, p. 113).

Proposed arms control treaties like that of the Treaty on the Prevention of the Placement of Weapons in Outer Space and of the Threat or Use of Force against Outer Space Objects (PPWT) have proposed to ban the placement of weapons in outer space as well as the threat or use of force against outer space objects. While the aim of this proposal is to target directly hostile space systems such as those satellites whose primary function is to incapacitate and immobilize, dual-use systems such as ADR and OOS circumvent this proposition by primarily acting as debris removal and servicing satellites. Jinyuan Su (2010) agrees, stating that "…it is contested that, to the extent that many on-orbit technologies are dual-use in nature, many space objects that have not been specially produced or converted for military purposes, and which are therefore not prohibited by the PPWT, could nevertheless be possible space weapons (Su, p. 86)." Thus, while it may be infeasible to impose an outright ban on such objects, it is possible to restrict their movements and locations. This idea of proximity thus aligns with avoiding inadvertent or unnecessary escalation; a concept that can be readily applied to the physical display of military power.

Costly Signaling

Costly signaling and bargaining behavior are means by which intentions are communicated, and its success is contingent a state's willingness to cross a certain threshold of armed conflict (Huth, 1999, p. 31). Signals are both actions and statements that can increase the risk of military conflict or the costs of capitulation, as well as convey the level of commitment a state maintains to either defend from or carry out its attack. Since costly signaling is dependent on whether or not a signal has been received and properly interpreted, some variables that could potentially degrade a signal's strength such as differing belief systems, or interference- intentional or not; and are points of inflection within the deterrence calculus that if present, are likely to result in deterrence failure. [22]

If the two actors perceive the same situation differently, then the probability of deterrence failure is high. Firstly, since defending states have an incentive to act as though they are determined to repel an attack, aggressor states may choose to disregard statements or actions that signal intent as if they were simply bluffs. Secondly, the assumption that it is always in a state's interest to clearly convey a signal is put into question when a situation arises where a state leader would be constrained in sending out a strong, coherent signal. In other words, it is not always a misperception of signal, but rather the deliberate manipulation of a signal that can cause deterrence failure. Fearon (1997) outlines a few situations in which signals are purposefully degraded on the part of political leaders:

> (a)… uncertainty about the degree of domestic political support for such a policy, (b) concerns that the ability to moderate the policies of allies will be weakened if unqualified support is given to them, or

---

[22]The concept of differing belief systems can be illustrated by Michael and Freedman's quote of Power : "The point is that what will deter us will not necessarily deter the Soviets, and what will make them accept risks will not make us accept similar risks. Moral principles would deter us strongly from launching a pre-emptive war unless there were no other way of averting certain aggression and still greater losses. But moral considerations and the prospect of losing Russian lives and cities would not deter the Soviets from launching an aggressive war if they saw no better way of achieving their objective (Freedman & Michaels, 2019)" In other words, in the event that there exists differing systems of requisite factors between two actors, costly signals are rendered null since the conditions being assessed are irrelevant with respect to one another. Differing perceptions on how an opponent actually behaves and how an actor believe an opponent will behave are thus factors that contribute to deterrence failure.

(c) the risk that rigid diplomatic and military policies may generate high political costs for the adversary in backing down (Huth, 1999, p. 31).

As Fearon notes, considering that there may be political costs to transmitting a fully realized signal, there are cases when the drawbacks of doing so outweigh committing to issuing a credible threat.

Other reasons for signal failure include the lack of perceptibility and deterioration of the strength of a signal. The lack of direct observation, or perceptibility, on part of the target or defender, is fairly straightforward: if a defender cannot perceive the signal then the signal becomes an ineffective means of deterring. Additionally, if the defender knows that the aggressor has an incentive to pretend to be resolved to use power, even if it is contrary to the aggressor's established patterns of behavior, it is representative of a phenomenon that contributes to unsuccessful signaling (Fearon, 1997, p. 69). Lastly, the strength of a signal can be effected by the strength of a domestic audience; in which case, Fearon states that democratic audiences are stronger. Signals put forth by democratic states, as opposed to autocratic states, are more credible because democratic leaders carry greater domestic political costs for retreating during a crisis (Fearon, 1994, p. 577).[23] These various factors that impede the transmission of coherent signals preclude successful credibility, and further demonstrate that the belief systems and patterns of behavior of an opponent must be taken into account when developing an effective strategy of deterrence.

[23] Critics of the claim that democracies are inherently more credible, point to the misinterpretation of data sets most commonly cited to test its hypothesis. For a detailed review see Downes and Sechser (2012) (p. 457).

Means of Costly Signaling: Words and Deeds

While not an explicit convergent factor in Huth's analysis, the means of costly signaling takes on an interesting form in Fuhrmann and Sechser's (2014) research into signaling alliance commitments. "To signal their alliance commitments, states have two basic tools: words and deeds (p. 919)." Their study refines the concept of credibility and takes it in a direction relevant to the question of whether or not treaty mechanisms in themselves, are capable of transmitting a signal strong enough to bolster the credibility of an inferential anti-satellite capability. Since history has shown that states occasionally default on their commitments to allies, the challenge of establishing credibility in alliances lie in convincing adversaries without escalating tensions to armed conflict.

A state may choose to issue a public statement, or a written declaration of its intention to defend another state.[24] Alternatively, a state could take actions to demonstrate its commitment such as the deployment of military forces on an ally's territory during peacetime. While these two methods may be used in tandem, Fuhrmann and Sechser conclude that words, within the context of public alliance commitments and formal defense pacts with nuclear weapon states, reduce the risk of being targeted in a militarized dispute. Deeds on the other hand, specifically in the form of playing host to a benefactor state's nuclear weapons, does not result in the same effect. As such, these types of states do not enjoy a reduction in the risk of military conflict (Fuhrmann and Sechser, 2014, p. 932).

[24] Others such as Leeds (2003b), and Johnson and Leeds (2011) have also found that alliances are valuable deterrents because in their announcement, are strong, public methods of signaling (Fuhrmann & Sechser, 2014, p. 932).

Fuhrmann and Sechser’s findings indicate that nuclear weapons may carry considerable advantages for extended deterrence, and that these benefits do not require the direct placement of nuclear capabilities within allied territory. It is important to bear in mind that this study specifies that signaled ‘words’ (i.e. formal alliances) are effective if the signaling state is a nuclear weapon state (NWS). Since nuclear weapons are have the unique ability to provide their possessors with great strategic benefits, and because they substantially increase the cost of war, they are regarded as ideal assets for those hoping to acquire a deterrent mechanism. In the case where an actor is unable to acquire a nuclear capability, the next logical option would be to form a defense alliance with a state who possesses a nuclear asset; especially since nuclear powers are probabilistically less likely to be attacked (Fuhrmann & Sechser, 2014, p. 921). Furthermore, the efficacy of such alliances are strengthened by the strategy of hand-tying; a tactic which aims to bolster credibility by means of risking one’s reputation through the public declaration of an alliance, whether that take on the forms of a defense pact or a multilateral treaty.

The same cannot be said of non-nuclear weapon states (NNWS) which, according to the study’s findings, are less effective in deterring military conflict (Fuhrmann and Sechser, 2014, p. 932). This phenomenon may be explained by several factors including the fact that benefactor states would station nuclear weapons abroad to defend states other than the host, or that the main objective in stationing nuclear weapons in a particular country is either to prevent that host from developing an independent nuclear capability, or to reassure allies as opposed to deter threats against them (Fuhrmann & Sechser, 2014, p. 929).

Recently, Japan and the United States signed an agreement that extends the U.S.'s security umbrella over Japan within the outer space domain. Formally titled "Framework Agreement Between the Government of Japan and the Government of the United States of America for Cooperation in Space Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies, For Peaceful Purposes," it is a move can be seen as an extension of Article 5 of the U.S.-Japan Security Treaty, which asserts the following:

> Each party recognizes that an armed attack against either party in the territories under the administration of Japan would be dangerous to its own peace and safety and declares that it would act to meet the common danger in accordance with its constitutional provisions and processes (*Japan-U.S. Security Treaty*, 1960).

The application of Article 5 to outer space comes at a time when the military activity in space is increasing at an exponential rate. China and Russia have doubled down on efforts to obstruct the use of space for other spacefaring actors by interfering with satellite communications, developing anti-satellite capabilities, and drafting treaties such as the PPWT that do not specifically outlaw dual-use satellites (Miki, 2023, para. 4). While Japan and the U.S. have long entertained a cordial and strong partnership, the public declaration of extending Article 5 serves to signal Japan's commitment to enhance its counterspace capabilities, and is a particularly loud signal directed at states that pose threats to regional stability such as China and North Korea.

Taken from Fuhrmann and Sechser's perspective, the strengthening of ties between Japan, a NNWS, and the U.S., a NWS, through the public enactment of a new security pact, further cements the strength of general deterrence as applied towards Japan's security- and in this case specifically, its security in outer space. One point worth noting is that while it will take some time to evaluate the true efficacy of this alliance, it may be difficult to acquire the data needed in the near future. This is due in part to the fact that reversible attacks (attacks that do not incur permanent damage) are perpetrated by both China and Russia on a daily basis, but permanent attacks are a matter of military and operational security, and are thus for the time being considered to be classified if it were actually the case that such attacks occurred (Rogin, 2021, para. 4). It is therefore reasonable to conclude that satellite attacks that deal out permanent damage are acts of escalation, and that reversible attacks are comparable to salami tactics since they remain below the threshold for escalation but can nonetheless degrade the integrity and security of outer space systems over time.

Reputations

The layman's notion of reputation does not stray far from that used in deterrence theory, in fact the U.S.'s credibility is also dependent on the maintenance of its reputation as perceived by allies and adversaries alike. In terms of deterrence, reputations can be viewed from the perspective of three different approaches. The first, also referred to as strong-interdependence-of commitments, is based on Schelling's argument that a defender's past record of behavior during times of international conflict is a strong indicator of their expected behavior in similar future situations. As discussed earlier, if such behaviors consistent, the credibility of a defender's deterrence posture is strengthened over time, and the reputation of that defender in showing resolve is causally linked to an aggressor's decision to challenge (Huth, 1999, p. 32).

The second approach, called the case-specific credibility position, argues that in reality, reputations matter little since credibility is more dependent on other factors such as military capabilities, interests, as well as the political constraints of leaders (Huth, 1999, p. 32). Case-specific credibility addresses an inherent weakness in the first; mainly that the strong-interdependence of commitments approach does not justify on what grounds more weight is given to reputation as opposed to other factors that demonstrate the resolve of the defender.

The third approach, also called qualified-interdependence-of-commitments, is what Huth identifies as an intermediate position between the first and second approaches (Huth, 1999, p. 33). The last approach takes into account that reputations could carry influence under certain conditions, and that it should never considered to be the full justification for signaled resolve; although it shouldn't also be completely disregarded. While the third approach makes sense logically, it does not detail under which specific conditions reputations can form and carry weight in signaling resolve. Though, perhaps as long as a defender's actions do not run in opposition to an aggressor's expectations, one could expect a strong correlation between past conflicts and expected resolve.

In determining whether or not a threat maintains deterrent power, the principles of credibility must be addressed. Out of the two general approaches to credibility- lexical and empirical, the discussion of the former in this chapter provides a basic theoretical framework from which to determine whether or not a deterrent threat is successfully signaled (communicated) and perceived, as well as if any external factors including reputations or divergent belief systems hamper the extent of a deterrent threat's credibility. The application of this type of analysis to credibility remains relevant for studying the deterrent success inferential anti-satellite capabilities since their nature precludes the level of visibility required for the analysis of deterrence from the nuclear camp. The inherent deficiencies in attempting to deter an aggressor solely with inferential ASAT can be remedied to a degree by supplementing such deterrent acts with a publicly-made, formal defense pacts with a NWS. The strong effect of hand-tying that accompanies public declarations serves to bolster an inferential deterrent by raising the reputational costs of breaking that pact. While other factors including the physical presence, proximity, and nature of deterrent (i.e. reversible or permanent) can potentially influence the success of a credible threat; they are not necessarily hard-lines that if not met, indicates an inevitable failure to establish credibility.

Chapter IV.

Nuclear Deterrence

In 1984 President Ronald Regan famously declared 'a nuclear war cannot be won and must never be fought', capturing the essence and method of nuclear deterrence (Black-Branch, 2020, p. 324). The nature of this strategy lies in its use of nuclear weapons to issue threats as a means to deter an adversary (Mueller, 2013, p. 44). Since the Cold War, no weapon has been more representative of a state's military power than that of nuclear weapons, and none have matched its level of sheer destructive power since. The iconic image of a mushroom cloud curling up into the atmosphere remains indicative of the unprecedented moral and ethical concerns that nuclear weapons brought to the table. Strategies that sought to control as well as operationalize this power manifested in paradigmatic tactics such as massive retaliation, flexible response, mutually assured destruction (MAD), 'no first use', and limited war.

Proponents argue that nuclear weapons are a necessary evil that serves to prevent armed conflict; a concept which at its surface appears to be contradictory given that nuclear war and peace are inconsistent with one another (Black-Branch, 2020, p. 325). Supporting this side of the argument is the fact that nuclear war has yet to occur, and their nonuse since the end of World War II demonstrates that nuclear weapons have been a pillar of peace between the great powers. Opponents however, argue that the dependence on such capabilities as a means of deterrence has become far too hazardous and ineffective. They point to the fact that nuclear weapons substantially raise the risk of war, international tension, environmental degradation, as well as bring about moral, legal and philosophical issues that stand in direct contrast to the inherent nature of nuclear power

(Black-Branch, 2020, p. 326). Some suggest that the proliferation of nuclear weapons should actively be curtailed, while others have proposed its complete abolition.

By its nature, nuclear deterrence relies on the stability of bipolarity, reinforced by the amassment of nuclear weapons. Thus, the extensive build-up of nuclear capabilities during the Cold War should come as no surprise: between 1947-1991, the United States manufactured roughly 70,000 weapons, and tested them more than 1,000 times. The USSR comparatively built approximately 55,000 nuclear weapons and tested them more than 700 times. These figures add up to an aggregate of 125,000 nuclear weapons- none of which were deployed for use in combat, nor accidentally detonated (Krepon, 2009, p. 4). Under these circumstances wherein the sheer amount of pressure continued to weigh on the international system, the extensive repository of nuclear weapons served to guard against any potential threats of annihilation. The mechanism underlying this world of nuclear deterrence was MAD and massive retaliation; a delicate balance between two powers teetering over the edge of a nuclear threshold. This is why the testing of weapons and the rollout of more advanced military capabilities such as bombers, missiles, and submarines during the Cold War were considered to be essential pillars of successful nuclear deterrence. They were effectively viewed as tools with which to defend against, if only in physical presence, the possibility that the Soviets would launch a first strike. Without such public exhibitions of power, the U.S.'s resolve to defend would not be as convincing to deter possible aggressive actions.

Deep-seated problems within the doctrine of massive retaliation stemmed from whether or not it was appropriate to use against small acts of aggression. Schelling in particular, remarked that there must be some type of automation if the enemy crossed the ‘red line’, which automatically triggered a response in kind. After all, for deterrence to be considered credible, the enemy must understand that you have the capability to attack and that you will commit to attacking. However, questions arose regarding whether or not this unquestionably meant that decisionmakers could leave this in the hands of artificial intelligence and taken out of the hands of an assumed rational human being- who may be tempted to reconsider pulling the trigger. Such a scenario would gradually lead to deterrence failure since the commitment to carry out an attack would be compromised.

The red line debate has long been present within deterrence literature because it associates closely with escalation thresholds. More specifically, the existence or lack of red lines represents the equivalent existence or lack of deterrent thresholds, as applied to the outer space domain. Until recently, only well-established state actors with deep pockets were able to acquire the technological means to deploy and operate satellite systems, making attribution a relatively manageable task (Kleiman & McNeil, 2012, para. 9). However, with the rise of the Internet as a platform to carry out satellite communications, a bevy of potentially adverse actors now had access to many low-cost tactics such as denying, degrading, counterfeiting satellite transmissions, leaking data including imagery, or compromising ground or space-based networks, that could be used to jeopardize a state’s satellite systems (Kleiman & McNeil, 2012, para. 10).

In concluding a 2010 wargame exercise focusing on deterring a combination of space and cyber-attacks, the question of red lines rose the forefront of issues that needed to be addressed for future military operations. Questions such as: what are red lines in space? How does an adversarial state understand what our red lines are in the outer space domain? Can we consider destroying a satellite a red line? Col. Roger M. Vincent, who had commanded the Air Force Space Innovation and Development Center stated, "…if you don't articulate those red lines to the adversary, they will never know when they get close, and if they don't know when they're close, how can they be deterred (Kleiman & McNeil, 2012, para. 20)?"

A potential (and albeit partial) remedy to red lines may be found in Chow's (2018) recommendations for treaty verification regimes, which uses proximity as a fundamental approach to hybrid arms control. Intended as a measure to restrict the types of weapons permitted in orbit, as well as a prohibit the placement of any weapons placed within 0.2 degree in longitude or inclination to another country's assets; proximity holds the potential to more clearly delineate red lines as applied to outer space deterrence (Chow, 2018, p. 116-117).[25] Of course, the literature on red lines more broadly represents the effort of deterrence scholars to resolve the ever-present conundrum of signaling and credibility; an issue that was not readily apparent during the nuclear age.

Nuclear deterrence forced decisionmakers to not only confront ideas of mortality, but to reevaluate what it meant to grapple with bipolarity. In this regard, the focus of the U.S. during the Cold War was singularly directed; affixed to but one great power, and engaged in extremes. The credibility of a threat made either by the U.S. or the Soviet Union within the context of the amassment of nuclear arms would be relatively stable, so

---

[25] Specifically, Chow recommends the categorization of space weapons into six groups with each subject to a tailored proximity restriction:

1. Space-based less than 10 km range ASAT- proximity demarcation to be determined by the DOD and the international community
2. Space-based less than 10 km range defensive weapons
3. Space based 10 km or more range ASAT
4. Ground-launch ASAT
5. Space-based weapons against terrestrial targets (Chow, 2018, p. 126)

long as neither side claimed a significantly large advantage over the other, and if both reserved the commitment to exact heavy costs if the other carried out a first strike.

Nuclear Posture

Trends that indicate the amassment of nuclear power continues to manifest today, signaling the persistence of nuclear deterrence, as well as the fact that deterrence as a strategy continues to be upheld by the status quo. The United States' stated nuclear posture asserts that "..our nuclear posture is intended to complicate an adversary's entire decision calculus, including whether to instigate a crisis, initiate armed conflict, conduct strategic attacks using non-nuclear capabilities, or escalate to the use of nuclear weapons on any scale (DoD NPR, 2022, p. 9)." Such an approach necessitates the rapid modernization of the U.S.'s nuclear forces, command, control, and communications (NC3) systems, in an effort to integrate deterrence strategies across Joint Forces, as well as have available, tailored options in order to retain adaptability within a rapidly evolving international environment (DoD NPR, 2022, p. 2).

Given this, it is reasonable to conclude that nuclear deterrence should persist as a viable strategy. Take for example the Treaty on the Non-Proliferation of Nuclear Weapons, which entered into force in 1970. Despite requiring the non-proliferation of nuclear weapons, there are still states not party to the Treaty including Israel, India, Pakistan, and the Democratic People's Republic of Korea (DPRK) that have since acquired nuclear capabilities (Black-Branch, 2020, p. 328). Additionally, since the Treaty's ratification half a century ago, there has been little sign of states bound to the Treaty drafting legislation that calls for the abolition of nuclear weapons, as well as lack of any complete and verifiable call for disarmament (Black-Branch, 2020, p. 329).

Members of the Big Five (P5): the United States, Russia, the United Kingdom, France, and China, continue to preserve their nuclear capabilities even though Article VI of the NPT explicitly requires states to undertake measures "…relating to cessation of the nuclear arms race…and to nuclear disarmament…and on a treaty on general and complete disarmament under strict international control (Treaty on the Non-Proliferation of Nuclear Weapons, 1968)." The United States for example, maintains the extension of its nuclear umbrella protection over its allies; most prominently through the North Atlantic Treaty Organization (NATO). As evidenced in both the 2018 and 2022 NPR, the need for extended nuclear deterrence to European allies- primarily against Russia in the Euro-Atlantic theater remains strong; especially in light of Russia's invasion of Ukraine that began earlier last year (DoD, 2018, p. 35).

In an address given on September 21st, 2022, after the annexation of Ukraine's four territories of Donetsk, Luhansk, Kherson, and Zaporizhzhia, Putin declared that in an effort to defend the territorial integrity and sovereignty of Russia, he would "..certainly make use of all weapon systems available to us," adding, that "… this is not a bluff (Vladimir, 2022, para. 33)." This particular transcript was taken from the Official Internet Resources of the President of Russia, an official Russian website managed by the Presidential Executive Office, and stands in somewhat of a contrast to a transcript taken from Reuters in which Putin is said to have declared that "we will defend our land with all the powers and means at our disposal (Reuters, 2022, para. 5)." While not substantial, the minute differences in wording between the two transcripts is a nod to the way states transmit signals, and the way others perceive them. The implicit intent behind the statement is nonetheless fairly consistent across both transcripts: that Russia is prepared escalate to the use of nuclear options if deemed necessary. Putin's short statement was alarming for Western analysts since the willingness to escalate in itself represents a threshold that none have crossed since the bombing of Hiroshima and Nagasaki. This particular instance is also an interesting case study on the efficacy of the nuclear taboo, a phenomenon hailed by proponents of nuclear deterrence as one of the most potent reasons for why nuclear weapons have not been used since 1945.

Another state antagonistic the United States is China; and while it has not yet demonstrated its nuclear capabilities to the extent that Russia has, it is still developing capabilities that point towards a growing nuclear arsenal with a great degree of resilience and effectiveness (DoD NPR, 2022, p. 4). In fact, the 2022 NPR notes that by the 2030's the U.S. will, 'for the first time in history', expect to face two major nuclear powers as strategic competitors (DoD NPR, 2022, p. 4). This potential challenge to the U.S. from both across the Pacific and Atlantic, could create a 'thinning' effect, where the U.S. must rely more strongly on strategic partnerships within each respective region so as to avoid directing an insufficient amount of attention to either side. During former President Trump's tenure, the Indo-Pacific saw a visible decline in regional stability; for rather than advocating for a more comprehensive, and unified collective policy to strengthen a Free and Open Indo-Pacific (FOIP), the Trump administration frequently alienated and berated allies over issues of host nation support costs and trade (Ford, 2020, para. 4). In preparing to defend against two major powers, the U.S. must ensure that its strategic alliances are fortified; not limited to increasing interoperability, intelligence sharing, and denial capability, enhancing early indicator and warning systems, as well as streamlining institutional barriers to collective R&D and cooperation (DoD NDS, 2022, p.14).

Thus, in light of new developments in technology and the dynamic of the international environment, nuclear deterrence remains an invaluable tool in the strategic toolkit. However it should no longer be expected to effectively deter unwanted aggressions alone, given the expansion in the types of capabilities now in play.

Nuclear Power in Outer Space

The technological innovations of recent times has significantly altered the international environment such that bipolarity has evolved into multipolarity. A host of different actors are carrying out tactics that seek to gain a strategic advantage in a world less cemented in the status quo of the past; a result of an increased amount of varying and cheaper methods to engage in conflict. Gray zone activities for example, fall below the perceived thresholds of military action, and incorporate strategies such as cyber and space operations, economic coercion, disinformation, as well as the use of proxy forces (DoD NDS, 2022, p. 6). Such tactics account for an increasing percentage of activities conducted by states against the United States. Yet, despite the rapid expansion of more ambiguous tactics and newer technologies, the significance of the nuclear option remains a staple of the U.S. military institution. Consequently, it remains clear that the U.S. continues to look towards nuclear power as a deterrent in spite of any moral qualms to its contrary. Strategic competition coming from China, Russia, as well as other states such as North Korea and Iran, who have all continued to bolster their nuclear weapons capabilities, provide enough of a justification for maintaining the nuclear option as a viable and necessary tool of deterrence (DoD NDS, 2022, p. 4). The 2022 NPR further articulates this point in its statement: “for the foreseeable future, nuclear weapons will continue to provide unique deterrence effects that no other element of U.S. military power can replace (DoD NPR, 2022, p. 1)."

An interesting change within the 2022 NDS that separates it from its predecessors is the use of the term 'deterrence by resilience'. While the 2018 NPR and 2021 Interim National Security Strategic Guidance both refer to 'resilience' in terms of either resilient nuclear weapons infrastructure, or resiliency against the harmful effects of climate change, the 2022 NDS takes it a step further by defining and differentiating the term entirely. Deterrence by resilience operates chiefly through the ability to "…withstand, fight through, and recover quickly from disruption (DoD, 2022, p. 8)." This effectively demonstrates the need to adapt to a rapidly changing environment that now sees multi-domain attacks against critical networks and infrastructure. The 2022 NDS goes on to declare:

> Because the cyber and space domains empower the entire Joint Force, we will prioritize building resilience in these areas. Cyber resilience will be enhanced by, for example, modern encryption and a zero-trust architecture. In the space domain, the Department will reduce adversary incentives for early attack by fielding diverse, resilient, and redundant satellite constellations (DoD, 2022, p. 8).

In recognizing that the space domain is becoming increasingly more prominent, the expansion of approaches to deterrence in kind represent an augmented deterrence strategy as opposed to an outright shift in focus from nuclear to space. Thus, the emergence of deterrence by resilience serves to supplement the existing architecture of the U.S.'s nuclear posture. This type of approach reflects the direction in which conflict in outer space is moving towards: just as nuclear deterrence *alone* can no longer effectively serve as a deterrence mechanism, inferential anti-satellite capabilities by themselves cannot function as a primary deterrent strategy in outer space, unless reinforced by some other means. Deterrence by resiliency can thus be considered to be supplementary reinforcement mechanism that guards against gray zone warfare and salami tactics in the newest theater of conflict.

Outer space presents some challenges to the nuclear deterrence doctrine. Issues are not limited to dual-use satellites, orbital debris, lack of governance and established behaviors, nor the inadequacy of the U.S.'s military posture to manage the massive volume of reversible ASAT attacks. It is still unclear as to what extent these types of salami tactics, on a case by case basis and rebranded to the space domain, represent increases in escalation. Though, one thing is for certain: the tremendous amount of reversible attacks taken as an aggregate collection of aggressive activity could result in at least *some* level of escalation. Forrest Morgan (2010) concurs, arguing that it is entirely possible for an aggressor to conduct revisable-effects attacks on space based intelligence, surveillance, and reconnaissance as a means to degrade an defender's capacity to respond to crisis without incurring a significant retaliatory response (Morgan, 2010, p. x). Furthermore, the nature of such tactics diminish the ability of a state to perceive intention, thus contributing to an increased probability of deterrence failure through misinterpretation.

So far, the application of nuclear power to outer space has been limited to nuclear fission reactors, which serve to power explorations into space that require greater volumes of energy than a typical fuel cells like those used in NASA's space shuttle (Anderson, 2015, para. 2). Thankfully, a nuclear fission reactor has yet to explode in space- either on purpose or by accident. However, if such an explosion were to take place, we could surmise that it would have a similar effect to that of the high-altitude bomb tests that took place during the 1960s. During the height of the Cold War, the United States launched a 1.4 megaton nuclear bomb into space (roughly 250 miles above the surface of Earth), in a test named Starfish Prime. The subsequent explosion released a powerful wave of light, heat, and dangerous, high-intensity radiation (e.g. X-rays and gamma rays), causing black-outs and inoperable communications networks (Staughton, 2016, para. 7-8). The Earth's magnetic field absorbed ionized radiation from the explosion, creating an artificial radiation belt that was stronger and much longer lasting than scientists had predicted. This "Starfish belt", lasted for at least 10 years, destroying Telstar 1- the first satellite to broadcast a live television signal, and Ariel-1, Britain's first satellite (Gutierrez, 2021, para. 16). As compared with the state of the world in 1962 when Starfish Prime was detonated, our modern society is highly dependent on computer chips and sources of electrical power; thus if in the unlikely event that a nuclear bomb detonates in outer space, its consequences would be comparatively more negative, and far-reaching.

To mitigate the possibility of the extension of nuclear conflict into outer space, treaties such as the Partial Test Ban Treaty (1963) and the Outer Space Treaty (1967) were ratified. The latter, which is formally known as the 1967 Treaty on the Principles Governing the Activities of States in the Exploration and Use of Outer Space OST), prohibits states party to its treaty in the use of nuclear weapons in outer space, as outlined in Article IV:

> states party to the treaty undertake not to place in orbit around the Earth any objects carrying nuclear weapons or any other kinds of weapons of mass destruction in orbit, install such weapons on celestial bodies, or station such weapons in outer space or in any other manner (United Nations, 1967).[26]

[26] 110 states have ratified the OST, and 89 have signed it but have yet to complete ratification (Kimball, 2020, para. 2).

While the OST is widely recognized as providing the basic framework on international space law, one of its major shortcomings is that it does not specify the prohibition of any other types of weapons in outer space. In this regard, many states argue that the OST is insufficient for safeguarding outer space as the common heritage of mankind, as it so advocates within its articles.

While one could argue that a nuclear weapons system doesn't necessarily need to be placed in outer space to still function as an effective deterrent, there are still alternative methods to circumvent the Treaty's provisions. For example, while the OST prohibits the *placement* of nuclear weapons in space, it does not explicitly outlaw an object that is nuclear in nature to *pass through* orbit. A co-orbital strategy such as the one implemented by Russia during the development of its anti-satellite program in the 1960s would clearly violate the OST as written, since it functions by arming and launching a weapon with conventional explosives (or in this scenario, nuclear weapons) into the same orbit as the target satellite. When the weapon approaches ten meters from the destination, it detonates, destroying the target satellite (Grego, 2012, p. 3).

A direct ascent strategy on the other hand, could theoretically pass through Earth's orbit and at some point during that journey, detonate. Of course, this 'loophole' does depend greatly on the definition of what constitutes when an object may be considered to be placed in orbit, and as well as the definition of 'orbit'. The OST also does not explicitly ban any other weapon types besides nuclear and mass destruction, opening up the possibility for satellites to be modified or repurposed for uses other than communications, science, navigation, etc. In other words, the primary deficiency of the

OST is that it’s scope is much too narrow to be considered effective in today’s international system where new advances in technology have generated new types of satellite capabilities.

Besides the international, legal restriction of physically placing nuclear weapons in space (which is of course, limited only to states party to the treaty), there are other reasons why nuclear deterrence, and deterrence in general, would not function successfully in the outer space domain by means of inferential anti-satellite capabilities. It seems apparent that, based on the principles of deterrence, it is not possible for a state to enhance its kinetic-physical ASAT capabilities in outer space without escalating to higher levels of conflict, since it would generally perceived as a build-up of armaments in preparation for an attack. The effective, successful implementation of deterrence in outer space thus requires strategically navigating credibility, as well as supplementing it where it is found lacking.

Credibility and Signaling Strategies in Outer Space

With the security of infrastructures placed in outer space being so critical to U.S. national security interests, and with the evolution of offensive capabilities within the international environment, the need to redefine the parameters of deterrence within a new context has been set to the forefront of national security priorities. The vast proliferation of space-based weaponry by state and non-state actors alike serve to further drive up the impermanence of power dynamics between relevant stakeholders. As noted earlier, it is now easier and more economically feasible for third-party actors including terrorists to acquire satellite technology in order to compete with, as well as threaten other actors traditionally considered to rank higher within the status quo. The idea of outer space becoming an extension of the nuclear arms race, put in this context, does not seem wholly unreasonable.

As it currently stands, deterrence is not equipped to explain or predict decisions to cooperate, engage in, or defer from conflict in outer space, since the range of inferential anti-satellite capabilities create challenges that wrest away the ability for states to establish credibility. This is not to say that deterrence will be become completely defunct. Considering the political heft with which classical deterrence governed the actions of the U.S. during the Cold War era, it is not outlandish to presume that the implications of a space deterrence doctrine should be equally impactful. Only that it must adapt in light of the increasing inefficacy of its core tenets.

Credibility and signaling can generally go awry in several ways including, but not limited to signal misinterpretation, blatant disregard of signal (bluffing), and lack of perception. The consequence of degraded signals are 1) the degradation of credibility, and 2) the inability to formulate an appropriate level of response. This of course, could cause a situation to escalate the conflict to dangerous and unwanted levels. One such possible solution would be to focus on the requisite thresholds for each tenet. In identifying appropriate, universally accepted thresholds for different satellite capabilities, the issue of perception could theoretically be mitigated. Any retaliatory threat would thereby be considered to be credible in *all* situations, allowing for more transparency on the part of both aggressor and defender to either escalate or capitulate. This line of reasoning agrees with Arce and Sandler's finding that in game modeling of PD, credibility always holds (2009, p.405).

However, the drawbacks of this approach can be as equally debilitating to credibility. Having an element of ambiguity in terms of whether or not a threat will be deployed is vital for the successful operation of deterrence: we only send signals we want the other side to know, we take into consideration whether or not those signals will be interpreted in the way we intend, we carefully craft a psychological environment for our opponent such that they are deterred from taking action against us. If this element is weakened by establishing that all retaliatory threats are credible, it would enable more consistent policy going forward and negate the possibility of communicated signals to be misinterpreted. However if an opponent knows this, they understand that they can at the very least, launch a debilitating first strike-as they are undeterred in their initial attack.

The concept of a structural escalatory ladder was the brainchild of Herman Khan (1965), who rejected the idea that 'war just happens.'[27] According to Khan, escalation is "…an increase in the level of conflict in international crisis situations," using a ladder metaphor for structurally representing the increasing rungs of conflict beginning with the 1^{st} and ending in either the 16^{th} or 44^{th} rung, with the later representing a generalized or abstract scenario (Davis & Stan, 1984, p 4-6). The purported advantage of such a structural approach to conflict analysis is its ability to clearly delineate each level of conflict, which effectively suggests a clear delineation of conflict thresholds. Kahn asserted that this enabled scholars to focus more rigorously on 'thinking about the unthinkable'. However, there are several problems with this approach that Davis and Stan address including it embodying a Western view of reality, describing conflict as one-

[27] This is in opposition to Schelling's concept of escalation, which took on the form of compellence, or the use of bargaining power to heighten levels of conflict by raising its stakes (P. K. Davis & Stan, 1984, p. 4).

dimensional, failing to make operational distinctions (i.e. differences between what the Soviet Union considers operational-tactical and operational-strategic uses of weapons), forcing a rank-order construct of conflict, and disregarding the intricate tempo of war (Davis & Stan, 1984, p. 4-7). However, Davis and Stan overlook another critical weakness in Khan's structural approach to conflict: that its rigidity provides an opportunity for adversaries to take advantage of knowing how an incumbent state may act based upon a given action. If for example, state B knows that state A will only choose to escalate conflict in outer space when state B's assets encroach a within a 0.2 degree radius of state A's assets, state B may choose to place its assets just outside of that radius just enough to pose a substantial threat to state A yet remain below the escalatory threshold for state A's retaliatory actions. This scenario assumes that state A's actions will hold true to its escalatory ladder, and that the rungs of the ladder are adequately conveyed to state B. Thus a structural approach to escalation may work against an incumbent state since deviating from it could be seen as incredible, and adhering to it may compromise critical infrastructure.

Another potential solution to the credibility and signaling problem of inferential ASAT capabilities would be to form a strong alliance mechanism through a defense pact or publicly stated alliance commitment, preferably with a NWS. This could also be in conjunction with the placement of a benefactor state's nuclear weapons on the beneficiary state's territory; albeit the advantages of the later may not be enough to warrant its implementation.[28] As discussed in the previous chapter, such an approach could be a significant supplementary means to bolster the credibility of inferential ASAT.

[28] For a more detailed explanation of why 'deeds' such as the placement of nuclear weapons on an ally's territory would not garner a significant advantage, refer to Fuhrmann and Sechser (2014).

Perhaps the only meaningful disadvantage to this approach would be the fact that it does not address the issue of reversible-effect attacks, which fall below the threshold for escalation. During the Cold War, grey zone warfare and salami tactics were also left unaddressed by strategies such as massive retaliation. This in turn gave way to flexible response, which called for a 'triad' of conventional, tactical nuclear, and strategic nuclear weapons.

In expanding the array of options available, the U.S. was strategically better equipped to deal with salami tactics since it also meant that the U.S. could respond more proportionately as well as more effectively to acts of aggression of varying levels. Flexible response not only created a more practical means to defend European Allies, but also enabled a shift in focus from assured destruction to "…countering infiltrations, incursions or hostile local actions without recourse to nuclear weapons (Freedman & Michaels, 2019, p. 383)." The challenge that salami tactics pose are not dissimilar to those presented by reversible-effect attacks. Considering that decisionmakers during the Cold War opted for a rebalancing of strategies, it is reasonable to conclude that a rebalancing may provide an effective means with which to address salami tactics in the space domain. This could include implementing a treaty mechanism along with a new 'triad' of SOF (special operations forces), space, and cyber forces that can provide an integrated defense to the threats ASAT pose (Howe, 2022, para. 1).

In addition to the issues associated with inferential capabilities, so too rests the conundrum of attribution, or the ability to identify who conducted an attack. It is extremely difficult to deter a threat, or even respond in kind when its origin is unknown or uncertain. Jamming or spoofing for example, interferes with satellite signals on commonly used frequencies such as GPS, to disrupt military or civilian satellite communication (Grego, 2012). Furthermore, these types of attacks closely resemble unintentional interference, making it nearly impossible to differentiate an attack from a benign occurrence. Firth notes that the U.S. Defense Intelligence Agency (DIA) occasionally jams its own communications satellites by accident, which only serves to emphasize the level of difficulty in attribution (2019; 38).

In consideration of each of the factors addressed, the question now shifts from whether or not an aggressor can be deterred to what kinds of attacks against which capabilities could be deterred, and under what circumstances (Morgan, 2010, p. x). Attacks that do not cause permanent damage to space systems such as dazzling and jamming warrant lower levels of response (alternatively phrased, proportionately lower levels of punishment through deterrence), than would attacks that could cause irreversible damage, as one would expect of a kinetic-physical attack. Along the same lines of reasoning, non-kinetic physical capabilities, in addition to dazzling and jamming, EMP's, HPM's, and neutral particle beams, that do not produce orbital debris would not provoke the same level of response as one that would discharge large amounts of debris throughout the space environment (Morgan, 2010, p. 17).

Morgan offers a general, preliminary framework for how an escalatory ladder may operate based on retaliatory actions, characterized by satellite type: SEWS (satellite early warning system), ISR (intelligence, surveillance, and reconnaissance), Weather, PNT (positioning, navigation, and timing), MILSATCOM, and SATCOM (satellite communication) (2010, p. x). In a conflict scenario, adversaries may prefer to target relatively low-threshold satellites such as GPS in the beginning of conflict by employing reversible-effects attacks, then as higher levels of conflict are breached, move up the escalatory ladder to more valuable satellite infrastructure such as ISR, SATCOM, and MILSATCOM. On the other hand, at lower levels of conflict, adversaries may opt for attacks limited in scope for fear of escalation, especially if reversible-effects attacks continue to yield benefits that, even if marginally, outweigh the costs of implementation.

While Morgan's template provides a good preliminary framework for an escalatory ladder, he does not take into consideration retaliatory threats as opposed to deterrent threats that obviate the need for the former; nor does he give attention to the weight costly signaling plays into this cost-benefit calculus. Ideally, an escalatory ladder would be categorized by the perceptibility of a deterrent threat, and then further sub-categorized by the type of satellite, as Morgan did, with a suggested proportionate response (defender's retaliatory threat). Furthermore, a graduated escalatory framework would benefit by taking into account the effects of supplemental mechanisms such as the declaration of a formal alliance.

In the next section, this thesis will propose an alternative framework that attempts to take into consideration the variability of perception as a result of inferential satellite capabilities, as well as incorporate into its calculus, a supplementary treaty mechanism that could theoretically boost deterrent thresholds in order to provide some 'cushion' room for conflict situations; potentially offering a more robust signaling mechanism for states to demonstrate their intent to deter an unwanted action. This type of mechanism could satisfy one of Zagare's propositions for resolving theoretical and empirical contradictions. It is also consistent with another promising line of research put forward by Fuhrmann & Sechser, which indicates that alliances can be a powerful means for signaling intent (2014, p. 919-920). Treaties also create a legally binding support mechanism that can in itself signal to potential adversaries that retaliatory measure will be taken if an article to the treaty has been violated, and can be made stronger through the effects of hand-tying. [29]

Concluding Remarks on Credibility

In an explosive, technological growth environment, it has become necessary to reexamine deterrence theory and its applications to outer space, especially when its most basic principles are being called into question. Factors such as the nature and type of actor, power dynamics, and motives that contribute to the degree of fluidity in the international environment affect a state's ability to signal credible deterrence threats (Koichi, 2016, p.25-26). Moreover, the largely inferential nature of non-kinetic physical, electronic, and cyber ASAT engender new customary norms that are now in the process of forming. In effect, the tenants of classical deterrence are being supplanted with concepts that are more tailored to the evolving space domain.[30] It is because of these emergent characteristics that the traditional doctrine is becoming increasingly ineffective as a mechanism to predict and model situations of interstate tension. States are now confronted with questions stemming from unaccounted discrepancies within international relations deterrence theory: how can a state utilize strategies of deterrence when the tools with which to implement them cannot inherently be perceived? And, what there to stop

[29] Research by Fuhrmann and Sechser has shown that alliances with nuclear states in particular seem to be able to deter violent conflict (2014; 932). This is addressed later in this discussion.
[30] These concepts refer to principles that are in direct contrast to classical deterrence. They do not require mutual dependence based in nuclear stability (in other words, a dyadic relationship), are not restricted to the highest levels of conflict (MAD is often referenced as the highest level of conflict), and can potentially withstand manifold power relations (no longer limited to state-actors).

an adversary in their tracks if they do not perceive any impending or ongoing use of force? In addressing these questions, developing a theoretical mechanism more suited for the use of inferred weapons capabilities in outer space is a pressing matter. Regarding international relations as a discipline, Bueno de Mesquita has argued:

> "…for too long the divorce of explicit, rigorously derived theory from explicit, rigorously derived evidence has hampered the study of international conflict…nearly twenty-five centuries after Thucydides wrote *The History of the Peloponnesian War*, we continue to debate whether a "balance of power" is good or bad, whether alliances are entangling or liberating, whether great powers behave differently from lesser powers, whether decision making patterns very markedly or are essentially the same in different centuries or in different cultures (Bueno de Mesquita, 1981, x)."

Other scholars have stated similarly that the discipline of international relations has lacked empirical support for theoretical concepts, with deterrence theory being no exception. All this to say that an emphasis on the marriage between theory and empirical support is long overdue. This discussion on deterrence theory in the context of outer space is not only crucial for its theoretical implications, but because it has the potential to inform national security strategies and shape the governance of ASAT in outer space as well as outer space jurisprudence.

Chapter V.

Game Theoretic Modeling

The study of rational choice and game-theoretic models of decision making which aim to empirically study deterrence, were borne in response to the tradition of examining case studies of war through lexical argumentation, or in-depth historical and theoretical analysis. Investigating the minute details of various events leading up to war- each harboring its own idiosyncratic causes and consequences, do little for the analysis of future conflicts and can often be proven invalid by examining other lexical case studies. Hobson's theory of imperialism, Morgenthau's interpretations of the balance of power, Organski's power transition, and Waltz's theory of polarities, are reflective of such extensive theoretical examinations (Bueno de Mesquita, 1981, p. 3). However, few such approaches have established a general hypothesis that serves to explain the nature and causes of war. This suggests that the lessons learned, while thought-provoking and perhaps useful for the study of history, fail to distinguish between recurring patterns and accidental happenings- a distinction without which one cannot hope to craft a general theory of war. Since deterrence is a facet of war, it goes without saying that such considerations are significant for its study as well.

Assumptions about state actors, motivations, and power can provide the grounds for an empirical analysis of rational decision making. As such, they lay out the conditions under which the deductions derived from it are necessarily true. In other words, assumptions are required in the rational choice approach because they always generate valid conclusions. However, the formation of assumptions requires the differentiation between what factors may be considered sufficient from those that are necessary (Bueno de Mesquita, 1981, p. 4). Since the initiation of armed conflict does not begin suddenly, it is within the buildup of tensions which form the context leading up to a war that a set of necessary conditions can be found and which must be satisfied in order for war to occur. Yet the sheer volume of conditions, necessary or not, may be so manifold and interact in such complex ways that distinguishing one from another becomes impossible to do. Much in the way that the social and political worlds are similarly complex, the distinction of assumptions must arise out of a meticulous and exact methodology such that the deductions that follow have no room for alternative interpretations. This is precisely what game theoretic modeling aims to accomplish; it allows the modeler to limit the scope of analysis thereby identifying necessary conditions for war and producing empirically-backed deductions.

Critics that question this method of empirical examination cite the inconsistency of employing a simplistic, formal model to capture the unique intricacies of the real world. However, it is in this simplicity that rigor and precision required by formal modeling come together to oblige the modeler to decide in an exact manner, what the assumptions of the argument are. Contrasted with lexical arguments, which can be interpreted in any number of different ways, the transition to modeling provides an opportunity to expose implicit assumptions in lexical arguments. Without formal modeling, it would be difficult to readily accept any given hypothesis as conceptually sound (Morrow, 2020, p. 6-7).

This chapter will begin with a discussion on the rational choice approach to formal modeling, then proceed to a brief overview of game theoretic modeling, including extensive and strategic form games, as well as examine the advantages and disadvantages of expected utility theory. Finally, this chapter will present a modified version of Bueno de Mesquita’s expected utility equation and test it against hypothetical cases that involve armed conflict in outer space, as well as historical cases that to an certain extent, that have similar components to what could result in such conflict.

The Rational Choice Approach

One way to formally model strategic situations is through rational choice approach to game theoretic modeling, which provides a method to formalize social structures. It explains how an actor's decisions are interrelated, how decisions are drawn to arrive at specific outcomes, and explores the effects of structure on individual decisions (Morrow, 2020, p. 1).[31] In terms of modeling, the rational approach assumes that actors have goals which they make an effort to achieve, that they have some degree of choice in making these decisions, and that they choose actions based upon whether or not they believe carrying out those actions will help achieve their goals. Moreover, the rational approach can be expressed in several ways; the first through extensive or strategic forms through game-theoretic models, and the second through mathematical equations or inequalities which are often used when testing expected utility theories.

Efforts to develop the rational choice model and empirically test deterrence strategies include Schelling (1960,1966), Snyder (1961), Ellsberg (1961), along with later works by Zagare (1987), Brams (1985), Nalebuff (1991), Powell (1990), Wagner (1982,1991), and Fearon (1994b) (Huth, 1999, p. 28). Bruce Russett's 1963 article "The Calculus of Deterrence", which examined nine factors for enhancing credibility, and

---

[31] Social scientists refer to structure as the social settings that influence an individual's choices. It can range from the factors that generate the outcomes of decisions (e.g. military capabilities), to whether an individual believes that they have choices (Morrow, 2020, p.1).

Peter DeLeon et. al.'s study entitled "Situational Analysis in International Politics" are examples of early efforts to develop a more robust empirical methodology (George & Smoke, 1974, p. 88-90).[32] Some of its early theoretical developments and limitations can be summarized below:[33]

> Brams and Kilgour (1988) have made a large contribution based on the game "chicken," but they do not fully resolve the credibility question. Powell (1987,1988,1989) successfully deals with the credibility problem, relying on a modified "chicken" and incomplete information, but his dynamic picture is somewhat contrived. Zagare (1985) and Langlois (1989) partially succeed in those two aspects but at the cost of using a game structure (the prisoner's dilemma) that failed to model the risk of losing control (Langlois, 1991, p. 802).

These studies did not take into account a variety of decision-making variables that are reflective of the complex, theoretical calculus of deterrence. Moreover, since these early studies vary widely in the degree and method in which they evaluate deterrence questions, they could not be directly compared to each other despite employing the rational-choice approach (George & Smoke, 1974, p.93). This suggests that methodologies for early rational choice game theoretic modeling were not yet sufficient for developing a general, coherent hypothesis of war.

In deterrence literature, the modeling of rational choice strategies often take the form of expected utility or two-by-two games (Morrow, 1985, p. 474). Utility theory,

---

[32] The results of these early studies were later determined to have some limitations. Based on his study using an arithmetic correlation technique, Russett concluded that out of the nine credibility factors, only economic interdependence (measured as a percentage of the pawn's imports provided by the defender) exhibited a strong correlation with successful deterrence. However, Russett himself acknowledged his study's limitations when he later stated that he made two substantial inferential leaps- one in terms of the existence of a relationship between defender and pawn, and the other in terms of their perception that their actions would be accounted for in the opponents calculations. DeLeon et. al.'s study on the other hand, employed a methodology (statistical-correlative) that could not capture the alterable nature of intervening variables (George & Smoke, 1974, p. 90-91).

[33] These early theoretical developments were largely the product of studies that promulgated from Brams' 1985 study which introduced the use of Chicken as a formal model to examine the Cuban Missile Crisis (Langlois, 1991, p. 803).

which is a derivative of game theory, is a mathematical method for representing decisions and its processes undertaken by these actors. Part of the aim of this calculus is to discern an actor's expected utility, or its preferences over the outcomes that represent their willingness to undertake risks in order to achieve desired outcomes, and is calculated by multiplying the utility of every potential outcome by the probability that it will occur if chosen, and then finally summing across all possible outcomes (Morrow, 2020, p.16). The result is such that actions with larger expected utilities are preferred over ones with lower expected utility. In order to construct an expected utility model, it is necessary to delineate "…the possible states of the world, the choices available to the actor, the utility function for each possible state, and the (subjective) probability distribution for the occurrence of each state of the world for each available choice (Morrow, 1985, p. 476)."

In the rational choice school of thought, the decision to refrain from or enter conflict hinges on this calculus. A potential aggressor will weigh the gains to be secured by the use of military force against its associated costs, and evaluates the likelihood that such force can be applied successfully. A high expected utility of initiating war compared to a low expected utility of maintaining the status quo means that it is more beneficial for an actor to initiate war despite its costs, than it is to abstain from action. Thus, the most effective strategies of deterrence are those that result in the decrease of the expected utility of using force while simultaneously not reducing the expected utility of the status quo. This is not to confuse the misconception that the values of expected utilities dictate a decisionmaker's preferences. Rather, preferences are used to construct utility values (Morrow, 2020, p. 33).[34]

---

[34] Utilities are values that represent an actor's preferences over actions, and they can be constructed such that they are consistent with observed behavior (Morrow, 2020, p. 33).

Figure 3: Quackenbush's Classical Deterrence Game (Extensive Form)

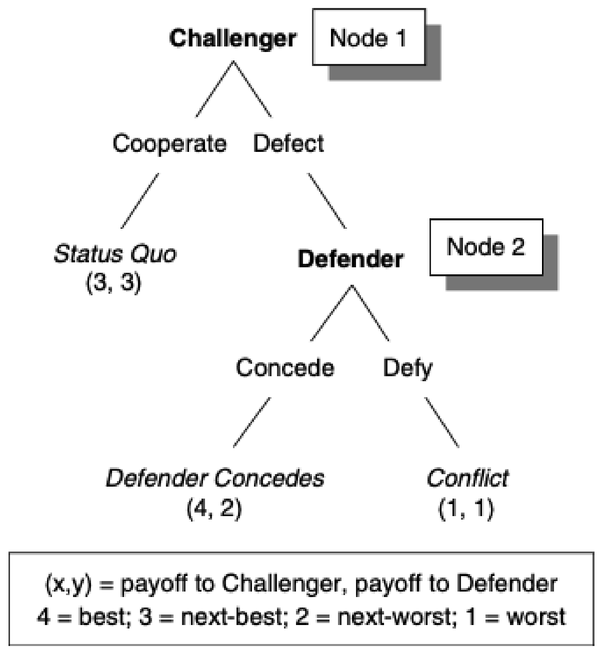

Figure 3 is an extensive form representation of a bilateral deterrence scenario (Quackenbush, 2011, p. 7)

Existing game theory in the rational choice approach such as expected utility modeling have traditionally used Chicken and Prisoner's Dilemma for empirical analysis of deterrence.[35] In typical game models, such as the one presented in Figure 3, the

[35] In Chicken, each player can choose between two strategies: to cooperate and to not cooperate (within deterrence literature this is modified to attack and submit). There are four possible resulting outcomes which are ranked from best to worst in terms of player preference: 1) both players cooperate, 2) one player cooperates and the other does not; which is the best outcome for the player who does not cooperate and the next-worst outcome for the player who does, and 3) bother players do not cooperate (Brams & Kilgour, 1988, p. 5). The Prisoner's Dilemma on the other hand is most often represented by the situation wherein two parties that are separated and cannot communicate, must choose between cooperating with each other or not. The highest pay out for any given player occurs when both parties choose to cooperate. The Prisoner's Dilemma likewise has a set of ordered preferences which determine the parties best course of action; i.e. the one that results in the best possible payout which are as follows: 1) mutual cooperation, 2) unilateral cooperation, 3) temptation from unilateral defection, and 4) punishment for mutual defection (Arce & Sandler, 2009, p. 387-388).

challenger- who's objective is to change the status quo, begins the game by deciding whether or not to attack or cooperate. It is the objective of the defender to deter the challenges posed by its counterpart and can respond by either conceding, or defying which results in the escalation to general conflict. The game outcomes are as follows: maintain the status quo, defender concedes, and conflict (in deterrence literature, this often denotes general nuclear conflict or war) (Quackenbush, 2011, p. 6).[36] Figure 3 also delineates a situation wherein B maintains a credible threat. Using backwards induction, we begin at Node 2 and deduce that B will choose to defy since conceding is a less preferable outcome (i.e. B receives a payout of 1 for conceding, and 2 for defying (conflict)).[37] A, knowing that B will choose to defy, A will opt to cooperate at Node 1 since A's payout for cooperation (3) is more preferable than if A were to Defect (2). From this, we are able to conclude that B has a credible threat and will therefore be successful in deterring A (Quackenbush, 2011, p. 9).

Both Chicken and the Prisoner's Dilemma are acknowledged as partially correct but nonetheless incomplete in terms of modelling an event such as a nuclear crisis (Langlois, 1991, p. 803-805). In a two-by-two outcome table for example, generalized nuclear war is the result of the most aggressive choices by the two competing parties; deciding between launching an attack or submitting. However in a real-world situation, strategies are made up of a series of actions and do not rely on a binary decision making

---

[36] Note that in strategic form games, it is common practice to refer to player 1 as the challenger and is addressed with masculine pronouns, while player 2 is the defender and is addressed with feminine pronouns. This is not intended to offend either or any gender by any means. The main purpose for this distinction is to simplify the discussion and to convey it in as a precise manner as possible.
[37] Backwards induction is a game theoretic strategy that is used to solve games. It begins at decisions located at the terminal nodes of the game, choosing the action that maximizes the utility of the player choosing at that node, and then working backwards sequentially through earlier nodes of the game. We can determine all player's optimal choices using this method (Morrow, 2020, p. 124).

system; a strategy may first involve conducting an air strike, but later escalate to the point where a limited nuclear strike is needed. It is thus difficult to believe that, within the parameters of the game which allows a decisionmaker to choose generalized nuclear war, a state would prefer to be the victim of a one-sided nuclear conflict, where the aggressor chooses nuclear war and the defender chooses to unconditionally submit.

An important difference between Chicken and the Prisoner's Dilemma is the ordering of preferences. In the former, defeat is preferred to nuclear war, whereas the opposite holds true in the latter. In other words, the Prisoner's Dilemma assumes that actions taken to escalate or retaliate are always the superior moves; thus a state is always going to be better off escalating than submitting. The representative inequalities of these preferences can be found at the bottom of each model in Figure 4, which illustrates a two-by-two strategic form game of the Cuban Missile Crisis where nuclear war is the result of the most aggressive choices made by two players.

Figure 4 Langlois' 2x2 Model of the Cuban Missile Crisis (Strategic Form)

The Cuban Missile Crisis as "Chicken"

		Soviet Union (State 2)	
		Withdrawal	Maintenance
United States (State 1)	Naval Blockade	(c_1, c_2) (Compromise)	(d_1, v_2) (Russian Victory)
	Air Strike	(v_1, d_2) (U.S. Victory)	(n_1, n_2) (Nuclear War)

with $v_i > c_i > d_i > n_i$ for $i = 1, 2$

International Crises as a Prisoner's Dilemma

		Soviet Union (State 2)	
		Submit	Attack
United States (State 1)	Submit	(c_1, c_2) (Compromise)	(d_1, v_2) (Soviet Preemption)
	Attack	(v_1, d_2) (U.S. Preemption)	(n_1, n_2) (Nuclear War)

with $v_i > c_i > n_i > d_i$ for $i = 1, 2$

The upper model depicts the Cuban Missile Crisis in the form of Chicken while the bottom model is a representation of the Cuban Missile Crisis through the Prisoner's Dilemma where v= victory (U.S./Soviet preemption), c= compromise, d= defeat, n= nuclear war. In "Chicken" the Soviet Union can be said to have been deterred if it chooses compromise regardless of whether the United States chooses to execute an air strike or a naval blockade. This is because in Chicken, mutual defection, is the worst possible outcome for both players. In the Prisoner's Dilemma however, deterrence is only possible if the Soviet Union elects to submit. If in the face of U.S. preemption, the Soviet Union decides against attacking, the latter's loss from taking no action is considered to be greater than if conflict escalated to nuclear war. In the real world, this could manifest in the preference of mutual destruction as opposed to unilateral destruction. In these types of limited-sequence games where moves are fixed, any threat is assumed to be credible since there exists no options for subsequent moves that could, through the process of backwards induction, render the present move incredible. Critics of Brams and Kilgour's "chicken" model such as Langlois, contend that the preferences associated with Chicken ignore strategies that remain the most desirable in all possible outcomes, since they can still arise even when actions are accidental or irrational because they nevertheless form subgame perfect equilibria (Langlois, 1991, p. 805). [38]

Instead of the limited sequence game of Chicken proposed by Brams and Kilgour, Langlois opts for the use of a continuous game structure that removes the emergence of subgame perfect equilibria, which occur at all levels of crises in the Brams and Kilgour model. His reasoning lies in the notion that in a fixed number of turns that occur in sequencing games, a player may be deterred from what could otherwise be its optimal first move. Thus, if the threat of retaliation is tested by employing a continuous structure, it may be exposed as a disadvantageous move. He notes that "…neither the United States nor the Soviet Union could in advance know exactly when the Cuban Missile Crisis would end," demonstrating that limited round games do not reflect real-world scenarios, which are never bound to a set number (Langlois, 1991, p. 805).

The classical model of deterrence as depicted in Figure 3, the defender faces a paradox: capitulation would provide a more preferable outcome than defection (the choice to escalate conflict). But, if the challenger knows that the defender will choose to

---

[38] Brams and Kilgour's Chicken model is an extended version of the original and introduces the concept of a probalistic threats. This allows the comparison of different levels of threats as well as the analysis of the relationship between preemption, retaliation, and game outcomes (S. Brams & Kilgour, 1985, p. 119).

capitulate, the challenger will always be incentivized to carry out a first strike attack, and deterrence will always fail. Zagare and Kilgour (2000) have named this predicament the ‘paradox of mutual deterrence’ (Quackenbush, 2011, p. 6).

One method for addressing this paradox is by what Schelling calls ‘leaving something to chance’. These types of threats raise the risk of a situation such that all control is lost, leading to an inevitable escalation to thermonuclear conflict. In this regard, a defender is able to rationally choose to simply raise the risk of war, as opposed to rely on an irrational choice for war (Quackenbush, 2011, p. 8). Schelling supports this idea by reasoning that credibility is contingent upon who’s interests are more greatly threatened and who will be more likely to take risks in order to protect those interest. However, an implicit assumption in this reasoning rests on a shaky foundation- that nuclear escalation also invites one’s own destruction. In other words, ‘leaving something to chance’ bears no credibility since a threat that signals one’s one destruction cannot be considered believable (Quackenbush, 2011, p. 8).

The second method for addressing the paradox of mutual deterrence is through a method involving a ‘burning bridges’ strategy. In the proposed solution, the defender would commit to making an irrevocable, hardline strategy. This has the effect of limiting the defender’s options by eliminating the ability to back down or defect even though that would be the preferred alternative (Quackenbush, 2011, p. 7). Such a strategy also carries the effect of strengthening the level of credibility a threat maintains since the defender is forced to a position where it must commit to carrying out its retaliatory threat. If in Figure 3, the defender concedes as opposed to initiating conflict and the challenger is aware of this, then the challenger has no reason to believe that the defender will carry out her threat. Thus, the defender's threat cannot be regarded as credible; it is *incredible*. Conversely, if the challenger knows that the defender prefers conflict over concession, the defender's threat is *credible*. Using backwards induction on Figure 3 allows us to obtain the optimal preferences for both players. If the defender decides to commit to a hardline strategy of attack (in Figure 3, defect), then the challenger must choose between maintaining the status quo (payout of 3), or defecting which results in conflict (payout of 1). Since the payout for cooperating is higher for both the challenger and defender it is the best mutual response, it denotes a Nash equilibrium. Quackenbush notes however, that this method does not result in a Nash equilibrium if the aggressor chooses to attack instead of maintaining the status quo and is thus not subgame perfect. In other words, it would be irrational for the defender to choose to concede after committing to a hardline strategy (Quackenbush, 2011, p. 7).[39] Thus, if the challenger decides to attack, the defender's only rational choice would be concession since the payout for conceding (2) is higher than conflict (1) for the defender. The later sequence cannot be considered a subgame perfect equilibrium because it entails the defender making an irrational choice (Quackenbush, 2011, p. 7).

Applying this concept generally, let's suppose that State A is contemplating whether or not to use ASAT capabilities to incapacitate one of State B's satellites currently in orbit. Assume that if State A chooses to use its ASAT capabilities, it will likely cause significant damage to State B's space exploration efforts as well as ability to operate in outer space. However, such an action on the part of State A has an equal probability of resulting in State B deploying their own ASAT capabilities in retaliation,

---

[39] A Nash equilibrium is a set of strategies comprised of player 1 and player 2's optimal choices such that neither player has an incentive to change its strategy (Morrow, 2020, p. 73).

causing significant damage to Country A's space exploration efforts and space-based capabilities in the future. To model this scenario, we can create a payoff matrix using strategic form with the following options for States A and B:

Table 1: Extended-Form Matrix of the Burning Bridges Strategy

	State B: Use ASAT	State B: Don't Use ASAT
State A: Use ASAT	(-5, -5)	(-3, -1)
State A: Don't Use ASAT	(-1, -3)	(0, 0)

Each payout within the matrix is represented in the format (State A's payout, State B's payout).[40] In this model, the 'burning bridges' strategy refers to the irreversible nature of the decision of whether or not to use ASAT capabilities. Once the decision to employ ASAT is made, the likely result will play out in the form of retaliation, eventually resulting in setbacks for both states' outer space endeavors. From this strategic form matrix, we can deduce that employing ASAT will result in negative payoffs for both countries, and that the most beneficial outcome-and thus most preferable outcome for both, is when they refrain from ASAT use. If State A were to launch a first-strike, it commits to making a credible threat simply because it would be irrational for State A to do so otherwise. Alternatively, if State A decides to launch a first-strike, then subsequently choose to defect after State B chooses to retaliate in response, State A's actions would be considered *irrational* and thus *incredible* because in such a scenario, the credibility of State A's first-strike would be rendered null. In this sense, the Zagare and Kilgour's connection between rationality and credibility is not only consistent within

[40] Note that the values used in this scenario are arbitrary and could be substituted for any other numbers depending on the situation.

informal strategic literature of deterrence, but with game-theoretic literature as well (Zagare & Kilgour, 2000, p. 68).

Credibility in Decision-Theoretic Deterrence

Like the theoretical flaws maintained by structural theorists, the empirical study of deterrence by decision-theoretic deterrence theorists is not without its shortcomings. Most of the theoretical analysis focus on general deterrence while quantitative analysis has focused almost exclusively on immediate deterrence (Quackenbush, 2011, p. 41). Fearon notes "…hypothesis that are valid for general deterrence should appear exactly reversed if we look at cases of immediate deterrence (Fearon, 2002, p. 15)."

The distinction between general and immediate deterrence can be found in the level of conflict each inhabits. Morgan (1983) defines immediate deterrence as a doctrine "that concerns the relationship between opposing states where at least one side is seriously considering an attack while the other is mounting a threat of retaliation in order to prevent it (Morgan, p. 30)." Immediate deterrence relates to decision making in crises and attempts to examine decisionmakers' attempts to prevent crises from escalating into generalized war (Quackenbush, 2011, p. 4). On the other hand, general deterrence "relates to opponents who maintain armed forces to regulate their relationship even though neither is anywhere near mounting an attack (Morgan, 1980, p. 30)." General deterrence in this sense, is much broader and logically proceeds immediate deterrence, because it addresses a dynamic between two states *before* conflict arises, whereas immediate deterrence focuses on the analysis of the crisis itself. Taken from this perspective, general deterrence is more pertinent for the understanding of international conflict than that of immediate deterrence (Quackenbush, 2011, p. 4-5).

In investigating the efficacy of inferential anti-satellite capabilities and potential supplementary mechanisms with which to increase the likelihood of deterrence success, this thesis seeks to focus its empirical study on direct-general and extended-general deterrence. Firstly, general deterrence inherently directs attention at the dynamics of the international system before any great level of conflict is breached. As such, the study of this particular iteration is well suited towards analyzing the impact of inferential ASAT; this is especially so because of the proliferation of reversible-effect attacks which often manifest as salami tactics. Secondly, the study of both direct and extended deterrence apply to the use of ASAT as taken from the perspective of this thesis's approach to the implementation of a treaty mechanism- an intrinsically 'extended' phenomenon due to its association with alliances.

One of the sticking points within decision-theoretic deterrence theory is centered on how one can accurately measure credibility. Despite the substantial amount of weight given to it as a requisite for deterrence within theoretical discussions, the development of credibility within decision-theoretic literature has been neglected. Schelling's work among others, converged around the idea of credibility existing in an opponent's mind, in fostering a strong ability to communicate signals, having the capability to enact threats, and in committing to carrying out such threats. Decision-theoretic literature has traditionally disregarded the need to formally define credibility because it was generally understood as simply whether or not a threat was believed; thus, threats that are not believed must be considered incredible (Zagare & Kilgour, 2000, p. 66). Zagare and Kilgour reference the Eisenhower Administration's threat to levy nuclear retaliation on the Soviet Union for small infractions on the status quo as an example of an incredible threat. However, this example does not fully capture credibility as a belief since it does not specify that the incredulity of the U.S.'s threat was simply due to a failure of the U.S.'s commitment to carry out such an escalatory act. Regardless, Zagare and Kilgour offer an elegant redress to the definition of credibility from a decision-theoretic perspective: they establish a strong correlation between the principles of credibility to rationality by arguing that threats are believable and therefore credible, only when they are rational to carry out (Quackenbush, 2011, p. 9). This relationship can be illustrated by noting that during the age of nuclear deterrence, establishing credibility by threatening to engage in nuclear warfare was a hollow endeavor, since one's adversary also understood the implicit irrationality of such threats (Zagare & Kilgour, 2000, p. 66). Based on a game of generalized mutual deterrence, Zagare and Kilgour conclude that when both actors maintain credible retaliatory threats, deterrence is rendered both stable and rational; an idea reinforced by Selten's (1975) perfectness criterion which measures the robustness of equilibrium solutions in order to determine whether they are likely to be observed in practice (2000, p. 77). The idea that credibility could be empirically examined through the lens of game-theoretic modeling was as Langlois asserted; a definitive way to help "…clarify which threats are credible, what retaliations can be implemented, which ones should be entirely left to chance, and what dynamics will result from such policy choices (1991, p. 802)." Given Langlois's assertion, it is important to understand the concepts of subgame perfect equilibria and incomplete information before moving on to modeling deterrence using expected utility.

Subgame Perfect Equilibria (SPE)

The concept of equilibrium is closely tied to the rational approach to examining deterrence through game-theoretic modeling. In equilibrium, no actor has an incentive to alter its behaviors on its own. Thus, an equilibrium is considered to be inherently stable because no actor, given its current knowledge and position, can take action to improve its own position through its own devices (Morrow, 2020, p. 8). In deterrence literature, the presence of equilibrium are one of the indicators of successful deterrence strategies; and subgame perfect equilibria in particular have been used as measurements for threat credibility.

In a similar vein to the ‘burning bridges’ strategy discussed earlier, Selten argues that when an actor cannot commit to carrying out a certain retaliatory action, behavior in a subgame should constitute a subgame equilibrium because under opposite conditions, there would exist a subgame wherein the solution precludes an equilibrium. If this were the case, at least one actor would have an incentive to deviate from the solution if the subgame was actually reached, and the solution would not be self-enforcing (Van Damme & Weibull, 1995, p. 24). In other words, being unable to commit to carrying out a retaliatory threat with costs above that of the initial threat cannot be considered credible as evidenced through the examination of game-theoretic models. Specifically that within the model, if any other scenario exists that does not involve a subgame perfect equilibrium, credibility is eroded because there are other options an actor can choose to deviate to, thus making the retaliatory threat incredible. Subgame perfect equilibria are necessary for the study of deterrence by game-theoretic modeling because it provides an empirical means with which to examine the credibility of threats, and a solution concept that can determine the most rational strategies for a country to pursue.

Incomplete Information

Traditionally, game-theoretic modeling has assumed from the outset that actors are rational decision-makers who, in making decisions, strive to maximize their payoffs. However, this represents an ideal world where each actor has full knowledge of the situation (game), is assumed to know the rules, as well as the payoff functions of other players (Van Damme & Weibull, 1995, p. 28). Reality very rarely conforms to this idealized environment. In order to model a dynamic more representative of the real world, game theorists introduce a situation of incomplete information into their models. Quite intuitively, incomplete information refers to a condition where at least one player dos not have complete information about the game's payoffs or the strategies of other players. It can impact the outcome of a game due to players' unable to make informed decisions, that is they must make decisions based on uncertain or incorrect information.

The importance in modeling incomplete information stems from answering the question of how it impacts the credibility of one's threat, how credible a threat must be to deter, as well as what is the connection between the credibility and magnitude of a retaliatory threat (Zagare & Kilgour, 2000, p. 111)? The game of generalized mutual game of deterrence as examined by Zagare and Kilgour also employed a version with incomplete information. Their findings suggest that when the credibility of each player's threat is sufficiently high, the probability of successful deterrence is high but by no means certain. In a Bayesian equilibrium, players have probalistic knowledge of each other's preferences between either capitulation or retaliation which is formed on the basis of beliefs about the private information (types) of other players (Zagare & Kilgour, 2000, p. 111). Thus, Bayesian equilibria are able to model incomplete information by outlining parameters that define each player's knowledge about each other's preferences between capitulation and retaliation.

A simple Bayesian equilibria can be modeled as a bilateral competition between two states- A and B, by using strategic form. In this game, suppose that both states A and B have the capability to launch dual-use ASAT. However, both states are uncertain whether their opponent will use these ASAT for peaceful or military purposes. A's type (i.e. private information) is its true intention to deploy ASAT for either peaceful or military purposes; information that B is not privy to. The game is played only once, that is, there are no subsequent stages for retaliation, and payoffs are determined by whether or not a state launches its dual-use ASAT capability. The payoff matrix for this game is as follows:

Table 2: Simple Extended-Form Model of a Bayesian Equilibrium

	State B: Launch	State B: Not Launch
State A: Launch	(-10, -10)	(-5, -5)
State A: Not Launch	(-1, -5)	(0, 0)

Here, a Bayesian equilibrium is a pair of strategies, one each for states A and B, such that neither can do better by opting to choose a different strategy given the preferences and beliefs of their opponent. Let's suppose that the prior belief of state B about the type of state A is 0.5, meaning that state B believes that state B has a 50% chance of using its dual-use ASAT for peaceful purposes. In this situation, we can deduce that a Bayesian equilibrium exists in which state A chooses "Not Launch" and state B choses "Not Launch". This is because given their beliefs, neither country can do better by opting for an alternative strategy. If state A chooses to Launch, state B chooses between either Launching or Not Launching; if state B Launches, it would expect to receive a payoff of -5, and -10 if it chooses to Not Launch. However, if state A does choose to Launch, state B should expect a payoff of 0 if it does Not Launch, and a payoff of -1 if it chooses to Launch. Thus, this equilibrium is Bayesian because it accounts for the player's beliefs about the game, as well as the fact that each player's strategy depends on the other player's type. Another important note regarding this simple model is that it also highlights the significance of dual-use capabilities within the context of anti-satellite capabilities; namely that the potential for peaceful use can result in a more stable outcome for the reason that both countries can choose to avoid the negative consequences of escalating conflict by opting to Not Launch.

Expected Utility Theory

The last concept within game-theoretic modeling to be discussed within the limits of this thesis is that of expected utility theory. Bruce Bueno De Mesquita's theorem on the subject is widely known as the foundational empirical work on modeling international conflict. As first introduced in The War Trap, the theorem attempts to formalize, test, and examine the causes of war. While the majority of academics who have reviewed Bueno De Mesquita's work find it 'pathbreaking', 'rigorous', and 'insightful', others have denounced it as nothing more than a "…self-fulfilling theory in which the idea that decisionmakers are 'expected-utility maximisers' is both an assumption and deduction (Booth, 1984, p. 296)." Despite the claims of critics, the theorem has nevertheless generated a plethora of literature examining the causes of war; an endeavor that is by no means insignificant nor of little consequence.

Before explaining the theorem itself, Bueno de Mesquita first lays out the assumptions under which it operates:

1. "War decision making is dominated by a single, strong leader,
2. leaders are rational expected-utility maximizers,[41]
3. differences in leaders' orientations about the likely behavior of other states in the event of war affects decision making, and
4. the power a state can use in a war declines as the site of war becomes geographically distant from the nation (Bueno de Mesquita, 1981, p. 20)."

The second assumption- that leaders are rational expected-utility maximizers, is premised on the subsidiary assumptions that a) a state's utility for another is a direct, positive function of the extent to which they share a common policy perspective, and b) a state's probability of success in either a bilateral or multilateral conflict is a direct, positive function of each party's power compared to that of the others parties. Consequently, Bueno de Mesquita assumes that a state's preferences are a reflection of utility values such that the perception of an agreement of interests is designated with a utility score of 1, and the perception of the opposite- of complete disagreement, is assigned a utility score of -1 (Bueno de Mesquita, 1981, p. 29-30). This is consistent with Zagare and Kilgour's perfect deterrence assumption that a player's utilities automatically

[41] Bueno de Mesquita's interpretation of rationality is such that, "a rational actor is one who compares options and orders them according to his preference for each…[and] a rational actor always chooses the outcome he considers most desirable (Bueno de Mesquita, 1981, p. 30) ."

represent both its relative preferences and its risk attitudes (Zagare & Kilgour, 2000, p. 71).

The third assumption addresses the fact that rational actors do not always respond in the same ways to the same set of information. In other words, each decisionmaker maintains their own set of preferences, especially when responding to situations of uncertainty, as well as how willing they are to take on risk. Bueno de Mesquita defines risk taking "…as the probability of success that a decisionmaker demands before pursuing a course of action," and uncertainty as "…the degree to which the probability of success of a course of action is unknown (Bueno de Mesquita, 1981, p. 33)."

To illustrate this, let us assume that no change in a state's policy will not result in any loss of power. In changing their policy, a decisionmaker may potentially emerge with more or less power than when he first initiated an action. Suppose we assign the value of 1 to the utility of winning and -1 to the utility of losing, and then assume that the probability of winning and losing are equally likely. Since the expected utility of the war is the sum of the utilities of all possible outcomes multiplied by their probabilities, the expected value would be delineated by the equation .5(1)+.5(-1)=0. In this situation, a risk averse decisionmaker would defer from war while a risk-acceptant leader would have no qualms about escalating conflict to war. This is because the latter prefers the chance of success over inaction despite there being an equal expected utility of both winning and losing. This can be represented in the following strategic form as follows:

Table 3: Strategic Form Game of a Simple Expected Utility Model

	Player 2 Wins	Player 2 Loses
Player 1 Wins		(1, 0)
Player 1 Loses	(0, 1)	

In this game, there are no (win, win) and (lose, lose) outcomes. The payouts are listed as player 1(P1), player 2 (P2) and are based on the equation EU(w)=1(1)-(0)(-1)=1; or the expected utility of player 1 winning equals the probability of P1 times P2's payout if P1 wins, minus the probability of P1 losing times P2's payout if P1 loses. If however, unlike the situation above where there is certainty (the decisionmaker was aware of what the probability of success was for each action), we can also enter into a situation of uncertainty, or incomplete information. The decisionmaker only knows that the probability of winning is .5+*k* where *k* is the measurement of uncertainty. It then follows that the probability of losing is 1- (.5+*k*), with the value of *k* being between -.5 and .5. A risk acceptant leader may assume that *k*= .5 such that the expected utility of initiating conflict is expressed as 1(1) + 0(-1)= 1 whereas a risk adverse leader would reach a value of 0(1)+1(-1)= -1 by assuming *k*= -.5 (Bueno de Mesquita, 1981, p. 34-35).

Bueno de Mesquita accounts for both bilateral and multilateral war by generating two different equations. He defines i's (aggressor) expected utility from a bilateral war with j (defender) as:

(1) $$E(U_i)_b = [P_i(U_{ii} - U_{ij}) + (1 - P_i)(U_{ij} - U_{ii})]_{t0} + P_{it0}[\Delta(U_{ii} - U_{ij})]_{t0 \rightarrow tn} + (1 + P_i)_{t0}[\Delta(U_{ij} - U_{ii})]_{t0 \rightarrow tn}$$

Where $[P_i(U_{ii} - U_{ij}) + (1- P_i)(U_{ij} - U_{ii})]_{t0}$ calculates the expected utility of player i under the assumption that the other player's type is known, and $P_{it0}[\Delta(U_{ii} - U_{ij})]_{t0 \rightarrow tn} + (1 + P_i)_{t0}[\Delta(U_{ij} - U_{ii})]_{t0 \rightarrow tn}$ calculates the expected utility of player i under the assumption that the other player's type is known. The variable $_{t0}$ represents the time at which the belief is formed and $_{tn}$ represents the time at which the game is played. Δ (U_{ij} – U_{ii}) thus calculates the change in utility from $_{t0}$ to $_{tn}$.

The second equation calculates i's expected utility from a multilateral war with j and k (all other states) as the following:

With k supporting i:

(2) $E(U_i)_{kl\ 1} = (P_{ik}\ U_{iki} + (1\text{-}P_{ik})\ U_{ikj})_{t0} + P_{ik\ t0}\ (\Delta U_{iki})_{t0 \rightarrow tn}$

$+ (1\text{+}P_{ik})_{t0}\ (\Delta\ U_{ikj})_{t0 \rightarrow tn}$

With k supporting j:

(3) $E(U_i)_{kl\ 2} = [(1\text{-}P_{jk})\ U_{iki} + P_{jk}\ U_{ikj}]_{t0} + (1\text{-}P_{jk})_{t0}\ (\Delta\ U_{iki})_{t0 \rightarrow tn}$

$+ P_{jk\ t0}\ (\Delta\ U_{ikj})_{t0 \rightarrow tn}$

Where: (note that for each i value, there is a corresponding j value)

- P_i = i's current perception of his probability of succeeding against j in a bilateral conflict.
- 1- P_i = i's current perception of his probability of losing against j in a bilateral conflict
- U_{ii} = i's utility for its own policies where U_{ii} = 1. i's utility for j's policies are such that U_{ij} varies between 1 and -1. (U_{ii} - U_{ij}) is thus i's perception of what could be gained by winning in a bilateral war with j, wherein i can impose new policies on j.
- Δ (U_{ii} - U_{ij}) $_{t0 \rightarrow tn}$ = i's perception of anticipated change in the difference between i's utility for its own policies and i's utility for j's policies as a function over time from $_{t0}$ (present time) to $_{tn}$ (a future point in time). Conversely, Δ (U_{ij} - U_{ii}) $_{t0 \rightarrow tn}$

is i's perception of anticipated change in how much j would want to alter i's policy; or I's perception of anticipated policy losses.

- U_{iki} = i's perception of the utility gained from k, or each third party actor.
- P_{ik} = i's perception of the likelihood of success against j, supposing that third party k supports j.
- 1-P_{ik} = i's perception of the likelihood of failure against j, supposing that third party k supports j.

Equation 2 represents i's calculus of the expected utility i gains from third party k to its success or failure, as well as the probability of i's winning and losing k's support (Bueno de Mesquita, 1981, p. 57). Equation 3 on the other hand, is i's estimation of the expected value third party k will contribute to j's success or failure, as well as the probability that i will still emerge as the victor even if k allies with j, and the converse probability that i loses if k allies with j. After summing across all k, Bueno de Mesquita deduces that "…whether i believes a third-party nation k will support i or j is determined…by i's belief about k's relative utility for i and j (1981, p. 58)." Given this, the expected gains or losses in a bilateral war must depend on the relative capabilities of both players, whereas the expectation of aid from all other third parties k depends on utilities. Bueno de Mesquita further asserts the following regarding k's relationship to i and j:

1) When $E(U)k_i$ (the perceived expected utility of k supporting i) > 0, i believes that k is more likely to support i than j.
2) When $E(U)k_i$ <0, i believes that k is more likely to support j than i.
3) When $E(U)k_i$ = 0, i believes that k is indifferent and therefore unlikely to support neither i nor j.

Thus, in summing the expected utility $E(U_i)$ for k across all third party states, i can determine an overall estimate of the expected impact of third party states. Drawing from the i-j-k relationship, i can expect similar outcomes for the expected value of support from k:

1) When the sum of $E(U_i)$ for k is positive, one can expect more support for i over j
2) When the sum of $E(U_i)$ for k is negative, one can expect less support for i over j
3) When the sum of $E(U_i)$ for k is zero, one expects that i believes that third party k's support is divided in half between i and j, k's effects are rendered insubstantial.

In this sense, it is still rational for i to go to war even if k throws its support behind j. This is because even if this summation is negative, and i believes that k will support j, it is not sufficient to say that from i's perspective, there is no value in going to war since it could be true that the expected utility of i is greater than the expected utility for k. Thus, the overall expected value of armed conflict could be positive, rendering the decision to go to war rational as permitted within the theory (Bueno de Mesquita, 1981, p. 59).

Under these conditions, we can model such a scenario by implementing Bueno de Mesquita's expected utility equation. Suppose that states i and j are in dispute over a who will be permitted to establish a moon base in a geographically favorable position on the Moon. State i has a lower expected utility of engaging in conflict than state j, and state j maintains a powerful military that boasts technology more advanced than that of State i's. However, state i has recently developed a non-kinetic anti-satellite system capable of disabling state j's communication and navigation satellites, which are crucial for its military operations. State i believes that if it were to use this weapon, it would drastically curtail state j's military advantage, and simultaneously increase its own chances of winning a conflict. Based on this belief, state i calculates its expected utility of going to war is still rational despite it maintaining a lower expected utility than state j. In this scenario, state j may choose to abstain from conflict due to the high cost, while state i sees an opportunity to initiate conflict, prevail, and reap the benefits of a more favorably-placed moon base. If we were to represent this as equation, it would depicted as the following:

$$E(U_i) = P_i\,(U_{ii} - U_{ij}) + (1- P_i)(\,U_{ij} - U_{ii}) + P_{it0}\,(\Delta(U_{ii} - U_{ij}))_{t0 - tn} + (1- P_i)t0(\Delta(U_{ij} - U_{ii}))_{t0 - tn}$$

Where:

- $E(U_i)$ is the expected utility for state i
- Pi is the probability of state i winning the war
- Uii is the utility for state i if it wins the war
- U_{ij} is the utility for state i if it loses the war
- Pit0 is the probability of state i winning the war at time t0
- $\Delta(U_{ii} - U_{ij})$ is the change in utility for state i from winning the war at time $_{t0}$ to $_{tn}$
- $\Delta(U_{ij} - U_{ii})$ is the change in utility for state i from losing the war at time $_{t0}$ to $_{tn}$ is the starting point of the war and $_{tn}$ is the end point of the war

This model demonstrates that even though State i has a lower expected utility of winning the war (U_{ii}) than state j (U_{ij}), but state A has higher probability of winning the war at time $_{t0}$ (Pit0) and also the change in utility for state i from winning the war at time $_{t0}$ to $_{tn}$ ($\Delta(U_{ii} - U_{ij})$) is greater than the change in utility for state j from losing the war at time $_{t0}$ to $_{tn}$ ($\Delta(U_{ij} - U_{ii})$) . Therefore, despite state i having a lower expected utility than state j, it is still rational for state i to go to war because the expected utility of going to war for state i is greater than not going to war. Based on this equation, we can also derive some predictions:

1) if P_i (the probability of success) is high, then the expected utility of State i engaging in conflict with State j will be positive, indicating that it is rational for State i to engage in conflict despite having a lower expected utility,

2) if U_{ii} - U_{ij} (the difference in utility if State i is successful) is high, then the expected utility of going to war will be more positive, further supporting the rationality of a conflict for State i,

3a) if P_{it0} (the probability of success multiplied by the initial time period) and $_{t0}$ (the initial time period) are both high, then the expected utility of engaging in conflict will be positive, as the potential gain from success increases with the length of time before the end of the war,

3b) conversely, if P_i is low, or U_{ij} - U_{ii} (the difference in utility if State i is not successful) is high, then the expected utility of initiating conflict will be negative, indicating that it is not rational for State i to initiate the conflict, and

5) if P_{it0} and $_{t0}$ are both low, then the expected utility of going to war will be less positive, as the potential gain from success decreases with the length of time before the end of armed conflict.

The rationality for going to war is thus not solely dependent on a state's expected utility. Rather, it is based on an array of variables including the change in expected utility over time, the high costs of war for an opponent, neutralizing capabilities, as well as any potential alliances. Indeed, there are certain circumstances under which an aggressor with a positive expected utility would be expected to defer from escalation, despite not having its demands fulfilled. If for example, one side maintains a deterrent counterthreat (the counter-threating state), it must also mean that it maintains a positive expected utility, since it would be irrational to take any action against the state who made the initial threat. Put alternatively, without positive expected utility, the counter-threatening state would need to behave irrationally in order to act against the original threat. In such a situation, one way in which states who maintain a positive expected utility can still lose in a conflict is to be confronted by a third party entity that can credibly threaten to impose a significant cost on them (Bueno de Mesquita, 1981, p. 170). The presence and ability for a third-party provide a deterrent threat can prevent conflicts from escalating into war, and can also help to bring a swift resolution to ongoing wars. This serves as an additional method, in addition to analyzing expected utility, to differentiate between disputes that have a high likelihood of escalating into war and those that are likely to be resolved peacefully.

Building upon his original model, Bueno de Mesquita ultimately formulated a revised expected utility equation that accounts for the inflexibility of risk-taking, and the separate evaluation of the perceptions of i and j; that is, how each player's situation appears as seen from the perspective of their opponent (De Mesquita, 1985, p. 156-157). The revised equation is:

$$
\begin{aligned}
E^i(U_{ij}) = \\
&[P_i (U^i_{si}) + (1- P_i)(U^i_{fi})] \\
&+ \sum_{k \neq i,j} (P^i_{kn} + P^i_{jkn} - 1)(U^i_{kni} - U^i_{knj}) \\
&- [(Q^i_{qi} U^i_{qi}) + (1-Q^i_{qi} Q^i_{bi})(U^i_{bi}) + (1- Q^i_{bi})(U^i_{wi})]
\end{aligned}
$$

Where each line of the equation delineates the following:

- 1st line: utility of i succeeding in a multilateral conflict.
- 2nd line: utility of alliance k supporting i.
- 3rd line: probability of i maintaining its policies.

However, some of the limitations with Bueno de Mesquita's expected utility theorem stem from his suppositions about power; specifically that it is derived only from indicators of what he calls tangible national capabilities. He admits that intangible factors, such as the willingness of citizens to support a war effort, or the reputational costs of reneging from a commitment, could easily influence calculations about the likely behavior of third parties (Bueno de Mesquita, 1981, p. 128). While he addresses two other shortcomings of his theorem in a revised formulation, he does not take into account the effects of inferential capabilities.

Considering these concepts on expected utility, this thesis hopes to contribute to the literature of deterrence by proposing to incorporate an expected utility model that examines the efficacy of deterrence within the context of the use of anti-satellite capabilities. Moreover, in limiting analysis to direct-general and extended general deterrence, the applications of this model can be applied to the proliferated use of salami tactics which do not necessarily meet the thresholds for overt conflict escalation. In the final chapter, this thesis explores a new expected utility model for deterrence in the space age.

Chapter VI. Deterrence and Expected Utility for Inferential Anti-Satellite Capabilities

The efficacy of deterrence theory has traditionally been limited to a nuclear weapons environment, where the demonstration of physical power signaled the ultimate form of retaliation and conflict escalation. However, it has failed to sufficiently adopt to a constantly evolving international environment, where new power dynamics between state and non-state actors are undergoing changes as each vies for a strategic edge over the other. Simultaneously, these new actors are employing an increasing amount of strategies that are antithetical to the nature of nuclear deterrence. The expansion of inferential anti-satellite capabilities for example, has altered the cost-benefit calculus of deterrence due to their inferential characteristics which render states unable to perceive aggression- a critical requisite for successful deterrence. However, the implications that such capabilities have over credibility and rationality as defined within the literature demonstrate that deterrence cannot operate in an environment where inferred capabilities are the primary means of retaliation. One approach to remedying the issue of credibility is to utilize a hand-tying strategy through an alliance mechanism. The transposition of these arguments to the empirical side of analysis is best understood by employing game-theoretic modeling. The expected utility theorem in particular, despite being well-suited for the study of deterrence, has several shortcomings that render it inadequate as a mechanism with which to study the efficacy of deterrence in outer space.

In recognizing these shortfalls, this thesis proposes to supplement a revised iteration of Bueno de Mesquita’s expected utility model with a treaty (alliance) mechanism with a nuclear weapon state. This model permits the analysis of salami tactics by introducing direct-general and extended-general deterrence assumptions, incomplete information, and subgame perfect equilibria to the model. By including these factors in this particular combination, we can account for: the dynamic of the international environment before conflict escalates into the initial stages of conflict; an intrinsic characteristic of salami tactics, the lack of credibility inherent in inferential anti-satellite capabilities, and the extent to which a treaty mechanism bolsters deterrent success.

Assumptions and Limitations

These models operate under several straightforward assumptions which align with the decision-theoretic camp of empirical analysis : 1) decisionmakers are assumed to rational, and always act to maximize their expected utility, 2) decisionmakers act in the interest of a single entity, as opposed to being divided into factions with conflicting interests, 3) conflict is costly, and the cost of conflict will always be weighed against its benefits, 4) decisionmakers are more likely to follow through on commitments if they are perceived as credible, 5) alliances in the form of treaty mechanisms effect the calculus of deterrence by increasing the credibility of a retaliatory threat, and 6) inferential anti-satellite capabilities necessitate retaliatory threats because alone, they are not sufficient for successful deterrence.

The first three assumptions align with De Mesquita’s in that they assume that the head of state is charged with the final say in deciding on whether or not to engage in conflict; and inversely that their disapproval is sufficient to refrain from conflict (Bueno de Mesquita, 1981, p. 20-21). Secondly, these assumptions are predicated on the logic that heads of state wish to maximize the net benefits they expect from any given foreign policy choice. In representing their country, head of state thus is intimately tied to the welfare of their society as a whole, and as follows, such a leader cannot rationally be expected to engage in conflict if their perception of war's net benefits are less than that of abstaining from conflict (Bueno de Mesquita, 1981, p. 29). The fourth assumption signals the importance of credibility within deterrence calculus. From a decision-theoretic approach, a threat is credible if it maintains a positive expected utility against the initial threat. However, one way in which this calculus can be overturned- that is, if a state with a positive expected utility of engaging in conflict is confronted by a third-party entity with the power to credibly threaten to impose a higher net cost, then it may be in the best interest of a state to defer from conflict (Bueno de Mesquita, 1981, p. 171). This leads us to the fifth assumption, and within the parameters of this thesis we further specify that a third-party treaty mechanism has the function of increasing the credibility of a retaliatory threat. The basis for the reasoning is the concept of words and deeds in costly signaling (refer back to Chapter III). In essence, the positive effect of costly signaling via a third-party treaty mechanism creates a strategy of hand-tying that serves to strengthen the credibility of a retaliatory threat. Lastly, assumption six is the inherent shortcoming of inferential anti-satellite capabilities underlying the argument of this thesis. As we have seen throughout the course of this discussion, inferential ASAT capabilities are not readily visible, and not easily recognized as a deterrent threat by adversaries due to their subversive nature. Thus, inferential ASAT *necessitate* retaliatory threats because their thresholds for deterrence are too high to be satisfied alone.

The limitations of this model stem from its relatively narrow parameters in that it only considers the effects of alliance mechanisms as a supplementary strategy for inferential ASAT capabilities. Admittedly there are other means with which to achieve this, such as through the use of kinetic-physical ASAT. However, the real-world application of such a proposition is, while not wholly irrelevant, highly condemned for its negative effects on the environment in outer space as well as on Earth. Thus, while the strategic evolution of deterrence in outer space may very well expand to include such capabilities in the future, this thesis will focus its attention on the strategic implications of inferential anti-satellite capabilities, as they are timely and relevant to the current international environment.

Modeling Inferred ASAT Deterrence

Table 4 below contains the expected utility models for inferential ASAT categorized by whether or not a NWS or NNWS maintains the treaty mechanism, and by whether or not the equation is examining demonstrable or inferred anti-satellite capabilities. This model is arranged such that i is the preemptor; deciding whether or not to launch a first-strike against j, the defender who will either maintain 1) demonstrable ASAT, 2) inferential ASAT, 3) inferential ASAT with a NNWS treaty mechanism, and 4) inferential ASAT with a NWS treaty mechanism. Each treaty mechanism for either NNWS or NWS are made both formally and publicly so as to make certain, the presence of hand-tying and burning bridges strategies which ascertain j’s commitment to carrying out the retaliatory threat. The expected utility values are set so that as a value approaches 1, there exists a perfect harmony of interests between i and j, and the status quo remains in-tact. If the values approach -1, then there is disagreement between the actor's perceptions of each other's strategies, resulting in a high probability of the outbreak of conflict. These parameters effectively mean that deterrence failure results from an outcome whose value is anywhere between 0 and -1, and deterrence success results from a outcome whose value is anywhere between 0 and 1.[42] For a chart outlining the definition of each variable used, see Appendix 1.

Table 4: Expected Utility Models for Inferential ASAT

Demonstrable ASAT	Inferential ASAT
1a. The Base Equation: $E^i(U_{ij})=$ $[P_i (U^i_{si}) + (1- P_i)(U^i_{fi})]$ $- [(Q^i_{qi} U^i_{qi}) + (1-Q^i_{qi} Q^i_{bi})(U^i_{bi})+(1-Q^i_{bi})(U^i_{wi})]$ $+ R^j_d (\sigma_1, \sigma_2)(1-R^j_d (\sigma_1, \sigma_2))$	1b. The Base Equation: $E^i(U_{ij})=$ $[P_i (U^i_{si}) + (1- P_i)(U^i_{fi})]$ $- [(Q^i_{qi} U^i_{qi}) + (1-Q^i_{qi} Q^i_{bi})(U^i_{bi})+(1-Q^i_{bi})(U^i_{wi})]$ $+ R^j_r (\sigma_1, \sigma_2)(1-R^j_r (\sigma_1, \sigma_2))$
2a. NNWS Treaty Mechanism Equation: $E^i(U_{ij})=$ $[P_i (U^i_{si}) + (1- P_i)(U^i_{fi})]$ $+ \sum_{k\neq i,j} (P^i_{kn} + P^i_{jkn} – 1)(U^i_{kni} - U^i_{knj})$ $- [(Q^i_{qi}U^i_{qi}) + (1-Q^i_{qi}Q^i_{bi})(U^i_{bi})+(1-Q^i_{bi})(U^i_{wi})]$ $+ R^j_d (\sigma_1, \sigma_2)(1- R^j_d (\sigma_1, \sigma_2))$	2b. NNWS Treaty Mechanism Equation: $E^i(U_{ij})=$ $[P_i (U^i_{si}) + (1- P_i)(U^i_{fi})]$ $+ \sum_{k\neq i,j} (P^i_{kn} + P^i_{jkn} – 1)(U^i_{kni} - U^i_{knj})$ $- [(Q^i_{qi}U^i_{qi}) + (1-Q^i_{qi}Q^i_{bi})(U^i_{bi})+(1-Q^i_{bi})(U^i_{wi})]$ $+ R^j_r (\sigma_1, \sigma_2)(1- R^j_r (\sigma_1, \sigma_2))$
3a. NWS Treaty Mechanism Equation $E^i(U_{ij})=$ $[P_i (U^i_{si}) + (1- P_i)(U^i_{fi})]$ $+ \sum_{k\neq i,j} (P^i_k + P^i_{jk} – 1)(U^i_{ki} - U^i_{kj})$ $- [(Q^i_{qi}U^i_{qi}) + (1-Q^i_{qi}Q^i_{bi})(U^i_{bi})+(1-Q^i_{bi})(U^i_{wi})]$ $+ R^j_d (\sigma_1, \sigma_2)(1- R^j_d (\sigma_1, \sigma_2))$	3b. NWS Treaty Mechanism Equation $E^i(U_{ij})=$ $[P_i (U^i_{si}) + (1- P_i)(U^i_{fi})]$ $+ \sum_{k\neq i,j} (P^i_k + P^i_{jk} – 1)(U^i_{ki} - U^i_{kj})$ $- [(Q^i_{qi}U^i_{qi}) + (1-Q^i_{qi}Q^i_{bi})(U^i_{bi})+(1-Q^i_{bi})(U^i_{wi})]$ $+ R^j_r (\sigma_1, \sigma_2)(1- R^j_r (\sigma_1, \sigma_2))$

[42] I have kept the variables Qiqi =.6 , Uiqi = -.6, Qibi = .8, Uibi = -.8, Uiwi = .6 Uifi= -1, Pikn= 1, Pijkn=1, Uikni= 1, and Uiknj = 1 consistent across all equations except for the first example in equation 1a, in order to evaluate the changes in i's perception of its probability of winning, i's perception of its utility for succeeding, and j's response function. These values are arbitrarily chosen based on how i is assumed to behave when confronted with different iterations of j's retaliatory threats.

The Base Equation

Equations 1a and 1b are base equations that serve as a foundation for measuring further iterations of the equation. In other words, they are used as a control which measures the efficacy of deterrence without the implementation of any treaty mechanism; be it with NNWS or NWS. These equations denote scenarios wherein j's retaliatory threat is either strictly demonstrable (1a), or inferential (1b), where line one represents the utility of i succeeding in a bilateral conflict against j, line two represents the probability of i maintaining its policies, and line three is a measure of j's threat credibility. The term R^j_d (σ1, σ2) signifies j's response function (retaliatory threat), or put alternatively, the degree of cooperativeness in response to i's preemptive threat, where the level of cooperativeness is indexed by $\sigma_1 = 1$ and $\sigma_1 = 0$, with the former indicating cooperative behavior and the latter uncooperative behavior. Thus, a higher measurement of cooperation suggests a higher likelihood of deterrence success. (1- R^j_d (σ1, σ2)) on the other hand, represents the threatened level of non-cooperativeness used by j in order to deter deviations from the original game path (i.e. the initial strategy wherein j's retaliatory threat is credible because it represents a subgame perfect equilibrium) (Arce & Sandler, 2009, p. 392).

Generally, we should expect deterrence failure for R^j_r and deterrence success for R^j_d because the threat credibility of j within the context of demonstrable ASAT should be higher than that of its inferential counterpart. Furthermore, if $R^j_d(\sigma_1, \sigma_2) < \sigma_2$, j's response is expected to maintain a lower level of cooperativeness than the initial strategy, and if $R^j_d(\sigma_1, \sigma_2) > \sigma_2$, then j acquiesces to i due to j maintaining a higher level of cooperation than the initial strategy. Lastly, if $R^j_d(\sigma_1, \sigma_2) = \sigma_2$, then the status quo is maintained. These inequalities also indicate that j should be at least as well off in carrying out its threat than committing to its initial strategy (Arce & Sandler, 2009, p. 396). Below are two scenarios where deterrence failure and success are expressed through the Base Equations 1a and 1b respectively.

1a: If we set the values of each variable arbitrarily as: $P_i = 1$, $U^i_{si} = 1$, $Q^i_{qi} =1$, $U^i_{qi} = 1$, $Q^i_{bi} = 1$, $U^i_{bi} = 1$, $U^i_{wi} =$, R^j_d (σ1, σ2) = 1, equation 1a yields: $Ei(U_{ij}) = [1(1) + (1-1)(1)] - [(1(1) + (1-1)(1)(1) + (1-1)(1))] + (1)(1-(1)) = 1 + 0 - 0 + 0 = 1$. These values indicate that i perceives a 100% probability of success in a conflict with j, 100% probability of maintaining its current policies, 100% probability of improving its policies, and 100% probability of both j's initial response and the outcome being successful in maintaining deterrence. This situation is interesting because it results in successful deterrence as a result of j's credible response function (i.e. demonstrable, or kinetic-physical ASAT) despite i perceiving its probability of succeeding in this bilateral conflict with a 100% likelihood of success, it fails. This can be explained by a couple of possible scenarios; the first is that i is irrational- thus despite j's credible retaliatory threat, i still believes it can succeed. The second is that i also maintains a credible, demonstrable deterrent threat that it believes can succeed in a prolonged conflict with j, and the reason why it fails in here is due to the fact that this equation is a limited stage model.

1b: If we set the values of each variable arbitrarily as: $P_i = .7$, $U^i_{si} = -.5$, $Q^i_{qi} =.6$, $U^i_{qi} = -.6$, $Q^i_{bi} = .8$, $U^i_{bi} = -.8$, $U^i_{wi} = .6$, R^j_r (σ1, σ2) = -1, equation 1b yields: $E^i (U_{ij}) = [.7(-.5) + (1-.7)(-.7)] - [(.6(-.6) + (1-.6)(.8)(-.8) + (1-.8)(.6))] + (-1)(1-(-1)) = -.475 + (-.368) + 0 = -.843$. Since this value is closer to -1, which indicates the conflict is more probable to result in deterrence failure by j, it justifies the supposition that inferential anti-satellite capabilities by themselves are not a sufficient means for producing successful deterrence outcomes. In this equation, given the values of the variables, i expects a 70% probability of succeeding in conflict against j, whose retaliatory response is comprised solely of inferential ASAT. Furthermore, since j's initial response function of R^j_r (σ1, σ2) = -1 indicates an incredible retaliatory measure, i is more likely to succeed in conflict against j whose retaliatory response is comprised only of inferential ASAT.

NNWS Treaty Mechanism Equation

Equations 2a and 2b depict a multilateral conflict where j supplements inferential ASAT threats through a treaty mechanism with a NNWS, where line one represents the utility of i succeeding in a multilateral conflict, line two represents the utility of NNWS alliance k supporting j, line three represents the probability of i maintaining its policies, and line four a measurement of j's threat credibility. The component of the treaty mechanism as represented within both sets of 2a/2b and 3a/3b are premised on the assumption that third-party alliance k maintains some degree of power.[43] It follows that ($P^i_{kn} + P^i_{jkn}$) must meaningfully exceed ($P^i + P^j$) in such a way that it is impossible for k to be indifferent between the strategic policies of i and j. This is a critical assumption because it allows this model to function on the basis of having a formal alliance supplement, as well as exerts a positive net effect on deterrence strategies. Equations 2a

[43] Power transition theorists Organski and Kugler (1980) have asserted that alliances are irrelevant within the context of armed conflict, when the dynamic of war is left between two of the most powerful nations. Their reasoning is premised on the fact that the warring states are too powerful in comparison with any other nation that alliances are rendered ineffective. Their logic follows the equation of $P_i + P_j = 1$, $P_{ik} \geq P_i$ and $P_{jk} \geq P_j$, such that $P_{ik} + P_{jk} \geq P_i + P_j$. Thus, if ($P_i + P_j$)= 1, then ($P_{ik} + P_{jk}$ -1) = 0 and ($P_{ik} + P_{jk}$ -1) (U^j_{ki} – U^j_{kj})= 0 (De Mesquita, 1985, p. 162-163). However, this thesis assumes that k does maintain some degree of power; as a result, it must be that $P_{ik} + P_{jk}$ exceeds $P_i + P_j$, demonstrating that k is not impartial between the policies of I and j.

and 2b also assumes that $P^i_{kn} > P^i_{jkn,}$ where $P^i_{kn} + P^i_{jkn} = 2$. In other words, the effect of third-party alliance k must always amount to a differential between inferential ASAT deterrence strategies both with and without a treaty mechanism. Furthermore, we should also observe that a NNWS treaty mechanism to produce a weaker deterrent than that of a NWS, and a stronger deterrent than one without any type of treaty mechanism. Suppose:

$(P^i_{kn} + P^i_{jkn}) = 2$

$(P^i_{kn} + P^i_{jkn} – 1) = 1$

$(P^i_{kn} + P^i_{jkn} – 1) (U^j_{ki} – U^j_{kj}) = Ø$

Where $(U^j_{ki} – U^j_{kj})$ is the net value i believes will be contributed by k to the conflict between i and j. In all cases where $(U^j_{ki} – U^j_{kj}) = Ø$, k will always have some positive net effect on the calculus of deterrence (De Mesquita, 1985, p. 161). The only instance when k may be indifferent between supporting i or j would be if both $(P^i_{kn} + P^i_{jkn} – 1) (U^j_{ki} – U^j_{kj}) = 0$. Such a result would indicate that k does not retain some degree of deterrent power in supporting either i or j, or that it simply has a weaker degree of power in comparison to i and j. However, this is an impossibility based on the assumptions of this model. Bueno de Mesquita likewise posits that :

> "So long as the difference between U_{ki} and U_{kj} is not zero, one can see that the multilateral component of the calculus is larger when k is assumed to be a major power than when k is assumed to be relatively weak. Therefore, a "powerful" third party is presumed to be able to have an impact on virtually all conflicts, whereas a weak one can have an impact only under a special set of circumstances [and] major powers, then, are more likely to play a role in conflicts than are lesser nations (1985, p.163)."

The assumptions underlying the calculus behind the deterrent effect of third party alliance k serves to highlight the importance of power in determining the characteristics, particularly that of power, of a third party's ability to impact a conflict. A powerful third party is seen as having a wider range of influence and the ability to affect virtually all conflicts, while a weak third party can only impact conflict under a specific set of circumstances. This notion could be interpreted to suggest that a more powerful third party state; by means of nuclear capabilities or otherwise, could indirectly bolster the deterrent power of a weaker state through an alliance mechanism. In aligning with the principles of realism, whose arguments are premised on the idea that power is the singular most important factor in determining the actions of states, Bueno de Mesquita is suggesting that more powerful states are more likely to be involved in conflicts due to their ability to influence conflict through their disproportionately large degree of power. Thus, the impact of third party states who maintain nuclear weapons capabilities should not be overlooked, and can provide insight into the nature of conflict in outer space.

In comparing equations 2a and 2b, the NNWS treaty mechanism is expected to lend equal amounts of deterrent power to state j in both scenarios of demonstrable and inferred response functions; this is of course the same for equations 3a and 3b with regards to a NWS treaty mechanism. Simultaneously, the presence of a NNWS treaty mechanism should supplement j's deterrent power such that its response function is more credible with it than without it, and that j's NNWS treaty mechanism, though effective, will not be more effective than a NWS treaty mechanism.

2a.1: If we set the values of each variable arbitrarily as follows: P_i = -.5, U^i_{si} = -.5, Q^i_{qi} =.6 , U^i_{qi} = -.6, Q^i_{bi} = .8, U^i_{bi} = -.8, U^i_{wi} = .6, R^j_r (σ1, σ2) = -1, U^i_{fi}= -1, P^i_{kn}= 1, P^i_{jkn}=1, U^i_{kni}= 1, and U^i_{knj} = 1, the equation would yield:

Ei(Uij) = [-.5 * (-.5) + (1 - (-.5))(-1)]
+ ∑_(k ≠ i, j)(1 + 1 – 1)(1 - 1)
[(.6 * (-.6) + (1-.6*.8)(-.8)+(1-.8)(.6))]
-1 * (1--1)
Ei(Uij) = [.25 + 1]+ 0[(-.6 + .12 + .2)]
Ei(Uij) = 1.25 - .48

$Ei(Uij) = 0.77$

Given the above set of values, equation 2a results in successful deterrence. Here, we see that i believes that it does not have a good probability of succeeding in a conflict with j, and that it's perceived utility of succeeding will not produce a net positive benefit. By changing the values of each variables, we can render an expected utility value that corresponds with a substantially higher probability of j deterring i while implementing a NNWS treaty mechanism. Suppose: : $P_i = -1$, $U^i_{si} = -1$, $Q^i_{qi} = .6$, $U^i_{qi} = -.6$, $Q^i_{bi} = .8$, $U^i_{bi} = -.8$, $U^i_{wi} = .6$, $R^j_r (\sigma 1, \sigma 2) = .5$, $U^i_{fi} = -1$, $P^i_{kn} = 1$, $P^i_{jkn} = 1$, $U^i_{kni} = 1$, and $U^i_{knj} = 1$, equation 2a then yields:

$$
\begin{aligned}
2a.2: Ei(Uij) = &[-1 * (-1) + (1 - (-1))(-1)] \\
&+ \sum\_(k \neq i, j)(1 + 1 - 1)(1 - 1) \\
&[(.6 * (-.6) + (1-.6*.8)(-.8)+(1-.8)(.6))] \\
&.5 * (1-.5)
\end{aligned}
$$

$$Ei(Uij) = [1 + 2] + 0 [(-.6 + .12 + .2)]= .25$$

$$Ei(Uij) = 3 - .48 + .25$$

$$Ei(Uij) = 2.77$$

The second iteration of equation 2a results in a much higher value indicative of j's success in deterring i. The only variables for which the values had been altered were P_i and U^i_{si}, with the original and revised values as -.5 to -1 for each. Since equation 2a.2 resulted in a much higher probability of deterrence success than equation 2a.1, we can deduce that as i's perception of the strength of j's response function increases, j is more likely to be successful in deterring i within the context of j's utilization of a NNWS treaty mechanism to supplement an demonstrable ASAT retaliatory measure.

2b: If we set the values of each variable arbitrarily as follows: $P_i = -1$, $U^i_{si} = -.5$, $Q^i_{qi} = .6$, $U^i_{qi} = -.6$, $Q^i_{bi} = .8$, $U^i_{bi} = -.8$, $U^i_{wi} = .6$, $R^j_d (\sigma 1, \sigma 2) = .5$, $U^i_{fi} = -1$, $P^i_{kn} = 1$, $P^i_{jkn} = 1$, $U^i_{kni} = 1$, and $U^i_{knj} = 1$. Substituting these values into equation 2a yields:

$$
\begin{aligned}
Ei(Uij) &= [-1 * (-0.5) + (1-(-1)) * (-1)] \\
&+ (1 + 1 - 1) * (1 - 1) \\
&- [(.6 * -0.6) + (1 - .6 * .8) * (-0.8) + (1 - .8) * .6] \\
&+ .5 * (1 - .5) \\
&= 1.5 - 0.3 - 0.96 + 0.25 \\
&= 0.59
\end{aligned}
$$

As evidenced by the lower expected value, we can infer that deterrence- while still probable, is far less likely in a situation where j employs inferential ASAT in lieu of its demonstrable counterpart. Note that in comparison to the results of equation 1b; and assuming that third party alliance k wields at least as much power or more than j, that a NNWS treaty mechanism bears some positive net gain on the deterrence success of j's strategies.

NWS Treaty Mechanism Equation

Equations 3a and 3b represent a multilateral conflict where j supplements inferential ASAT retaliatory threats through a treaty mechanism with a NWS. These equations are formatted in the same manner as equations 2a and 2b, thus each line of the equation bears the same function with line one representing the utility of i succeeding in a multilateral conflict, line two representing the utility of a NWS alliance k supporting j, line three as the probability of i maintaining its policies, and line four as a measure of j's threat credibility. In this scenario, the treaty mechanism functions such that $P^i_{kn} > P^i_{jkn}$, where $P^i_{kn} + P^i_{jkn} = 3$. This should yield a result in which the value of k exceeds that of the NNWS treaty mechanism:

$$(P^i_{kn} + P^i_{jkn}) = 3$$

$$(P^i_{kn} + P^i_{jkn} - 1) = 2$$

$$(P^i_{kn} + P^i_{jkn} - 1)\ (U^j_{ki} - U^j_{kj}) = \varnothing$$

Similarly to equation 2a and 2b's representation of k's expected impact on j's response function, the value of the treaty mechanism can be adjusted to account for the higher degree of power and thus credibility j is able to employ. Here, we see a scenario wherein k yields a value of 3 instead of 2; thus, based on the principle of transitivity, we should then be able to observe a higher expected utility of j's deterrent efforts supplemented by a NWS treaty mechanism, not only in comparison to its NNWS counterpart, but also in comparison to a deterrence strategy wherein j employs an inferential anti-satellite capabilities without any type of treaty mechanism.

3a: If we set the values of each variable arbitrarily as follows: P_i = -.5, U^i_{si} = -1, Q^i_{qi} =.6 , U^i_{qi} = -.6, Q^i_{bi} = .8, U^i_{bi} = -.8, U^i_{wi} = .6, R_{jd} (σ1, σ2) = -.5, U^i_{fi} = -1, P^i_{kn} = 1, P^i_{jn} =1, U^i_{ki} = 1, and U^i_{kj} = 1, the equation would yield:

Ei(Uij) = [-0.5(-1) + (1 + 0.5)(-1)]
+ [0]
- [(0.6(-0.6) + (1-0.6*.8)(-.8) + (1-.8)(.6))]
+ (-0.5)(1+0.5)
= -1.7 + (-0.52) - (-0.52) = -1.7 + 1.04 + 1.04
Ei(Uij) =1.97.

3b: And, if we set the values of each variable arbitrarily as follows: P_i = -1, U^i_{si} = -1, Q^i_{qi} =.6 , U^i_{qi} = -.6, Q^i_{bi} = .8, U^i_{bi} = -.8, U^i_{wi}= .6, R_{jr} (σ1, σ2) = -.5, U^i_{fi}= -1, P^i_{kn}= 1, P^i_{jn}=1, U^i_{ki}= 1, and U^i_{kj} = 1, the equation would yield:

Ei(Uij) = [-1 * -1]
+ [(1 + 1 - 1) * (1 - 1)]
- [(.6 * -.6) + (1-.6 * .8) * -.8 + (1-.8) * .6]
+ (-.5 * (1 + -.5))
Ei(Uij) = 1 - 0 - (-0.24 + (-0.12) + 0.4) + (-0.25)
Ei(Uij) = 1 - (-0.24 + -0.12 + 0.4 - 0.25)

$E_i(U_{ij}) = 1 + 0.01$
$E_i(U_{ij}) = 1.01$

Based on the equation of 3a and 3b, we can infer by comparison to equations 2a and 2b, that NWS treaty mechanisms yield stronger deterrent supplements to inferential anti-satellite capabilities than that of NNWS treaty mechanisms. Moreover, across all equations, demonstrable ASAT capabilities generated a higher probability of success than that of its inferential counterpart. Thus, the notion that a the credibility of j's retaliatory threat is an increasing function of successful deterrence remains consistent within both theoretical and empirical literature.

Implications

It is also possible to render these equations as a game-theoretic model playing out between two hypothetical players. Suppose that the U.S. and China are each gradually taking actions to establish strategic presence in outer space, both equipped with their own inferential anti-satellite capabilities. China's primary method of establishing its presence is through the use of salami tactics, and China's goal is to gradually undermine U.S. power by slowly taking over strategically favorable locations in orbit around the Earth. The U.S. maintains an alliance with the United Kingdom (NWS) which should serve to enhance the U.S.'s deterrent credibility. This game is played with incomplete information, meaning that both players do not know exactly what the other player knows or will do. The payoffs for each player are determined by the outcome of their actions. The U.S. would prefer to deter China, as it would be an indication of China's inability to fully gain any strategic, orbital 'territory' in outer space. China on the other hand, prefers to gradually inhabit these orbital locations since it would allow it to establish a strong presence in space, as well an increase its power and influence. Table 5 is a simple payoff matrix of the possible outcomes of the game and the payoffs for each player.

Table 5: Simple Payoff Matrix

U.S. Strategy	China Strategy	U.S. Payoff	China Payoff
Deterrence	Salami Tactics	High	Low
Weak Response	Salami Tactics	Medium	Medium
Surrender	Salami Tactics	Low	High

In this game, the U.S. would prefer to play the deterrence strategy, as it would result to the highest payoff. However, incomplete information makes it difficult for the U.S. to ascertain that China will respond to the deterrence strategy in its desired way. If the U.S. plays a weak response, it may lead to a medium payoff for both players, but the U.S. would still be in a position of risk since it could end up losing more ‘territory’ in outer space. If the U.S. surrenders, it would result to a low payoff for the U.S. and a high payoff for China.

Next, we can define our player’s types within the context of Bayesian games and incomplete information. In this case, the players are the U.S. and China, and their types could be their level of commitment to their respective strategies (i.e. deterrence or salami tactics). Let T be the type space, with T= {D,W,S} for the U.S. (representing deterrence, weak response, and surrender), and T = {ST, NST} for China (representing the use of salami tactics, and the non-use of salami tactics). China will either commit to using salami tactics with probability p, or not using salami tactics, with probability 1-p, and China believes that that the U.S. is either committed to deterring them with probability q, or not deterring them with probability 1-q. Given these beliefs, the players’ strategies can be represented as a function of their types and beliefs. In other words, the U.S. strategy is a function of its type and China's belief, denoted as U (t, p), where t is the U.S. type and p is China's belief. The China strategy is a function of its type and the U.S. belief, denoted as C(t, q), where t is China's type and q is the U.S. belief.[44]

Finally, we can define the expected payoffs for each player. The expected payoff for the U.S. is a function of its strategy and China's strategy, denoted as EU(U,C) = ∑p(t) * ∑q(t) * U(t, p) * C(t, q), where t is the type of each player and p(t) and q(t) are the prior probabilities. The expected payoff for China is similarly defined, denoted as EC(U,C) = ∑p(t) * ∑q(t) * U(t, p) * C(t, q). Furthermore, given the expected payoffs, each player can use Bayesian reasoning to update their beliefs and choose their optimal strategies, leading to a Nash equilibrium in the game. In applying this scenario to equation 3b, which takes into account the probabilities (P_i) of each strategy being chosen by the U.S., the expected payoffs (U^i_{si}, U^i_{fi}) for each strategy, the probabilities and expected payoffs of China's strategies for each of its strategies: (Q^i_{qi} ,Q^i_{bi}), and (U^i_{qi}, U^i_{bi}, U^i_{wi}) respectively, and (R^j_r (σ1, σ2)), which represents the uncertain outcome of the game due to the incomplete information, we can deduce that the best strategy for the U.S. would be to play deterrence, as it would lead to the highest expected payoff (U^i_{si}) of high. However, the U.S. cannot be certain of China's response, so a weak response (U^i_{fi}) may lead to a medium payoff for both players, and surrender (U^i_{wi}) would result in a low payoff for the U.S. and a high payoff for China. Alternatively for China, the best strategy would be to

[44] Note that in this game, subgame perfect equilibria can occur when both players maintain certain beliefs about the actions of the other player and act accordingly. For example, if the U.S. believes that China will respond to a deterrence strategy with salami tactics, then the U.S. may choose to play a weak response strategy instead, leading to a medium payoff for both players. This could be a subgame perfect equilibrium if both players continue to play their chosen strategies in subsequent rounds.

use salami tactics, as it would lead to a high expected payoff (U^{i}_{qi}) if the U.S. surrenders, or a medium payoff (U^{i}_{bi}) if the U.S. plays a weak response.

Conclusion

The proliferation of anti-satellite capabilities presents new challenges to the space governance regime. Deterrence, which has traditionally been confined to the nuclear context, is becoming increasingly unable to address the emerging dynamics of the international environment. International conduct is no longer enjoyed solely by nation-states who maintain the highest levels of economic, political, and military power. Non-state actors who seek to upset the status quo are increasingly employing new technologies which can disrupt the operations of society at an unprecedented low cost. In the newest theater of conflict, inferential anti-satellite capabilities are enabling spacefaring actors to gain a disproportionate amount of influence. Moreover, such technologies have substantially altered the cost-benefit calculus of deterrence based on their inherently non-demonstrable nature, causing issues to arise within deterrence theory regarding credibility. Yet, inferential ASAT are the primary means with which conflict in outer space is likely to take form in the immediate to near future, thus emphasizing its salience for the study of deterrence. In order to address these challenges, this thesis proposed an alternative framework that sought to mitigate the effects of inferential ASAT on credibility through the implementation of a NWS treaty mechanism. In affixing hand-tying and burning bridges strategies, the game-theoretic approach to resolving the credibility issue could potentially resolve deterrence failure in situations where the conditions outlined in this thesis are met. The expected utility equations and their modeling provided an empirical analysis that supported this thesis's theoretical arguments and yielded results consistent across both theoretical and game-theoretic literature. Suggestions for future research include an examination of the effects of space situational awareness on signaling strategies, and an in-depth analysis of general-extended deterrence within the context of salami tactics. This discussion is not only meaningful for its theoretical applications, but because its successful implementation implies that deterrence as a theory is highly adaptable, resilient, and will continue to remain relevant in formulation national space strategies going forward.

Appendix

Variable Chart of Expected Utility Models for Inferential ASAT

Variable *all variables have a j/i counterpart	Definition
P_i $1 – P_i$	i's perception of the probability of i succeeding in a conflict with j i's perception of losing
U^i_{si} U^i_{fi}	i's perception of its utility for succeeding (gains) i's perception of its utility for losing (loss)
Q^i_{qi} $1 - Q^i_{qi}$ Q^i_{bi} $1 - Q^i_{bi}$	The probability of i maintaining its current policies in the absence of a challenge by the other actor The probability of i not maintaining its current policies *Q_q= 1.0 so that in a no challenge situation, the adversary maintains its policies i's perception of the probability that its policy will improve i's perception of the probability that its policy will worsen
U^i_{qi} U^i_{bi} U^i_{wi}	i's perception of its utility of maintaining its current policies i's perception of the utility of its policies improving i's perception of the utility of its policies worsening
R^j R^j_r R^j_d	j's response function to i's preemptive action j's response function in terms of inferential antisatellite capability j's response function in terms of demonstrable antisatellite capability
σ_1 σ_2	i's initial strategy j's initial strategy
P_{ikn} P_{ik}	i's probability of success if non-nuclear weapons state (NNWS) alliance k supports i i's probability of success if nuclear weapon state (NWS)alliance k supports i

U^i_{kni}	The utility of NNWS alliance k supporting i
U^i_{ki}	The utility of NWS alliance k supporting i

References

Anderson, R. (2015, June 22). *What would happen if a nuclear-powered shuttle exploded?* http://curious.astro.cornell.edu/about-us/150-people-in-astronomy/space-exploration-and-astronauts/general-questions/929-what-would-happen-if-a-nuclear-powered-shuttle-exploded-beginner

Arce, D. G., & Sandler, T. (2009). Deterrence: Credibility and Proportionality. *Economics & Politics*, *21*(3), 384–408. https://doi.org/10.1111/j.1468-0343.2009.00352.x

Astorino-Courtois, A., & Bragg, B. (2018, February). *Commercial Space Actors: Disruptors or Solid Partners | NSI*. https://nsiteam.com/commercial-space-actors-disruptors-or-solid-partners/

Atherton, K. (2021, September 30). Why DARPA's New Hypersonic Weapon Could Pack Such a Devasating Punch. *Popular Science*. https://www.popsci.com/technology/darpas-hypersonic-missile-test-explained/

BAE Systems. (n.d.). *Space Resiliency: Managing Evolving Threats in Crowded and Contested Environments*. BAE Systems | United States. https://www.baesystems.com/en-us/feature/space-resiliency-managing-evolving-threats

Black, J. (2018, March 12). *Our reliance on space tech means we should prepare for the worst*. Defense News. https://www.defensenews.com/space/2018/03/12/our-reliance-on-space-tech-means-we-should-prepare-for-the-worst/

Black-Branch, J. (2020). Precarious Peace: Nuclear Deterrence and Defence Doctrines of Nuclear-Weapon States in the Post-Cold War Era. In J. L. Black-Branch & D. Fleck (Eds.), *Nuclear Non-Proliferation in International Law—Volume V: Legal Challenges for Nuclear Security and Deterrence* (pp. 323–350). T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-347-4_16

Booth, K. (1984). The War Trap (Book Review). *International Affairs*, *60*(2), 296–297. https://doi.org/10.2307/2619065

Brams, S. J., & Kilgour, D. M. (1988). Deterrence versus Defense: A Game-Theoretic Model of Star Wars. *International Studies Quarterly*, *32*(1), 3–28. https://doi.org/10.2307/2600410

Quarterly

Brams, S., & Kilgour, D. M. (1985). *Optimal Deterrence*. https://www-cambridge-org.ezp-prod1.hul.harvard.edu/core/journals/social-philosophy-and-policy/article/optimal-deterrence/68960D83E7CD09AE3B7B9A51C7B2B0F0

Bueno de Mesquita, B. (1981). *The War Trap*. Yale University Press.

Chow, B. G. (2018). Space Arms Control: A Hybrid Approach. *Strategic Studies: SSQ*, *12*(2), 107–132.

Chris Flaherty. (2023, January 2). *@ChrisFlaherty8 Tweet* [Tweet]. Twitter. https://twitter.com/ChrisFlaherty8/status/1610036042798886913

Cole, A. (2014, May). *Asymmetric Risks in Space*. American Security Project. https://www.americansecurityproject.org/asymmetric-risks-in-space/

Davis, L., & Filip, L. (2015, March 12). *How Long Does it Take to Develop and Launch Government Satellite Systems*. ICEAA. https://www.iceaaonline.com/ready/wp-content/uploads/2014/03/

Davis, P. K., & Stan, P. (1984). *Concepts and Models of Escalation*. RAND Corporation. https://www.rand.org/pubs/reports/R3235.html

De Mesquita, B. B. (1985). The War Trap Revisited: A Revised Expected Utility Model. *The American Political Science Review*, *79*(1), 156–177. https://doi.org/10.2307/1956125

DIA. (n.d.). *Challenges to Security in Space*. Retrieved April 12, 2021, from https://www.dia.mil/Military-Power-Publications/

DIA. (2018). 2018 Nuclear Posture Review. *Federation Of American Scientists*. https://fas.org/issues/nuclear-weapons/nuclear-posture-review/

DIA. (2022). *Challenges to Security in Space*. https://www.dia.mil/Military-Power-Publications/

DoD. (2022). *2022 National Defense Strategy*. U.S. Department of Defense. https://www.defense.gov/News/Publications/

Dotto, C., Lendon, B., & Yeung, J. (2022, December 26). *North Korea's Record Year of Missile Testing is Putting the World on Edge*. CNN. https://ix.cnn.io/dailygraphics/graphics/20221223-nk-missile-range-map/index.html?initialWidth=857&childId=responsive-embed-20221223-nk-missile-range-map&parentTitle=North%20Korea%27s%20record-breaking%20year%20of%20missile%20testing%20is%20putting%20the%20world%20on%20edge%20%7C%20CNN&parentUrl=https%3A%2F%2Fwww.cnn.co

m%2F2022%2F12%2F26%2Fasia%2Fnorth-korea-missile-testing-year-end-intl-hnk%2Findex.html

Downes, A. B., & Sechser, T. S. (2012). The Illusion of Democratic Credibility. *International Organization*, *66*(3), 457–489. https://doi.org/10.1017/S0020818312000161

Eftimiades, N. (2022, May 5). Small satellites: The implications for national security. *Atlantic Council*. https://www.atlanticcouncil.org/in-depth-research-reports/report/small-satellites-the-implications-for-national-security/

Erwin, S. (2019, November 14). Air Force: SSA is no more; it's 'Space Domain Awareness'. *SpaceNews*. https://spacenews.com/air-force-ssa-is-no-more-its-space-domain-awareness/

Every Satellite Orbiting Earth and Who Owns Them. (2022, January 18). Dewesoft. https://dewesoft.com/daq/every-satellite-orbiting-earth-and-who-owns-them

Fearon, J. (2002). Selection Effects and Deterrence. *Intenational Interactions Vol. 28*. https://www-tandfonline-com.ezp-prod1.hul.harvard.edu/doi/abs/10.1080/03050620210390

Fearon, J. D. (1994). Domestic political audiences and the escalation of international disputes. *The American Political Science Review*, *88*(3), 577.

Fearon, J. D. (1997). Signaling foreign policy interests: Tying hands versus sinking costs. The Journal of Conflict Resolution, 41(1), 68–90.

Firth, N. (2019, August). How to fight a war in space (and get away with it). MIT Technology Review, 122(4), 36–39.

Ford, L. W. (2020, May 5). The Trump administration and the "Free and Open Indo-Pacific." Brookings. https://www.brookings.edu/research/the-trump-administration-and-the-free-and-open-indo-pacific/

Freedman, L., & Michaels, J. (2019). The Evolution of Nuclear Strategy: New, Updated and Completely Revised. Springer.

Fuhrmann, M., & Sechser, T. (2014, October). Signaling Alliance Commitments: Hand-Tying and Sunk Costs in Extended Nuclear Deterrence. https://onlinelibrary-wiley-com.ezp-prod1.hul.harvard.edu/doi/full/10.1111/ajps.12082

Garcia, M. (2021, May 26). Space Debris and Human Spacecraft [Text]. NASA. http://www.nasa.gov/mission_pages/station/news/orbital_debris.html

George, A. L., & Smoke, R. (1974). Deterrence in American foreign policy: Theory and practice. Columbia University Press. http://hdl.handle.net/2027/heb00720

Grego, L. (2012, March). A History of Anti-Satellite Programs | Union of Concerned Scientists. https://www.ucsusa.org/resources/history-anti-satellite-programs

Grey, J. (n.d.). Earth satellite | Definition & Facts | Britannica. Retrieved June 5, 2022, from https://www.britannica.com/technology/scientific-satellite

Gutierrez, B. (2021, July 15). Why the U.S. once set off a nuclear bomb in space. Science. https://www.nationalgeographic.com/science/article/why-the-us-once-set-off-a-nuclear-bomb-in-space-called-starfish-prime

Harrison, T., Johnson, K., Roberts, T., Way, T., & Young, M. (2020, March 30). Space Threat Assessment 2020. https://www.csis.org/analysis/space-threat-assessment-2020

Harvey, F. P. (1998). Rigor Mortis or Rigor, More Tests: Necessity, Sufficiency, and Deterrence Logic. International Studies Quarterly, 42(4), 675–707. https://doi.org/10.1111/0020-8833.00102

Hencke, D., & Gibson, O. (2005, December 3). Protest to Libya after satellites jammed. The Guardian. https://www.theguardian.com/uk/2005/dec/03/politics.libya

Howe, E. (2022, November 15). The US’s New Tool for Deterrence Isn’t Ready. Defense One. https://www.defenseone.com/threats/2022/11/uss-new-tool-deterrence-isnt-ready/379750/

Hughes, R. B. (2016). The Autonomous Vehicle Revolution And The Global Commons. The SAIS Review of International Affairs, 36(2), 41–56.

Huth, P. K. (1999). Deterrence and international conflict: Empirical findings and theoretical debates. Annual Review of Political Science, 2(1), 25–48.

Japan-U.S. Security Treaty. (1960, January 19). Ministry of Foreign Affairs of Japan. https://www.mofa.go.jp/region/n-america/us/q&a/ref/1.html

Jaramillo, C. (n.d.). In Defence of the PPWT Treaty: Toward a Space Weapons Ban. Project Ploughshares. Retrieved November 16, 2021, from https://ploughshares.ca/pl_publications/in-defence-of-the-ppwt-treaty-toward-a-space-weapons-ban/

Kimball, D. (2020, October). The Outer Space Treaty at a Glance | Arms Control Association. https://www.armscontrol.org/factsheets/outerspace

Kleiman, M., & McNeil, S. (2012, March 5). The Space Review: Red lines in outer space. https://www.thespacereview.com/article/2038/1

Klein, J. (2020, June 5). Op-ed | To deter attacks on satellites, U.S. needs a strategy to identify bad actors. SpaceNews. https://spacenews.com/op-ed-to-deter-attacks-on-satellites-u-s-needs-a-strategy-to-identify-bad-actors/

Koichi, A. (2016, December). Complex Deterrence Theory and the Post-Cold War Security Environment. NIDS Journal of National Defense. http://www.nids.mod.go.jp/english/publication/kiyo/e2016.html

Krepon, M. (2009). Better Safe Than Sorry: The Ironies of Living with the Bomb. Stanford University Press. https://doi.org/10.11126/stanford/9780804760638.001.0001

Krepon, M. (2013). Space and Nuclear Deterrence (Anti-Satellite Weapons, Deterrence and Sino-American Space Relations, pp. 15–40). Stimson Center. http://www.jstor.org/stable/resrep10894.6

Lafleur, J. (2011, February). Extension of a Simple Mathematical Model for Orbital Debris Proliferation and Mitigation. https://ssdl.gatech.edu/publications/extension-simple-mathematical-model-orbital-debris-proliferation-and-mitigation

Langlois, J.-P. P. (1991). Rational Deterrence and Crisis Stability. American Journal of Political Science, 35(4), 801–832. https://doi.org/10.2307/2111495

Lupovici, A. (2010, September). The Emerging Fourth Wave of Deterrence Theory—Toward a New Research Agenda. https://www-jstor-org.ezp-prod1.hul.harvard.edu/stable/40931133?seq=1#metadata_info_tab_contents

Mann, A., & Pultarova, T. (2022, April 14). Starlink: SpaceX's satellite internet project | Space. Elizabeth. https://www.space.com/spacex-starlink-satellites.html

Matthews, J. (n.d.). The Decline of Commercial Space Launch Costs – Insights to Action | Deloitte US. Deloitte United States. Retrieved May 24, 2022, from https://www2.deloitte.com/us/en/pages/public-sector/articles/commercial-space-launch-cost.html

Mazarr, M. J., Chan, A., Demus, A., Frederick, B., Nader, A., Pezard, S., Thompson, J. A., & Treyger, E. (2018). What Deters and Why: Exploring Requirements for Effective Deterrence of Interstate Aggression. RAND Corporation. https://www.rand.org/pubs/research_reports/RR2451.html

Mearsheimer, J. J. (1990). Back to the Future: Instability in Europe after the Cold War. International Security, 15(1), 5–56. https://doi.org/10.2307/2538981

Miki, R. (2023, January 12). Japan, U.S. extend security treaty into space to protect satellites. Nikkei Asia. https://asia.nikkei.com/Politics/International-relations/Japan-U.S.-extend-security-treaty-into-space-to-protect-satellites

Morgan, F. E. (2010). Deterrence and First-Strike Stability in Space: A Preliminary Assessment. RAND Corporation. https://www.rand.org/pubs/monographs/MG916.html

Morgan, P. (1980). Deterrence: A conceptual analysis. Beverly Hills, Calif. : Sage Publications. http://archive.org/details/deterrenceconcep0000morg

Morrow, J. D. (1985). A Continuous-Outcome Expected Utility Theory of War. The Journal of Conflict Resolution, 29(3), 473.

Morrow, J. D. (2020). Game Theory for Political Scientists. Princeton University Press. http://muse.jhu.edu/book/74150

Most, B. A., & Starr, H. (2015). Inquiry, Logic, and International Politics: With a New Preface by Harvey Starr. University of South Carolina Press. http://muse.jhu.edu/book/42288

Mountin, S. (2014). The Legality and Implications of Intentional Interference with Commercial Communication Satellite Signals. International Law Studies, 90(1). https://digital-commons.usnwc.edu/ils/vol90/iss1/12

Mueller, K. P. (2013). The Absolute Weapon and the Ultimate High Ground: Why Nuclear Deterrence and Space Deterrence Are Strikingly Similar - Yet Profoundly Different (Anti-Satellite Weapons, Deterrence and Sino-American Space Relations, pp. 41–60). Stimson Center. http://www.jstor.org/stable/resrep10894.7

Poponak, N., & Siewert, J. (n.d.). Space: The Next Investment Frontier. Retrieved May 16, 2022, from https://www.goldmansachs.com/insights/podcasts/episodes/05-22-2017-noah-poponak.htm
l

Quackenbush, S. L. (2010). General Deterrence and International Conflict: Testing Perfect Deterrence Theory. International Interactions, 36(1), 60–85. https://doi.org/10.1080/03050620903554069

Quackenbush, S. L. (2011). Understanding General Deterrence. Palgrave Macmillan US. https://doi.org/10.1057/9780230370791

Reuters. (2022, September 30). *Extracts From Putin's Speech at Annexation Ceremony*. https://www.reuters.com/world/extracts-putins-speech-annexation-ceremony-2022-09-30/

Rogin, J. (2021, December 1). Opinion | A shadow war in space is heating up fast. *Washington Post*. https://www.washingtonpost.com/opinions/2021/11/30/space-race-china-david-thompson/

Sadat, M., & Siegel, J. (2022, August 2). Space traffic management: Time for action. *Atlantic Council*. https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/space-traffic-management-time-for-action/

Sagan, S. D., & Waltz, K. N. (1995). *The Spread of Nuclear Weapons: A Debate*. W.W. Norton.

Schelling, T. (1981). *The Strategy of Conflict*. https://www.hup.harvard.edu/catalog.php?isbn=9780674840317

Smallsats by the Numbers 2022. (2022). https://brycetech.com/reports

Space: Investing in the Final Frontier. (2020, July 24). Morgan Stanley. https://www.morganstanley.com/ideas/investing-in-space

Start-Up Space Report 2022. (2022, April 6). BryceTech. https://brycetech.com/reports

Staughton, J. (2016, March 5). *What Would Happen If A Nuke Exploded In Space?* Science ABC. https://www.scienceabc.com/eyeopeners/happen-nuke-exploded-space.html

Su, J. (2010). The "peaceful purposes" principle in outer space and the Russia–China PPWT Proposal. *Space Policy*, *26*(2), 81–90. https://doi.org/10.1016/j.spacepol.2010.02.008

Tarver, E. (2022, March 11). *Horizontal Integration vs. Vertical Integration: What's the Difference?* Investopedia. https://www.investopedia.com/ask/answers/051315/what-difference-between-horizontal-integration-and-vertical-integration.asp

The Great Brazilian Sat-Hack Crackdown | WIRED. (2009, April 28). https://www.wired.com/2009/04/fleetcom/

UCS Satellite Database. (2022, January 1). Union of Concerned Scientists. https://www.ucsusa.org/resources/satellite-database

United Nations. (n.d.). *The Outer Space Treaty*. Retrieved April 15, 2021, from https://www.unoosa.org/oosa/en/ourwork/spacelaw/treaties/introouterspacetreaty.html

United Nations. (1968, July). *Treaty on the Non-Proliferation of Nuclear Weapons (NPT)*. https://www.un.org/disarmament/wmd/nuclear/npt/text/

United Nations. (1976, September 15). *Convention on Registration of Objects Launched into Outer Space*. https://www.unoosa.org/oosa/en/ourwork/spacelaw/treaties/introregistration-convention.html

Van Damme, E., & Weibull, J. W. (1995). Equilibrium in Strategic Interaction: The Contributions of John C. Harsanyi, John F. Nash and Reinhard Selten. *The Scandinavian Journal of Economics*, *97*(1), 15–40. https://doi.org/10.2307/3440825

Vladimir, P. (2022, September 21). *Address by the President of the Russian Federation*. President of Russia. http://en.kremlin.ru/events/president/news/69390

Witze, A. (2022, May 26). *'Unsustainable': How satellite swarms pose a rising threat to astronomy*. https://www.nature.com/articles/d41586-022-01420-9

WOO, K. C. (2020). *Overview of Perfect Deterrence Theory* (Implications of Perfect Deterrence Theory for South Korea, pp. 10–12). Asan Institute for Policy Studies. http://www.jstor.org/stable/resrep26115.4

Zagare, F. C. (2004). Reconciling Rationality with Deterrence: A Re-Examination of the Logical Foundations of Deterrence Theory. *Journal of Theoretical Politics*, *16*(2), 107–141. https://doi.org/10.1177/0951629804041117

Zagare, F., & Kilgour, M. (2000). *Perfect Deterrence*. https://www.cambridge.org/core/books/perfect-deterrence/B0AC9644F74842ACEC1B387A34457F61