



Attribution–NonCommercial–NoDerivs 2.0 KOREA

You are free to :

- **Share** — copy and redistribute the material in any medium or format

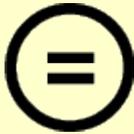
Under the following terms :



Attribution — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



NonCommercial — You may not use the material for [commercial purposes](#).



NoDerivatives — If you [remix, transform, or build upon](#) the material, you may not distribute the modified material.

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

This is a human-readable summary of (and not a substitute for) the [license](#).

[Disclaimer](#) 

국제학석사학위논문

**Explaining North Korean Nuclear
Strategy Under Kim Jong-Un: A
Comparative Study of a Coercive
Bargaining and “Madman” Model**

김정은 체제 하의 북핵 전략: 강압적 거래와 “매드맨”
모델의 비교 연구

2018년 2월

서울대학교 국제대학원

국제학과 국제협력전공

이 성 호

Master's Thesis

**Explaining North Korean Nuclear
Strategy Under Kim Jong-Un: A
Comparative Study of a Coercive
Bargaining and “Madman” Model**

A Thesis Presented

By

Andrew Roskos-Ewoldsen

A thesis submitted in fulfillment of the requirements for the
degree of Master of International Studies in the subject of
International Cooperation

February 2018

Graduate School of International Studies

Seoul National University

Seoul, Republic of Korea

ABSTRACT

Explaining North Korean Nuclear Strategy Under Kim Jong-Un: A Comparative Study of A Coercive Bargaining and “Madman” Model

Andrew Roskos-Ewoldsen
International Cooperation Major
Graduate School of International Studies
Seoul National University

Policy-makers and political scientists agree that North Korea intends to become a nuclear weapons power, but there is a lack of consensus on how North Korea intends to make this goal a reality. For the past few decades, North Korea has often slowed down or sped up its pursuit of nuclear weapons in unexpected and seemingly irrational ways. Explaining North Korea’s varying nuclear strategy has become even more pertinent under the rule of Kim Jong-Un, whose country grows closer to intercontinental and second-strike nuclear capabilities every day. In attempt to evaluate some of the theoretical approaches to this research topic, this study examines and compares the explanatory power of two theoretical models as applied to North Korean nuclear strategy: a coercive bargaining model created by Victor Cha and a “madman” model created by Avidit Acharya and Edoardo Grillo. The coercive bargaining model, based on power transition theory and prospect theory, rationalizes North Korea’s acceptance of risky nuclear strategy choices through its

perception of its own position in the international system. The “madman” model, based on two-sided incomplete information game theory, extols the logic of North Korea “acting crazy” to illicit greater strategic outcomes for itself. A congruence test is used to see whether the predictions of the models co-vary with North Korean transitions in nuclear strategy from fast to slow, or vice versa, from the years 2011 to 2017.

Keywords: Nuclear strategy, International Security, International Relations, North Korea, United States

Student Number: 2016 – 25053

TABLE OF CONTENTS

ABSTRACT	i
TABLE OF CONTENTS	iii
INTRODUCTION	1
LITERATURE REVIEW	5
RESEARCH DESIGN	24
COERCIVE BARGAINING MODEL	36
THE “MADMAN” MODEL	57
CONCLUSION	93
APPENDIX	97
REFERENCES	107
ABSTRACT (KOREAN)	112

Introduction:

North Korea's goal to become a nuclear state has been a critically important issue both practically and theoretically for decades. The threat of North Korea possessing operational nuclear weapons endangers millions of lives, and the state's decisions regarding those weapons have enormous security implications for the Northeast Asian region and beyond. In recent years, the stakes have not changed substantially since Kim Jong-Un became the state's leader, but rather intensified as North Korea has advanced its nuclear weapons capabilities and aggravated its relations with other regional actors. Policymakers, national intelligence, and scholars alike have a vested interest in explaining why and how North Korea is trying to achieve its nuclear ambitions. The why question is both important and considerably challenging to study; North Korea is a famously opaque nation that is renowned for blocking quality information on its activities. Although numerous studies have explored the topic with diffusive results, a general consensus seems to have formed in the formal literature: whether for deterrence purposes, regime survival, the personal pride of its leaders or other reasons, North Korea has shown a consistent preference to go nuclear. Scholars using various methodological approaches and theoretical assumptions agree: North Korea wants the bomb (Cha and Kang, 2003; He and Feng; 2013; Hymans, 2008). Perhaps more perplexing is the how question, or what actual strategy North Korea has employed to reach its goal to go nuclear over the years. North Korea has been described as having "two faces", varying between cooperative and

rebellious stances on its nuclear program, and is notorious for its bizarre and seemingly irrational behavior typified by aggressive nuclear-related tests or threats that could provoke outside intervention and the entrance into and eventual abandonment of denuclearization agreements with key rival states (Cha and Kang, 2003). This variance has also presented a significant challenge to the predominant IR theories while simultaneously creating an opportunity for alternative theories. Neorealist and neoliberal institutional theoretical explanations that rely on traditional rational actor model assumptions have had difficulty explaining North Korea's acute changes in nuclear behavior; if North Korea wants the bomb, then why has it stopped a full pursuit of nuclear proficiency and signaled a willingness on multiple occasions to negotiate and denuclearize? In other words, why does North Korea's nuclear strategy vary? Subsequent research focusing on small adjustments to rational actor assumptions and decision-making under conditions of risk has provided innovative and convincing analysis (Cha and Kang, 2003; He and Feng, 2013). This paper seeks to continue in this line of research, applying and testing three theoretical models in a more contemporary context to explain perplexing and varying nuclear behaviors by treating them as components of North Korea's overall nuclear strategy to obtain nuclear weapons proficiency. Thus, this paper is a study of nuclear strategy, using a security studies' definition of strategy where intelligence, diplomacy, and military strength are used to determine outcomes by a state and narrowed to an analytical focus on activity related to North Korea's nuclear weapons program (Luttwak, 2009). The study includes the

application of Victor Cha's (2003) coercive bargaining model and Avidit Acharya and Edoardo Grillo's (2015) "madman" model to the case of North Korean nuclear strategy from 2011-2017, capturing both the variance in strategy we seek to explain and the years of Kim Jong-Un's rise to power and subsequent role as the country's supreme leader. The results will hopefully provide distinct answers to the how question discussed above, and will allow us to evaluate and compare the effectiveness of these theoretical models' ability to explain North Korea's nuclear strategy under the current Kim Jong-Un regime.

The models were selected for their distinct independent variables explaining North Korea's varying nuclear strategy and their ability to produce testable hypothesis, and to use their results to compare and evaluate each model's explanatory powers. The first two models, the coercive bargaining model and the political legitimacy-prospect model, have been applied to North Korea before and their application to a contemporary set of cases can help evaluate their utility as the theories have aged. The coercive bargaining model, as formulated by Victor Cha, uses a combination of power transition theory and prospect theory to explain that North Korea chooses its nuclear strategy based on its risky position as a weak, revisionist state surrounded by stronger, rival states (2003). North Korea uses its pursuit of nuclear weapons, and accepts the subsequent risks of doing so, as a bargaining tool to coerce its rival states into revising the status quo that they would otherwise be unable to affect, which explains the "two faces" of North Korean nuclear strategy. Alternatively, the "madman" model is a game theoretic model of war and peace

that expounds the dynamics and benefits of using a “madman” strategy, or acting as if one is crazy to force another country to agree to terms in bargaining that a rational state would be unable to achieve (Acharya and Grillo, 2015). In other words, this model extolls the costs and benefits of “acting crazy” to get your opponent to give in to exorbitant demands, and in the North Korean context explains the how Kim Jong-Un could be using the “madman” strategy when making its nuclear decisions.

It is important to note that this paper does not intend to make a definitive claim on the explanatory power of the models or the theories that they are based on in general, but merely to compare their explanatory power for this case study. Much of the scholarly work on North Korea to date acknowledges the possibility of overly deterministic theorizing for a notoriously convoluted case, so the same logic should be applied here in that the findings of the study should only be seen as a preliminary attempt to explain North Korean nuclear strategy and to expand research possibilities in the IR field. As such, this paper will use a congruence test as the basis for testing the models, or observing whether North Korea’s nuclear strategy is actually congruent with the predictions of the models. This means the findings do not indicate causality and leaves open the possibility that other variables are responsible for producing North Korean nuclear strategy in the case study, but as He and Feng (2013) explain, it is extremely difficult to access the kind of information on North Korea required to rule out alternative explanations and provide enough evidence for acceptable certainty of findings.

This paper will follow the following structure: the first section provides a basic literature review of scholarly work relevant to the research questions of how state's pursue their nuclear goals and what factors impact nuclear strategy. This section will also briefly summarize how these theoretical developments have manifested in studies of North Korea and its nuclear strategy. The second section introduces the research design, describing the case parameters and reasons for its selection along with the details and justification of the dependent variable. This section uses a framework to operationalize the dependent variable that the models will try to explain in the rest of the paper. The fourth and fifth sections will explain the theoretical background and assumptions of the respective models, list their hypotheses and apply them to the case study, and provide an analysis. The paper's final section will present a conclusion on the study's contributions, limitations, and implications for future research.

Literature Review:

The main question this section is concerned with is the following: how do states formulate their nuclear strategies? A review of the formal literature offers a bounty of answers, and for the sake of brevity this literature review will be broken up into three subsections: defining strategy, which will explore the development of the term as it is used in social science; structural theories of IR, which will detail the predominant theoretical

premises for both why and how states pursue nuclear weapons; and alternative theories on North Korean nuclear strategy, which will detail how certain research has developed or deviated from the previously accepted premises in significant ways to study North Korean nuclear strategy. In sum these subsections will provide a necessary overview of the theoretical field relevant to the research question and some significant research on North Korean nuclear strategy to date.

Defining Strategy

This paper is a study of nuclear strategy, but what exactly does nuclear strategy mean? The exact definition of the concept used in this study will be covered later, but firstly, it is worth briefly noting the historical development of the concept of “strategy” as a subject of research by social scientists. This will help to frame the discussion of the formal literature and the various theories that answer the questions of why and how states go nuclear.

Carl von Clausewitz and his book *On War* have been extremely influential in the development and study of the concept of strategy (Hart, 1967). In his book, Clausewitz discusses strategy within the context of war, as he was most concerned with studying the nature of war and how it was conducted (Von Clausewitz, 1832; Hart, 1967). The art of war, according to Clausewitz, was achieving a political goal through military conflict with another country and this attainment of the political goal of war could be achieved through

the combination of two different levels of war: strategy, which entailed the use of battles; and tactics, which was the art of winning battles (Von Clausewitz, 1832). Thus, strategy is often associated with wartime and security studies, higher stratification of using military force compared to tactics, and the use of military force as a means to an objective end (Hart, 1967).

B. H. Liddell Hart is also considered a key contributor to developing the term strategy in security studies literature, juxtaposing his own thoughts on the concept with Clausewitz directly. While Hart maintains the same analytical focus as Clausewitz on war and its nature, he differs significantly in his definition of strategy and what the term applies to. In his book *Strategy*, Hart claims that Clausewitz's definition centered on battles unnecessarily ties the end of war, or the achievement of a political goal, with fighting the enemy (1967). On the contrary, Hart argues that strategy can be expanded to include how military decisions are made to avoid the costs of war as much as possible, and generals or military leaders use their military force to achieve the goals of war as efficiently as possible. His subsequent definition of strategy, "the art of distributing and applying military means to fulfill the ends of policy", reflects his attempts to broaden the concept to more accurately encompass military-related decisions made during war (Hart, 1967). Hart also made a lasting contribution to the field with his introduction of the concept of grand strategy, which broadens the resources considered under strategy beyond the military. Grand strategy, or "policy which guides the conduct of war", covers a

government's entire approach to fulfilling the policy goals of war, including economic and diplomatic resources that can be employed for these ends (Hart, 1967). This concept further stratified the levels at which a government can bring about its policy objectives during war. Grand strategy and its manifestations has become commonplace in security studies in attempting to paint a full picture of a government's strategy for attaining a political goal.

Hart's definition of strategy and grand strategy have laid the foundations for other scholars to expand the concept further, such as the work of Edward Luttwak. In his book *Strategy, The Logic of War and Peace*, Luttwak asserts that studying grand strategy in wartime alone is folly, and further points out that strategy can alter depending on the logic presented in wartime versus peacetime (1987). Luttwak went on to update and expand the definition of grand strategy, extending the analytical focus beyond distinctions of war and peace to advocate a certain level of government coordination of resources. Luttwak states: "Grand strategy is simply the level at which knowledge and persuasion, or in modern terms intelligence and diplomacy, interact with military strength to determine outcomes in a world of other states with their own "grand strategies"..." (2009). This represents a significant step in the clarification of the term grand strategy, as well as the lower-level term strategy, in acknowledging strategy as the government's use of its available resources to achieve a goal.

As the concept of strategy has expanded beyond the military, and then beyond the dichotomy of wartime and peacetime, we can arrive at a defensible idea of what strategy looks like. Based on these progressions in definitions, it's reasonable to think of strategy in the Luttwak definition of a state's comprehensive efforts to achieve an outcome. In considering nuclear strategy, we can then logically point to a government's coordination and use of resources to fulfill its nuclear goals, whether that be through economic, military, or diplomatic means.

Structural Theories of IR

To answer the question of how states formulate their nuclear strategies, it is important to first know the fundamental assumptions we make about states, the way they think, and why they desire nuclear weapons in the first place. Beginning with mainstream international relations, there are numerous theories explaining state behavior within what Graham Allison (1999) describes as the rational actor model or paradigm. Most notably, neorealism and neoliberal institutionalism have become predominant in the IR field for their comprehensive and systematic explanations of the character of international life (Allison and Zelikow, 1999). These theories have been convincingly and consistently used to explain state behavior and action in the international environment in countless cases and studies, and have demonstrated great value over and over as parsimonious, generalizable, and reproducible theories. In other words, neorealism and neoliberal institutionalism

within the rational actor paradigm form the base assumptions for most scholarship about states and their behavior, including their nuclear strategy. While these two schools have significant differences, neoliberal institutionalism shares most of the core assumptions of neorealism with an added assumption about the structural role institutions can play in constraining state behavior in addition to anarchy. Since this emphasis on institutions mainly concerns the circumstances underlying international cooperation, and considering the two theories are described as two-sides of the same coin (Allison and Zelikow, 1999), this review will mainly focus on core neorealist assumptions due to their direct relevance to a state's desires for nuclear weapons and their subsequent strategies.

Neorealism seeks to explain international events by taking the foundations of classical realism and adding emphasis on the structural aspects of the international system. According to Graham Allison and Philip Zelikow (1999) there are five main tenets of neorealism: (1) nation-states are unitary actors, and the key units in international relations; (2) states are rational actors who analyze decisions based on a cost and benefits analysis and choose the best option; (3) the international environment is naturally anarchic and jungle-like where all are actors compete with one another to survive and thrive (4) the true goal of states is to gain security and relative power over other states, states act politically to achieve these goals; (5) system-level factors in the international system determine the character and consequences for international life.

Kenneth Waltz, one of the pioneers of neorealism, argues that anarchy is the most important and defining characteristic of the international environment. Waltz's core claim was that anarchy and differences in relative power of states is the primary causal mechanism for international actions taken by states. Within Waltz's strand of neorealism, states are all equal and identical in nature save for one important value: the aggregate power of states. The differences in aggregate power are naturally produced by the anarchic system. Classical realists like Hans Morgenthau first developed this idea of anarchy but did not place the emphasis on structure that Waltz did. Waltz proposes that because the anarchic system creates differences in power, the system will try to balance itself as self-interested and rational states try to secure their own safety and power relative to other states (2001). While Waltz makes the assumption that states are self-interested and rational in seeking power and security, it is the system that determines state behavior. Consequently, the pressures of the international system dictates that states are perpetually seeking to rebalance (Waltz, 2001). In sum, Waltz's idea is that the system is constantly trying to balance itself, and this leads to assumptions about state behavior. Due to differences in aggregate power, when states are faced with a more powerful state these systemic pressures force states to counterbalance. Neorealist scholars are concerned with how states try to counterbalance, and primarily focus on military and economy as measures of power. Waltz argued that states improved security primarily through military

power and alliances, with the strongest security measure being nuclear weapons (Waltz, 2001).

This forms the basis for what is known as balance of power theory, where survival-seeking states are constantly responding to power imbalances in the anarchic international system (Waltz, 2001). Great and weaker powers use internal military build up and alliances for their own security, and thus alignments of states are temporarily made which determine the character of international politics. The international system, depending on the alignment of powerful states, are most commonly known as unipolar, bipolar, and multipolar systems (Waltz, 2010). This theory explains a state's desire to go nuclear given that nuclear weapons represent the greatest security measure, as well as informing us as to the strategy a rational nation-state is expected to take in the face of anarchic pressures: when nuclear weapons represent the most efficient path to security, a state should pursue nuclear weapons development as quickly and efficiently as possible. Thus, a state should be expected to make its nuclear decisions and carries out its policies with this strategy of efficiency in mind, and this expectation has been widely accepted without much inspection (Hymans, 2006).

Other scholars have slightly modified Waltz's assumptions of neorealism and formed different strands, but focus on the same core assumptions. Concerning the state neorealist scholars agree that: unitary states are the principal actors in international life, that these states are rational in calculating possible decisions and choosing the maximum

valued option, and that their main objectives are survival and power. They also agree that the system is the causal mechanism for state behavior, and that anarchy and differences in power define the international system. From these variations that broaden and deepen the explanatory power of neorealism have emerged. For instance, John Mearsheimer's (2001) neorealist strand "offensive realism" makes the same core assumptions as Waltz did but conceived how the distribution of power in the system affected states differently. Waltz argued that the anarchic system will force states to try to achieve power for the sake of survival relative to other states, and once enough power and security is attained to ensure survival the state will be content. In contrast Mearsheimer argued that the pressure of the international environment makes it impossible for states to know whether they have attained enough power to ensure survival so they will therefore continue to maximize their utilities and relative advantage over others. This constant offensive state behavior is still caused by the anarchic system, but its effects on state behavior are different than the more defensive neorealism of Waltz.

Stephen Van Evera (1984) also contributed to the discussion of offense versus defense in states based on technological differences in states. Technological limitations or advances might also determine whether a state is building up power to defend itself or dominate others. This distinction is important for how other states respond to their increased offensive or defensive capabilities. Stephen Walt (1987) added that state behavior and agency were important alongside structural factors. Walt argued that how

states interpret the pressures of the system is important for the establishment of intentions and behavior. States respond to systemic threat, not power imbalance, and such threats are determined by objective military power but also state intention.

Power transition theory is often considered a distinct research paradigm independent from neorealism, but it borrows many of its core assumptions from neorealism and provides a structural explanation of international politics distinct but comparable to balance of power theory (Organski, 1958; DiCicco and Levy, 1999). Similar to balance of power theory, power transition theory assumes nation-states as rational, unitary units, focuses its analysis on power differences between states, and points to structural variables as the key to explaining state behavior. As in balance of power theory, the international system globally and regionally often takes a hierarchal form based on the alliances and military build ups of nation-states, with the key difference being that in power transition theory states do not necessarily act out of purely survival-seeking preferences (DiCicco and Levy, 1999). In addition to the minimum-level goal of security, states can act based on their satisfaction with the current power alignment of the international system and whether it benefits them or not, determining whether they are status quo or revisionist states (Tammen et al., 2000). In other words, in addition to or instead of the structural variable anarchy depending on the use of the theory, states are affected by another structural variable: the international system under a dominant nation. A dominant nation forms a system that perpetuates its power preponderance through military

build up and alliances, and states are either in favor of this system (status quo) or against it (revisionist) (Tammen et al., 2000). Thus, nuclear weapons can represent a rational response for revisionist states in particular to respond to structural differences in power by building up their own military capabilities (Cha and Kang, 2003).

In terms of nuclear strategy, these developments in neorealist theory point to an increasing number of variables that states respond to when deciding whether to go nuclear or not and how to do so if it becomes a rational choice. States, in the neorealist tradition, can respond to structural variables such as anarchy or the international system through offense/defensive orientations, perceptions of threat, and satisfaction with the status quo when deciding whether to go nuclear or not. Still, the expectation remains that once a state decides to go nuclear, they pursue nuclear capabilities as directly and efficiently as possible while it is rational to do so.

Three major observations can be established based on these various strands of neorealism. Neorealism (as well as neoliberal institutionalism) is a top-down theory that contends the anarchic nature of the system and aggregate differences in power are the causal mechanisms of international events. As Keith Shimko (1992) summarized: “while there is disagreement among neorealists about what exactly anarchy causes, there is agreement that anarchy causes it.” Secondly, neorealism also assumes that states are unitary actors who are rational. In regards to nuclear strategy, this means that the decision to go nuclear must be a rational one where a costs and benefit analysis reveals pursuing

nuclear proficiency as the most efficient security choice. Thirdly, while structural theories answer the question of why states pursue a nuclear weapon, they do not subsequently spend much analytical attention to how a state achieves its nuclear goals once they decide to go nuclear. It is assumed that when a combination of rationality and the effects of systemic variables like anarchy produce the situation where it is in a state's best security interest to pursue nuclear weapons, then the state's nuclear strategy should reflect the most direct path of that pursuit.

A significant example of this structural approach applied to North Korea and its nuclear behavior can be seen in Victor Cha and David Kang's *Nuclear North Korea: A Debate on Engagement Strategies* (2003). David Kang in particular employs the assumptions of balance of power theory to explain how the growing gap in military capabilities between North Korea and the rival states surrounding them, mainly the US and South Korea, has caused North Korea to seek deterrence by going nuclear (2003). Kang empirically supports this structural approach by comparing the economic size, defense spending, and military capabilities of North Korea to other key states to display the growing power gap in the international system, while also arguing that more domestic or psychological approaches to the case are basically ad-hoc explanations due to lack of information and an inability to prove different outcomes in a counterfactual scenario (Cha and Kang, 2003). This international system and its asymmetrical power alignment stacked against North Korea support the idea that it is rational for North Korea to close the gap

between itself and its rivals, or balancing, through the most efficient means possible: nuclear weapons capabilities. Kang notes that while North Korea's attempts to balance may seem threatening, the deterrence of the United States and its allies in turn remains consistent in its ability to marginalize the rationality of a preemptive use of force by North Korea (Cha and Kang, 2003). Kang ends his appraisal of the North Korean situation by pointing to neoliberal institutionalism, or the ability to constrain North Korea's behaviors through systemic incentives, in order to deescalate the tensions on the peninsula (Cha and Kang, 2003). This use of structural theory convincingly answers why North Korea would seek nuclear weapons capabilities, but as other authors have pointed out, Kang fails to address the variance in strategy North Korea employs to reach its nuclear goals (He and Feng, 2013).

Alternative Theories on North Korean Nuclear Strategy

While these structural explanations of state behavior have been effectively deployed in many cases, the existence of some atypical cases of nuclear strategy have resulted in theoretical approaches that modify or deviate from the baseline theories described above. In particular, North Korea's past erratic nuclear behavior indicated an inconsistent and/or varied nuclear strategy that is not easily explained. This subsection will summarize how scholars have modified or deviated from the structural approaches above

in significant ways, and subsequently how scholars have explained North Korea's erratic nuclear strategy.

In the book *Nuclear North Korea: A Debate on Engagement Strategies* discussed above, Victor Cha approaches North Korean nuclear strategy with a structural theoretical argument similar to that of his counterpart David Kang (2003). Instead of employing balance of power theory, Cha uses power transition theory to explain why North Korea wants to go nuclear (Cha and Kang, 2003). The appraisal of North Korea's status in the international system is somewhat similar in that Cha empirically supports the idea that the power odds are stacked against North Korea, and thus it should rationally seek to revise the status quo. Cha takes a pioneering step; however, in arguing that in some situations of high risk such as North Korea's, a state cannot always perform a perfect cost and benefit analysis as expected of a rational actor. Cha argues that prospect theory, an empirically supported psychological theory on decision-making under conditions of risk, is a logical substitute in this case due to the high risks and short-term time horizons North Korea must operate in under such difficult systemic pressures (Cha and Kang, 2003). According to prospect theory, which was first developed in laboratory experiments by Daniel Kahneman and Amos Tversky (1979, 2000), actors make their decisions based on a reference point that frames, or splits, their available choices into a domain of gains or a domain of losses. Depending on the domain a choice is framed in, an actor responds differently to the risks and probabilities associated with a choice. In other words, actors treat prospective win and

losses differently. Actors, for instance, are known to be risk-averse in a domain of gains and risk-acceptant in a domain of losses, and this implies key decision-making differences compared to the rational actor who does not rely on a reference point for help in evaluating choices (Kahneman and Tversky, 1979). Cha takes this theory and combines it with power transition theory, using North Korea's structural position in the international system as the reference point, and splits North Korea's nuclear strategy into two options, or "two faces", where they can choose to try to engage with other states and institutions diplomatically or unilaterally pursue its nuclear weapons development (Cha and Kang, 2003). Cha asserts that due to North Korea's constantly decreasing position in the international system, North Korea frames its nuclear strategy choices in a domain of losses, meaning they are more risk-acceptant. Thus, North Korea uses its aggressive nuclear weapons development, threatening diplomacy, and military demonstrations, despite the high risk of loss through sanctions, intervention, or war involved with this strategy, to try to provoke the United States and South Korea into compromises that revise the power status quo of the region in what is known as "coercive bargaining" (Cha and Kang, 2003). This marks a departure from the typical rational explanation by arguing that it may in fact be rational for North Korea to use a varying nuclear strategy that accepts high risks in pursuing its nuclear goals; a strategy not normally expected by a rational state. This argument that North Korea is doing the best it can to improve its position in the international system under difficult

circumstances is a compelling example of how mainstream IR theory has been modified to explain North Korea's strategic variance.

Following in Cha's footsteps of using prospect theory as a reasonable substitute for rational actor assumptions, Kai He and Huiyun Feng's (2013) political legitimacy-prospect model uses neoclassical realism and prospect theory to analyze the first and second nuclear crisis in North Korea. Like power transition theory, neoclassical realism is a structural realist theory on international relations and serves as the theoretical underpinning of the explanation to North Korea's varying nuclear strategy. Neoclassical realism shares many of the basic assumptions of structural realist theory, agreeing on the critical point that the distribution of power in the international system is the main independent variable for shaping a state's foreign policy, and thus its strategy (He and Feng, 2013). The main difference comes in how this system-level variable affects states, or put in another way what level of analysis is important to study. Neorealist theory focuses solely on "third image", or system-level variables, while neoclassical realist approaches assert that system-level variables like anarchy can also be responded to indirectly through "intervening variables" at the "first image" (individual) and "second image" (domestic) levels of analysis (He and Feng, 2013). In applying this to North Korean nuclear behavior, or what states do, He and Feng argue that North Korea responds to the power distribution of the international system through "political legitimacy", or external and internal threats to the North Korean regime's authoritative control of the country (2013). Invasion from a

rival state like the United States, for instance, represents an external threat to political legitimacy while famine and economic decay might represent internal threats to North Korean leaders' political control and security. He and Feng go on to describe North Korea's two methods of pursuing nuclear weapons: "publically and provocatively" or "secretly and deceptively", with the former considered high-risk strategy and the second low-risk (He and Feng, 2013). They argue that North Korea chooses between the two based on whether it thinks its political legitimacy is threatened or not, using the first and second nuclear crises to demonstrate how North Korea's high external security threats coincide with its aggressive nuclear policies and how its deceptive bargaining behavior inversely coincides with North Korean leaders having strong political legitimacy (He and Feng, 2013). Compared to Victor Cha's neorealist approach, this case study adds a significant analytical focus on North Korea's decision-making and leadership while maintaining the assumption that systemic variables are at the heart of North Korea's nuclear strategy.

The final research program that will be introduced here is crisis bargaining. Although there are a number of theories and paradigms that deal with strategy and nuclear strategy, crisis bargaining in particular has potential for examining North Korea because of the crisis-like nature of its disputes with other actors over its nuclear weapons program and the constant bargaining-style negotiations North Korea conducts regarding its nuclear weapons. Crisis bargaining as pioneered by Powell (1987), Banks (1990), and Fearon

(1995) focuses on the interactions between sovereign states and the dynamics of power, interest, cooperation, and conflict inherent in those interactions (Snyder and Diesing, 2015). By studying the specifics of state-to-state interaction, particularly in a “crisis” described as “a sequence of interactions between the governments of two or more sovereign states in severe conflict, short of actual war, but involving the perception of a dangerously high probability of war”, we can subsequently learn about larger subjects such as international politics and relations (Snyder and Diesing, 2015).

As James Fearon’s influential paper “Rationalist Explanations of War” (1995) exemplifies, crisis bargaining generally involves the use of game theory to mathematically map out the possible choices, outcomes, payoffs, or penalties of interactions between states in a crisis. Game theory, first developed by Von Neumann and Morgenstern (1945), is a theory on the behaviors of decision makers in “games” or situations of strategic interdependence. Game theory allows political scientists to precisely define and test theoretical assumptions through mathematical models, and then applying these models to case studies to see how accurately the empirical evidence matches the model. James Fearon’s work demonstrated how game theory could be used in the context of crisis bargaining to test how rational states will strategically act to get what they want when their interests conflict with another state, and how these interactions can lead to various bargaining agreements or ultimately war (1995). Fearon’s paper demonstrated the conditions necessary for war, even if all of the conflicting, rational states would both

prefer to avoid the extreme costliness of war, and in doing so compellingly modified realist assumptions on the role of anarchy in conflict and also set a standard for research on how states behave when they disagree enough on an issue to the point that use of force is a serious possibility. Much of the crisis bargaining research paradigm has subsequently copied or modified this approach by writing their own models to investigate various theoretical assumptions (Acharya and Grillo, 2015). Significantly, crisis bargaining research and game theoretic models have investigated the assumptions of systemic-level theories such as predictive power of balance of power theory in crisis situations amongst others (Fearon, 1994; Snyder and Diesing, 2015). Crisis bargaining research represents a departure from the mainstream structural IR approaches discussed above because of their dynamic study of rational states' strategy in interactive situations. This allows for depth and comprehensiveness in its explanations of strategy, a subject that is mostly ignored or assumed away in other schools of research, while also incorporating the theoretical assumptions of IR theory. In terms of North Korea's nuclear strategy, the country is often used as a member of large number case study applications of game theoretic models and is regarded as a strong candidate for in-depth case-study research due to its history of nuclear crises and ongoing conflict of interests over its nuclear weapons program (Acharya and Grillo, 2015).

Research Design:

Case Selection

North Korea's nuclear strategy under Kim Jong-Un is a natural target for an in-depth case study. Most of the significant analysis on North Korean nuclear strategy has focused on the first and second nuclear crises, so a contemporary study has the potential to illustrate if and how their nuclear strategy has changed under new leadership while also testing how durable different theories' explanations are. Additionally, Kim Jong-Un's position as supreme leader is unlikely to change in the short term and an analysis of his country's nuclear strategy under his regime can potentially contribute to an ongoing security issue (Hee-gwan Chin et al., 2013). Findings from the analysis might be generalizable as well, as North Korea can exemplify patterns in strategy that are common in other places as well, such as situations where state decision-making is typified by high risk or prior crisis. Methodologically, while quantitative, and large number studies could also be effective in studying nuclear strategy, the models being used in this paper are more conducive to a comparative approach. An in-depth case study approach to the North Korean case allows the paper to consider a diverse range of independent variables, and can ultimately lead to the deductive application and evaluation of each model's assumptions.

But why study North Korean nuclear behavior under Kim Jong-Un specifically? This paper focuses on North Korean nuclear behavior from the years 2011 to 2017,

beginning with the year Kim Jong-Un is known to have begun his ascent to power and subsequently established rule over the North Korean government up (Hee-gwan Chin et al., 2013; He and Feng, 2013; Chanlett-Avery, Rinehart, and Nikitin, 2016). This time period was chosen for a few reasons. Firstly, as mentioned previously, two of the models being used have previously been applied to North Korea and we could benefit from the application of these models to a fresher case. Secondly, as information on North Korea has become more reliable over time, albeit inconsistently, it makes sense to use a case with the most empirical evidence available to substantiate independent variable claims. Most critically, as found in previous time periods of North Korean history, North Korean nuclear strategy under Kim Jong-Un displays an observable variance that can be operationalized as dependent variables. The strategic variance in North Korea is extremely valuable as a research puzzle because it so difficult to explain, and therefore we can rigorously test the assumptions of theories by assessing how effectively they can explain such a difficult a problem. The three theoretical models used in the paper can use their assumptions to tell a causal story that results in the observed variance, and allows us to evaluate how well they performed against expectations.

Dependent Variable

The introduction established that the variance in North Korean nuclear strategy under the Kim Jong-Un regime is the dependent variable, but it is important to flesh out

exactly what this means. It is critical to make sure that the way we measure the variance is valid and reliable, or ensuring that our dependent variable is measuring what we want to measure and is consistent with its measurements.

This paper is concerned with North Korea's nuclear strategy under Kim Jong-Un so far in that it displays a tendency to vacillate between a willingness to cooperate with international actors on nuclear disarmament and a converse non-willingness. In attempt to capture the timing and dynamics of these two strategies, North Korean nuclear strategy as the dependent variable has been operationalized into a dichotomous variable with the following possible values: fast and slow. This variable isolates the subject of the study, which is strategic variance, and assumes that North Korea has a preference for nuclear weapons that it is trying to fulfill. Since it is beyond the scope of the paper to investigate why North Korea wants nuclear weapons capabilities, or if it truly really wants them, or what it hopes to do with them, the dependent variable is simply a reflection of the two observable ways North Korea attempts to reach its goal to become a nuclear state. North Korean nuclear strategy can be categorized as fast, fast but slow-signaling, slow but fast-signaling, or slow using the following two-variable framework, which attempts to comprehensively capture the variance we are trying to explain: (1) speed; (2) signaling. Each variable in the framework has a dichotomous value choice that can be determined qualitatively based on evidence on the North Korean government's activity regarding its nuclear weapons. Evidence for determining the values of the dependent variable are drawn

from secondary academic sources, verified news outlets, intelligence reports, and policy papers.

Speed is a straightforward measure of one key component of North Korea's nuclear strategy. It refers to the speed and manner in which North Korea pursues its goal to become a nuclear state, or to attain a stockpile of nuclear weapons with second strike and intercontinental firing range capability (Hee-gwan et al., 2013). Speed of nuclear weapons development can be measured qualitatively as either fast or slow values. Slow indicates that North Korea is not openly pursuing nuclear weapons production. It could be secretly using its nuclear programs or not using them at all, but ultimately North Korea is not actively using its nuclear facilities and resources to pursue nuclear weapons development in ways obvious to the outside world via official inspection, intelligence, or espionage. Fast indicates that North Korea is openly and fully using its nuclear facilities, meaning that their nuclear activity is obvious upon inspection, intelligence collection, or espionage of their activity. This variable is designed to capture the natural disposition of North Korea's nuclear efforts: is the state pursuing nuclear weapons development openly regardless of cost, or in other words are they pursuing nuclear weapons as fast as possible? Activities like testing nuclear weapons, opening new nuclear facilities, or public displays of nuclear military capability would indicate a fast value while shut down nuclear facilities, absence of nuclear-related tests observable to outside actors, and adhering to denuclearization agreements would constitute verifiably slow nuclear strategy.

The signaling variable is taken from the crisis-bargaining literature, in which James Fearon (1997) explicates how states use costly signals to express their foreign policy interests credibly to other states, whether in a crisis situation or in grand strategy. Costly signals, or “messages, gestures, or actions that are costly enough that only an actor of a certain type would be able, or willing, to them carry out”, can be considered credible if it is costly to the state to back down or make empty promises (Fearon, 1997; Yarhi-Milo, Kertzer, and Renshon, 2017). Thus, a state receiving the signal can determine that the sending state’s signal legitimately represents its foreign policy interests because the costs associated with the signal would deter an unresolved or bluffing state from making it. Fearon (1997) indicates two ways in which a state can signal its foreign policy intentions to another state: tying hands or sinking costs. Tying hands “means taking an action that increases the costs of backing down if the would-be challenger actually challenges but otherwise entails no costs if no challenge materialize”, and typically take the form of public official statements made by the state that would affect the state’s prestige and reputation if it were not carried out (Fearon, 1997). Sinking cost signals are “actions that are costly for the state to take in the first place but do not affect the relative value of fighting versus acquiescing in a challenge...” indicating that the ex ante, or before the outcome, costs of the signal are high enough to be deemed credible (Fearon, 1997). Fearon (1997) offers arms build up and military demonstrations of power as examples of this ideal type of signaling. In sum, there is empirical evidence to show that states can credibly

signal its strategic intentions by making costly signals, or creating audience costs that could be suffered by not following through with a threat or offer (Fearon, 1997). As other works have shown (Jervis, 1982; Weisiger and Yarhi-Milo, 2015), reputation is a significant factor for states and affect how carefully they send and use costly signals. We can therefore reasonably use North Korea's costly signals as a component in North Korea's nuclear strategy, observing the messages North Korea wishes to convey to other actors about its nuclear program.

The signaling variable in this framework can have two valuables determined qualitatively: fast-signaling or slow-signaling. If North Korea makes a costly signal to other actors about its nuclear intentions via tied hands or sunk-cost signals, then they can be categorized as signaling whether North Korea is seeking to speed up or slow down its nuclear weapons development. The fast-signaling value indicates that North Korea is signaling that it intends to speed up its nuclear weapons to a full, open pursuit. This might include threats to test or use missiles or nuclear weapons, leave denuclearization negotiations or agreements, or start war over their sovereign right to pursue nuclear weapons; examples of tying hands. Examples of sunk-costs could include military demonstrations, actual nuclear weapons development-related tests, or limited acts of force. The slow-signaling value consists of situations where North Korea signals that it is willing to slow down its nuclear weapons development or stop it completely. Observable examples include whether North Korea directly or indirectly, and either by own its own initiative or

in response to the actions of other actors, verifiably indicated to an international actor its willingness to engage in negotiations over its nuclear disarmament. Another signal would be North Korea taking verified steps to slow down or hurt its ability to pursue nuclear weapons as quickly as possible; a sunk-cost signal.

The values of the framework may overlap, such as a situation where North Korea simultaneously scores a fast speed value and a fast-signaling value by testing a nuclear weapon, but each measures a distinct element necessary for capturing the essence of North Korean nuclear variance. The speed value captures what North Korea is doing with in regards to its nuclear activity at that moment, whereas the signaling value is an indication of what North Korea wants to do in the future. Additionally, the speed value may be partially dictated by other actors; North Korea might score a fast speed value while actually desiring to go slow because it requires cooperation and agreement with another actor, the United States for instance, who in that instance refused participation. The components can co-vary together or create distinct groupings, which are captured in the four broad values of the dependent variable. Fast nuclear strategy consists of a fast value in speed and a fast-signaling value in signaling. Fast but slow-signaling strategy is comprised of fast speed and slow-signaling values. Slow but fast-signaling strategy is the opposite, with a value combination of slow speed and fast-signaling. Finally, slow nuclear strategy indicates slow speed and slow-signaling values. These dependent variable values across the two components of the framework are consistently observable during the years 2011-

2017. The four values for the dependent variable and the dichotomous framework provide a nuanced measure of North Korea's nuclear behavior that captures the varying transitions of strategy North Korea vacillates between.¹

In this study, the models will try to explain the transitions between the existing dependent variable values during the case study time period. A transition between the four strategy types does not have to be ordered; North Korea can switch between any strategy at any time. It is important to note that this approach assumes that when North Korea transitions to a different dependent variable value, or a different type of nuclear strategy, this strategy remains the default until the state observably transitions to a new type of strategy. Again, it is the goal of the models to explain North Korean nuclear strategy *variance*, and thus the study of the *transitions* in types of nuclear strategy is most important. Based on these dependent variable values, North Korea under Kim Jong-Un experienced four transitions from the years 2011 to 2017.

Transition 1: Right after Kim Jong-Un took power, on February 29th, 2012, North Korea transitioned from a fast but slow-signaling nuclear strategy to a completely slow nuclear strategy when it agreed to the Leap Day Agreement with the United States (Chanlett-Avery et al., 2016). Prior to the agreement, North Korea was engaged in a full-out pursuit of nuclear weapons capabilities characterized by various nuclear-related tests, illegal trading for nuclear resources, and an express desire to pursue nuclear weapons

¹ See Table 1 in the Appendix

publically (Hee-gwan Chin et al., 2013; Chanlett-Avery et al., 2016). Thus, a fast but slow-signaling strategy type captures that North Korea was still taking a defiant, bold posture with its nuclear activity but had expressed interest in negotiation in bilateral talks with the United States. After the agreement was announced by both the United States and North Korea, North Korea appeared to slow down its nuclear weapons pursuit from a full-out pursuit as promised. The agreement was to include the shut down of operations at the Yongbyon uranium enrichment plant, admittance of outside nuclear inspectors into North Korea, and a moratorium on nuclear-related tests in exchange for 240,000 metric tons of food aid from the United States (Chanlett-Avery et al., 2016).

Transition 2: Less than a month later that year on March 16th, 2012, North Korea announced that it planned to launch a satellite later that year to celebrate the former North Korean leader Kim Il-Sung's legacy (Chanlett-Avery et al., 2016). Other international actors, including the United States, deemed this to be a possible act of nuclear weapons development considering the shared technology of satellite launching and intercontinental ballistic missile technology, and thus a violation of the Leap Day agreement (Chanlett-Avery et al., 2016). Considering the audience costs created by threatening to launch a satellite while in an agreement with the United States, North Korea showed an observable transition to slow but fast-signaling nuclear strategy.

Transition 3: On April 13th, 2012, North Korea attempted to launch its weather satellite into orbit, using rocket engines that are also used for their ballistic missiles

(Chanlett-Avery et al., 2016). The United States subsequently cancelled its food aid plans to North Korea, signifying the end of the Leap Day Agreement. Although the launch was a failure, North Korea subsequently continued launching rocket-powered satellites and tested a nuclear device while reopening its Yonbyon facilities (Chanlett-Avery et al., 2016). This demonstrates an observable turn based on the dependent variable framework, from slow but fast-signaling to fast nuclear strategy. It is clear North Korea had resumed full out pursuit of its nuclear weapons development.

Transition 4: On January 10th, 2015, after a long period of using the fast nuclear strategy that included the testing of nuclear weapons and missile systems, various threatening statements, and a continued lack of participation in any treaty or agreement on nuclear weapons, North Korea contacted the United States to officially reopen bilateral negotiations (Davenport, 2017). This came at a time when North Korea recently announced plans to test a fourth nuclear device in the previous year. The offer entailed a moratorium on nuclear weapons testing in exchange for the cancelling of the annual US-ROK joint military exercise to decrease tensions. While the United States rejected the offer of negotiation, this demonstrated a transition from fast nuclear strategy to a fast but slow-signaling strategy based on the offering of a costly signal. The gesture of reopening negotiations for a denuclearization deal is a tying hands signal as North Korea trying to engage with the United States creates audience costs and hurts their ability to bargain in the future if they make an empty offer.

Transition 5: On February 7th, 2015, Kim Jong-Un personally oversees a missile test following the rejection of its offer to the US (Davenport, 2017). This can be characterized as a sunk-cost signal, where the official reporting of Kim Jong-Un's personal oversight of the tests displaying an arms build-up signal of North Korea's intentions to pursue an unequivocal fast nuclear strategy. Following the tests, North Korea continued fast speed nuclear development with various intercontinental ballistic missile tests and a nuclear detonation in January of 2016. Additionally, in May of 2016 North Korea held a congress for its only major political party, where Kim Jong-Un announced its intentions to have nuclear weapons and to refrain from using them unless North Korea's sovereignty was encroached upon (Davenport, 2017). This would indicate a full resumption of fast nuclear strategy by North Korea, ending its costly signaling of openness to slow down its nuclear weapons development.

Transition 6: On July 6th, 2016, North Korea releases an official statement indicating its willingness to resume denuclearization talks with the United States and South Korea (Davenport, 2017). The statement laid out significant conditions for any agreement, including the denuclearization of the entire Korean peninsula as well as the removal of security threats from the US. Considering the high level government spokesman used to communicate the message, which is generally only used when North Korea hopes to convey a sincere message, the statement could be considered a costly, credible signal of tying hands (Carlin, 2016). Thus, North Korea transitioned from a fast

nuclear strategy to a fast but slow-signaling nuclear strategy once again by not necessarily slowing down its nuclear weapons development at the time but expressing a desire to slow it down pending the interest of other actors. The United States responds by designating Kim Jong-Un and other top North Korean officials as having committed human rights abuses, an indirect rejection of the offer (Carlin, 2016). Also, South Korea and the United States announce their decision to deploy a missile defense system in South Korea intended to stop intercontinental ballistic missiles (Davenport, 2017).

Transition 7: The final transition occurs on August 3rd, 2016, when North Korea fires a medium range ballistic missile off the coast of Japan in a clear sunk cost signaling of its willingness to resume its fast nuclear strategy in full (Davenport, 2017). Shortly after North Korea detonates its fifth nuclear test, and continues ICBM missile testing as well as public demonstrations of missile capabilities. In 2017, North Korea announce plans to test more missiles in response to remarks by United States President Trump, conduct their sixth nuclear test, and Kim Jong-Un personally advocates for its pursuit of nuclear weapons in light of threats President Trump makes towards North Korea regarding their nuclear activity (Davenport, 2017).

These seven transitions capture the strategic variance North Korea has displayed under the Kim Jong-Un era, and provide a systematic and empirically evidenced dependent

variable that can be used to test and compare the hypotheses of theoretical models.² We can see that depending on the cooperation of the United States, and to a lesser extent South Korea, that North Korea varies its strategy from a completely fast pursuit of nuclear weapons to an occasional willingness to slow down its nuclear activity. In the following sections, the coercive bargaining model, the political legitimacy-prospect model, and the crisis bargaining “madman” model will be applied to explain these observable transitions.

Coercive Bargaining Model:

As briefly discussed in the literature review, Victor Cha’s coercive bargaining model was applied to explain North Korea’s nuclear behavior during the rule of Kim Jong-II, and was intended to contrast the previously inductive explanations of the state’s behavior with a more systematic approach (2003). Cha attempted to explain North Korea’s “two faces”, or variance in its nuclear policies, by using both power transition theory, based on neorealist assumptions, and prospect theory. Cha put forth a “coercive bargaining” explanation that argued North Korea, due to its relatively weaker position in the international system compared to other regional actors, could choose a rational strategy of provoking other states like the US and South Korea with its nuclear weapons. The rationale for these provocations would be that small-scale provocations would lead the

² See Table 2 in the Appendix

other states to compromise and lessen the gap in relative power between them. To support these arguments, Cha uses neorealist principles to serve as the basis for North Korea's nuclear strategy while using prospect theory to reflect the risky decision-making involved in going nuclear. Below, the key theoretical assumptions of the model, independent variables these assumptions point to, and the logic of the model in practice will be explained before presenting hypothesis that will be applied to our case study.

Theoretical Background and Assumptions

The coercive bargaining model is based on the theoretical assumptions of power transition theory and prospect theory. Power transition theory, as described in the literature review, assumes that the power distribution of the international system is the main causal driver for state behavior. Thus, the model is an explanation of North Korea's nuclear strategy by focusing on how the power distribution as an independent variable affects the strategy that North Korea takes. Power transition theory also categorizes states within the international system as status quo or revisionist, based on whether the international system benefits them or not (Tammen et al., 2000). In the original model, and consistent with most approaches in IR theory, power distributions are determined by comparing economic size, defense spending, and military capabilities (Cha and Kang, 2003). If a state is in an unfavorable position in the international system, as a revisionist state it will try to change the power balance in the system to its own advantage through military build up of its

capabilities or through alliances (Waltz, 1987; DiCicco and Levy, 1999). A state may choose its strategy of closing the power gap through various ways, but offense and defensive balance, or whether military capabilities favors the attacker or defender, plays a significant role in this decision-making according to the model (Van Evera, 1984; Cha and Kang, 2003). This model also replaces the standard expected-utility model, or rational actor assumption, with the alternative prospect theory to explain a state's decision-making process and how a state can make rational decisions in a seemingly non-rational way (Chan and Kang, 2003). This is based on the idea that international politics and security issues in particular present inherently high-risk, highly uncertain situations that prospect theory specializes in. To understand the dynamics of the model, prospect theory's core assumptions and how it differs from the standard rational actor assumptions will be detailed below.

Prospect theory, as developed by Daniel Kahneman and Amos Tversky (1979), assumes that an actor's decision-making and preferences revolve around a reference point that the actor use to evaluates assets and choices. This is in direct contrast to the expected-utility rational actor model where the costs and benefits are weighed completely objectively. Based on this idea that the reference point is central, prospect theory offers a general process of choice making. First, in an editing phase an actor identifies a reference point through a process called framing, using available information like expectations, aspirations, social norms, and social comparisons. The actor also assigns value and

probability of outcomes to available options around this reference point. Second, in the evaluation phase where essentially the values of possible outcomes are combined with their weighted or compared probabilities, and the actor uses the product to make decisions. In other words, actors do not have a consistent and transitive set of preferences and their behavior is not always conducive to maximizing these preferences. Consequently, behavior becomes non-linear or irrational. Instead of making choices that attempt to maximize total welfare or wealth for instance, actors make choices according to probable gains and losses of actions that revolve around a given reference point.

Another key assumption in prospect theory is that actors care more about changes to assets than net asset levels. Essentially, actors are more concerned with gains and losses that are given meaning by a reference point than the total or net worth of an option. This claim invites a critical subset of assumptions about gains and losses. First, actors treat gains and losses differently. They tend to overvalue losses compared to relatively equal gains. This affects actors' behavior in what is called loss aversion. For instance, an actor will probably not engage in an activity that has an equal chance of success or failure since the actor tends to overweight the failure. Typically, 50-50 gambles are not taken. Regarding gains and losses, the endowment effect refers to the notion that actors tend to overvalue assets they already possess as opposed to comparable assets they do not have. The endowment effect implies that losing an asset is worse than acquiring an asset of comparable value. This means that an actor might work or pay more to keep an asset than

they ever would to attain that asset if it did not belong to them, again pointing to what seems like irrational behavior. A third sub-assumption called status quo bias merely implies that actors tend to treat the costs of moving away from the status quo as losses and the benefits of moving away as gains. This is based on the notion that actors overvalue losses so they tend to remain at the status quo. The last sub-assumption in regards to gains and losses is called risk orientation, which simply states that actors are risk-averse in the domain of gains and risk-acceptant when it comes to losses, with a reflection effect around the reference point. This explains a tendency of actors' choices; if the possible outcome is a gain then the actor will probably not take much risk to achieve it, but if there is a possible loss then the actor will be far more inclined to take risks in order to avoid it.

Finally, prospect theory assumes that framing is absolutely crucial to choice making. Framing is the identification of a reference point, and given the role of the reference point in shaping an actor's valuations of assets and choices the actual process of establishing the reference point is extremely important. Most instances of framing the reference point depend on the setting or situation, and usually the status quo, or the pre-determined situation, is made to be the reference point. A change in frame is just as important as establishing the original frame, which can easily result in reference point, preference, and value changes. This is known as preference reversal. For instance, the difference between using the words "intervention" and "invasion" can mean all the difference in shaping an actor's preferences, values, and choices. It is also vital to take into

account dynamic situations where the reference point and frame might have a high likelihood of changing, and an actor's behavior might be altered substantially with each successive choice. In a sequence of choice making, the instant endowment effect may occur, which is described as an actor instantly accommodating to previous gains and treating these acquired gains as their own assets. The important point to recall is that actors value losing an asset much more than gaining an asset of comparable value.

Based on these assumptions and effects, the coercive bargaining model can explain the manner in which a revisionist state attempts to change the power balance status quo in its favor based on the risks associated with a revisionist strategy and how the strategy is framed.

The Model

The main logic at work for the coercive bargaining model is that North Korea finds itself in a losing situation due to its weak position in the international system's power distribution. Its main rivals, including the United States and South Korea, have been pulling farther and farther away from North Korea in terms of economic size, defense spending, and military capabilities, which are the most common indicators of power in international relations theory (Cha and Kang, 2003). Additionally, North Korea is technically still at war with South Korea and the United States, and the US has shown an express desire for a regime change and the denuclearization of North Korea (Cha and

Kang, 2003). Thus, North Korea finds itself in an untenable situation where the growing power gaps between itself and its rivals indicates that it is constantly losing its security and the future will eventually bring an unacceptable outcome. According to the model, this is the exact scenario necessary for preventive or preemptive action by a revisionist state to change the status quo. Preemptive action, as described by Cha (2003), is “when a state perceives aggression by another as imminent and acts first to forestall the impending attack,” while preventive action is “when a state is motivated to attack first, or otherwise suffer increasing inferiority in capabilities vis-à-vis the opponent over time.” In the North Korean case, this preventive or preemptive action takes the form of its aggressive nuclear strategy to develop nuclear weapons despite the high risk involved in doing so. Aggressively pursuing nuclear weapons can have the gains of attaining nuclear deterrence capabilities or forcing the United States and South Korea to compromise to a level that it was unwilling to before, thereby changing the status quo in North Korea’s favor. Thus, it becomes rational in certain situations to pursue nuclear weapons. The main mechanism responsible for North Korea’s decision to employ an aggressive, coercive bargaining strategy as Cha calls it, or in this study classified as fast nuclear strategy, versus a less risky nuclear strategy such as not seeking nuclear weapons at all, pursuing them in secret, or trying to change the status quo through other means as captured by slow nuclear strategy, is their perception of their power position in the international system. The model uses prospect theory to split the action choices of using preventive/preemptive action or

not into a domain of gains, neutrality, or losses based on the reference point, or status quo, of their power position. Remembering the concept of risk orientation, we know that actors treat potential gains and losses differently. In a domain of gains, actors are risk-averse, meaning that the calculation of probable losses is considered more strongly than the calculation of probable gains. In a domain of losses, actors are risk-acceptant, meaning they are more willing to accept a risky choice to avoid a loss to their current status quo. Applied to North Korea, we can equate the preemptive/preventive action choice with a fast nuclear strategy considering the ability of nuclear weapons to alter the status quo in a losing situation. We also assume that fast nuclear strategy is highly risky because of the probability of significant losses of nuclear testing and other provocations inherent in fast nuclear strategy, such as military intervention or sanctions from the United States, South Korea, and the international community. Based on North Korea's reference point, or status quo within the international system, the model indicates that North Korea will choose its nuclear strategy based on the riskiness of the nuclear strategy and how favorable its position is in the international system.³ In terms of empirically determining the framing of the nuclear strategies with domains of gains, neutrality, or loss, Victor Cha gives the following four criteria that can be qualitatively determined: (1) fulfillment of ideational, national objectives, (2) economic and military well-being, (3) status in the international system, (4) availability of allies (Cha and Kang, 2003).

³ See Table 3 in the Appendix

Hypothesis Set for the Coercive Bargaining Model

The coercive bargaining model points us to the power distribution of the international system, and whether North Korea's position is getting better or worse, as the main independent variables for its choice of nuclear strategies. Each of the four strategies represents an ordinal level of risk that corresponds to how much risk North Korea is willing to make in regards to accomplishing its nuclear goals, dependent on the domain the strategies are framed in. We can thus put forward the following hypotheses:

H1: North Korea will use a fast nuclear strategy when it is framed in a domain of losses and offense has the advantage. This is when North Korea is in a losing situation in regards to its power position, meaning its rival states are increasing the power gap to an untenable to degree. Therefore, North Korea accepts the most risk with the highest reward by pursuing nuclear weapons capabilities as quickly as possible to revise the power imbalance. North Korea must believe that despite the high risk of failure, the nuclear weapons capabilities and subsequent deterrence they seek are achievable.

H2: North Korea will employ a fast but slow-signaling strategy when it is framed in a domain of losses but defense has the advantage. Similar to H1, North Korea's increasing power gap makes the current situation untenable and thus they should pursue nuclear weapons as fast as possible even with the likelihood that they will fail. Due to the increased likelihood of failure via US intervention, internal failure, or other risks that

indicates an advantage for North Korea's opponents, North Korea will try "coercive bargaining" where it uses its nuclear weapons development as a bargaining tool to seek compromises from the US and other rivals.

H3: North Korea will employ a fast but slow-signaling strategy when it is framed in a neutral domain but offense has the advantage. This means that North Korea's position is remaining relatively constant or a non-losing level for North Korea, but pursuing nuclear weapons presents an advantage that North Korea can leverage for increased gains, North Korea will attempt to coercively bargain for a decrease in the power gap between itself and its rivals.

H4: North Korea will use a slow but fast-signaling strategy when it is framed in a neutral domain but defense has the advantage. Like H3, North Korea deems itself in a non-losing situation in regards to its power position but the calculation of risk and payoffs from going nuclear are not favorable. Thus, North Korea will hedge and pursue a slow nuclear strategy to avoid the high risks of fast nuclear weapons pursuit, but signal that they may change to an active pursuit in order to increase possible gains from compromise.

H5: When North Korea is in a domain of gains, or when it perceives its position in the international system to be improving relative to the status quo, and offense has the advantage it will employ either a slow but fast-signaling or a slow nuclear strategy. Because its position is improving, and as described with the endowment effect, North Korea will be risk-averse and try to avoid losses to the state's newly increased position and

therefore will take the low-risk slow strategy. However, if the gains from the prospects of going nuclear are high enough, it may signal that it is willing to pursue nuclear weapons actively to further increase its bargaining gains.

H6: When North Korea is framed in a domain of gains and defense has the advantage, the state will use a slow nuclear strategy. This indicates that like in H5 the power distribution is changing favorably for North Korea, and that it will protect these increases in power balance by not taking risky behavior. Since defense has the advantage, the increased risk of fast nuclear strategy eliminates the possibility that they will signal for a pursuit of nuclear weapons.

These six hypotheses correspond to the different scenarios under which the model would expect North Korea to employ one of the four types of nuclear strategy observed in the case. Thus, we should be able to observe changes in North Korea's power distribution status in the international system and the corresponding affects this has on its perceptions of nuclear strategy risks to explain the seven transitions that have occurred during the Kim Jong-Un regime. If the independent variable of changing power balance co-varies in line with the changes in the dependent variable as expected, we can reasonably assume the model adequately explains the variance by the preliminary congruence test standards used for this study.

Application to the Case

Transition 1: Fast but Slow-Signaling to Slow, February 29th, 2012

Prior to February 29th, 2012, North Korea displayed a fast but slow-signaling nuclear strategy as it negotiated bilaterally with the United States while it simultaneously introduced and consolidated Kim Jong-Un's rise to leadership. North Korea had in 2011 expressed an interest in returning to the negotiating table about denuclearization after developing its uranium enrichment facilities and increased tensions with South Korea following shelling of the South Korean Yeonpyeong island in 2010 (Davenport, 2017). This coincided with the death of Kim Jong-Il, the former leader of North Korea, and the official announcement of Kim Jong-Un as the new supreme leader of the country in late 2011 (Davenport, 2017). Shortly after, North Korea and the United States announced their Leap Day Agreement, marking the transition into a period of slow North Korean nuclear strategy. For a brief period, North Korea seemed to agree to slow its nuclear weapons development with its moratorium and adhere to the deal faithfully. This change in nuclear strategy should thus indicate a change in how North Korea interpreted its position in the international system and an offense versus defense balance of preemptive/preventive action, or pursuing nuclear weapons. We should expect to see evidence indicating a change in North Korea's power position to switch its framing of nuclear strategy choices from a domain of losses to a domain of gains with defense having an advantage, as predicted in H3 and H6.

Based on a variety of state and scholarly resources, it is fairly easy to establish that North Korea is operating in a domain of losses before the Leap Day Agreement based on its position in the international system. From an ideational standpoint, Kim Jong-Un's nascent rise to power was consolidated on the prestige of his father and grandfather's legacy (Hee-gwan Chin et al., 2013). In other words, Kim Jong-Un had to prove he could lead the country in the way his predecessors had. Considering its economic and military wellbeing and alliance strength compared to the power of a combined United States, South Korea, and Japan, North Korea was in a far inferior position in terms of capabilities and alliance (Goo and Lee, 2014). The four criteria for determining a decisional frame all indicate that North Korea, as it had been for around two decades prior, was in a losing situation compared to its rivals. In terms of offense and defensive advantage of a preemptive/prevent action like developing and using a nuclear weapon, the United States, South Korea and other major regional rivals had a nuclear deterrence and military strength advantage despite North Korea's advances in nuclear technology (Chanlett-Avery et al., 2016). This would indicate that in accordance with H3, North Korea was pursuing a fast but slow-signaling nuclear strategy due to its risk-acceptant behaviors and attempt to coercively bargain a deal with its rivals to change the status quo. Thus, a change to slow nuclear strategy would indicate a change to a domain of gains, where North Korea believes its status quo position in the system is improving and thus would not take a risky nuclear strategy like a fast one to jeopardize its gains as explicated in H6. The Leap Day

Agreement must then logically have such a status quo changing payoff for North Korea through its coercive bargaining that it is willing to slow down or stop its nuclear development.

This does not seem to be the case; however, considering that the payoffs given to North Korea from the United States were 240,000 metric tons of food aid and the implicit increase in serious negotiations with higher stakes in exchange for the freezing of its nuclear facilities and moratorium on nuclear-related testing (Chanlett-Avery et al., 2016). While the risks to the nuclear weapons program was not high, the payoff of the Leap Day Agreement do not significantly change the power imbalance in the international system to North Korea's benefit. The power gap in military, economy, and alliance capabilities had not been significantly changed enough to warrant a slow nuclear strategy as expected in H6. In other words, the model would expect North Korea to coerce much higher, much more significant payoffs from its coercive bargaining efforts with its nuclear strategy.

Transition 2: Slow to Slow but Fast-Signaling, March 16th, 2012

The threat by North Korea to launch a weather satellite in honor of Kim Il-Sung marks the transition from slow to slow but fast-signaling nuclear strategy. So why would North Korea change course and threaten the potential gains from its new Leap Day Agreement? The coercive bargaining model would point to a change to a domain of gains with offensive advantage (H5) or a neutral domain with a defensive advantage (H4) in

regards to its reference point of its own power status in the system. So is there evidence to support either of these hypotheses?

For the same reasons listed for the first transition, there is little evidence to suggest the payoffs of the Leap Day Agreement is enough to change the structural power disparities between North Korea and its rival states significantly to the point that it perceives its position in a domain of gains. The food aid had not been delivered at that point, and further negotiations had not taken place to increase the payoffs of further agreements (Davenport, 2017). Likewise, there is little evidence to suggest that North Korea saw itself in a neutral domain, or a non-losing situation, considering nothing had structurally changed about its prior losing power position because of its Leap Day Agreement payoffs. It is conceivable that North Korea was bluffing, or attempting to coerce a stronger payoff from the United States by threatening a satellite launch, but the audience costs of announcing a planned launch and the risk involved in losing the Leap Day Agreement do not support this logic. Therefore, the model's hypotheses corresponding to this type of nuclear strategy are not empirically supported to explain this transition.

Transition 3: Slow but Fast-Signaling to Fast, April 12th, 2012

Less than a month after threatening to launch their satellite, North Korea actually did so in a failed attempt to use ballistic rocket technology to push a weather satellite into Earth's orbit. This started a clear transition to fast nuclear strategy, and was followed by

uninterrupted pursuit of nuclear weapons development through various tests and government policies supporting its nuclear program. According to the model, a fast nuclear strategy indicates that North Korea sees itself in a domain of losses with an offensive advantage (H1). We are therefore looking for evidence to suggest that North Korea must have framed its nuclear strategy choices in a domain of losses based on its losing power situation, and thought a high risk solution like a nuclear weapons program provided its best chance of changing the status quo if successful.

As in the cases of the previous two transitions, there is an abundance of evidence to suggest that along economic, military, and alliance indicators, North Korea was in a drastically inferior power position compared to its rival states and would thus frame their choices within a domain of losses. Considering that North Korea subsequently launched a satellite into orbit successfully following the failed launch along with other progressively impressive missile tests, tested another nuclear device in 2013, and the available intelligence on North Korean nuclear technology at the time, there is empirical support for the idea that North Korea saw its ability to alter the status quo through the attainment of functional nuclear capabilities (Davenport, 2017; Kim and Cohen, 2017). In this case, the transition to a fast nuclear strategy seems supported by the rational choice to take a high-risk strategy in an untenable, losing circumstance.

Transition 4: Fast to Fast but Slow-Signaling, January 10th, 2015

After three years of nuclear behavior consistent with a fast nuclear strategy, such as nuclear device tests and missile testing designed to increase North Korea's missile range and ability to carry nuclear weapon payloads, North Korea conveys a rare public offer to suspend its planned fourth nuclear test in exchange for the cancellation of the annual US-ROK joint military exercises that have bothered North Korea in the past (Davenport, 2017). This public announcement through its official news agency, which claims North Korea offered the United States, was legitimate enough to illicit a rejection by the United States' State Department spokeswoman Jen Psaki, claiming that the nuclear issue and the military drills were separate issues (Davenport, 2017). Nonetheless, this indicates a gesture or offer with real audience costs that put North Korea's prestige on the line, since they would be compelled to agree or suffer reputation and possibly conflict costs if the United States had agreed and North Korea rescinded its offer. Thus, North Korea suddenly changed from fast to fast but slow-signaling strategy, dependent on negotiations with the United States. According to the model, we should expect North Korea to change to this type of nuclear strategy when it sees its strategic choices in a domain of losses with a defensive advantage (H2) or in a neutral domain with offensive advantage (H3).

There is yet again little evidence to support North Korea seeing its position in a non-losing or neutral position considering the consistently large gaps in economic, military, and alliance capabilities between North Korea and its rival states. Instead, as has been the case for the entirety of the Kim Jong-Un era up to this point, North Korea finds itself in the

same losing situation considering the imbalance of power in Northeast Asia (Goo and Lee, 2014; Pollack, 2017). Therefore, H3 is not supported, but there is evidence to suggest that the offense-defense balance had shifted in favor of defense (Pollack, 2017). From its adversaries, South Korea launched its own missile tests in 2014 in a defensive response to North Korea, while the United States expanded its sanctions on nuclear-related trade entities and individuals (Davenport, 2017). Additionally, North Korea's strongest ally China declared a red line on North Korea's nuclear weapons program and states its commitment to denuclearization of the Korean peninsula, presenting another obstacle to offensive advantage. Thus, the second hypotheses aligns with the change in strategy by North Korea, explaining this change through North Korea's interpretation of risks attached to a fast nuclear strategy.

Transition 5: Fast but Slow-Signaling to Fast, February 7th 2015

The return to a full, fast nuclear strategy was highlighted by the costly signal of launching an anti-ship missile with the leader Kim Jong-Un reported as present for the test (Davenport, 2017). Kim Jong-Un personally being present for an aggressive military test, open to the world, indicates North Korea's foreign policy plan to continue its pursuit of nuclear weapons with no intentions of backing or slowing down based on the costs this would do to the national and personal image of North Korea and Kim Jong-Un respectively. Fast nuclear strategy indicates an offensive advantage in a domain of losses

(H1), and there should be evidence to support the idea that North Korea believed a preventive/preemptive action was viable through its nuclear weapons development.

There is evidence to support the hypothesis, based on the subsequent tests North Korea conducted within the year and based on the intelligence and scholarly estimates of its nuclear technology. North Korea conducted several missile tests that would further its nuclear weapons capabilities significantly such as a successful submarine-based ballistic missile launch, several intermediate-ranged ballistic missile tests, and another nuclear detonation that the country claims is a hydrogen bomb (Davenport, 2017). Additionally, intelligence and scholars had reason to believe that North Korea's intercontinental ballistic missiles were nearly operational (Chanlett-Avery et al., 2016; Kim and Cohen, 2017). This is consistent with the model's first hypothesis expectations.

Transition 6: Fast to Fast but Slow-Signaling, July 6th, 2016

On July 6th, 2016, a high-level North Korean government spokesman released a statement declaring North Korea's openness to resume denuclearization talks provided conditions were met on the agreed definition of denuclearization and the ceasing of threatening behavior between countries (Carlin, 2016). This message constitutes a costly signal wherein North Korea would suffer audience costs if it backed down from its willingness to cooperate, yet the United States responded by adding Kim Jong-Un and other North Korean leaders to its designations of human rights abuses (Carlin, 2016). Despite the negotiating failure, this signal represents a change in nuclear strategy from fast

to fast but slow-signaling, which is expected when North Korea is located in a domain of losses with defense having the advantage (H2) or a neutral domain with offensive having the advantage (H3).

The evidence almost exactly mirrors that of the fourth transition, where there is little evidence to support along economic, military, or alliance capability indicators that North Korea had improved its power position enough to consider itself in a non-losing situation in the international system. However, while clearly located in a domain of losses, developments in South Korea and the United States again shifted the advantage to defense. Increased sanctions by the United States and the United Nations Security Council on North Korean nuclear-related trade, and the newly planned implementation of the Terminal High-Altitude Area Defense battery (THAAD) in South Korea further pushed the offensive-defensive balance in the defensive direction (Davenport, 2017). This evidence seems to support the hypothesis that North Korea would hedge their nuclear weapons strategy to open up coercive bargaining options in a losing situation as the viability of nuclear-based preemptive/preventive action decreased.

Transition 7: Fast but Slow-Signaling to Fast, August 3rd, 2016

North Korea returned to completely fast nuclear strategy through the costly signaling of arms build up, or a message through missile testing in August of 2016. This indicated North Korea was once again unwilling to back down on its nuclear ambitions and was followed by a slew of submarine-based missile tests, ICBM tests, and a fifth

nuclear test (Davenport, 2017). The return to fast nuclear strategy is explained in the model's first hypothesis, where North Korea's strategic choices are framed in a domain of losses and offensive has the advantage.

The successful testing of SLBM's, ICBM's, and a nuclear device indicate North Korea is closer than ever to second-strike capability and viable preemptive/preventive action through nuclear means (Armstrong, 2017; Kim and Cohen, 2017; Pollack, 2017). While the THAAD system was implemented in South Korea, North Korea has estimated strike range of up to Los Angeles, Denver, and Chicago (Davenport, 2017; Kim and Cohen, 2017). The fast nuclear strategy reflects that North Korea, while in a losing position, is rationally justified in its pursuit of viable nuclear capabilities that could drastically alter the power status quo through military build up. The hypothesis provides a reasonable explanation for North Korea's transition to fast nuclear strategy.

Analysis

In sum, the coercive bargaining model provides a compelling, structural explanation for North Korea's strategic nuclear variance. Adjusted for a new time period, the model provides a logical story for why North Korea transitions between seeming as if they want to cooperate on denuclearization or wanting to go full nuclear. North Korea's drastically inferior power capabilities and position in the international system force them to take high-risk strategies, like fast or fast but slow-signaling strategies, to change the

status quo in their favor due to the likeliness of the future being worse if they do not act. Based on the offensive-defensive balance, North Korea will likely pursue nuclear weapons at full speed when it thinks the outcome is viable enough, but hedges to a coercive bargaining strategy when defense gains an advantage. The coercive bargaining buys them more time through a possibly status quo changing agreement while still allowing them to pursue nuclear weapons capabilities without slowing down. The model is weak; however, in explaining why North Korea agreed to such low payoffs in the Leap Day Agreement and upon agreeing to it, explaining what changed in the short time after the agreement to compel them to switch back to a fast nuclear strategy. This indicates that while structural, power-based explanations may play an important role in determining North Korea's nuclear strategy, there are likely other factors at play and thus it cannot provide a full picture of North Korea's rationale. In the very least, the ways states interpret and respond to the main causal driver of anarchy and its complementary effect on power imbalances in the international system may be more diverse than a focus on capabilities and offensive-defensive balance alone would indicate.

The “Madman” Model:

Another way of looking at North Korean nuclear strategy is through crisis bargaining theory, where we can produce explanations by examining the strategic

interactions between states on a conflicting issue. North Korea, the United States, and South Korea disagree on whether North Korea should be allowed to have nuclear weapons enough to threaten the use of force or use it in small doses to the point that war is possible. In this situation, the two sides alone cannot resolve the dispute without giving up and both are better off by coming to an agreement to split the higher payoffs of cooperation rather than fighting a war over it, which would be the most costly resolution to both sides. The dynamics of any agreement are complicated by the fact that no one is forcing them to make an agreement; sovereign states are not obliged to cooperate unless they agree to on their own terms in the anarchy of international politics. In other words, they are playing a non-cooperative bargaining game to see who gets the best piece of the pie they would share through an agreement that settles the conflict. This is a different way of looking at nuclear strategy from the coercive bargaining model for instance because it recognizes the impact of strategy by other important players beyond North Korea, and stresses the importance of state's choices more than structural approaches will allow. Game theory is able to map out the entire catalogue of possible strategic interactions and outcomes of a bargaining game, and thus the ideal or probable strategies for each player, from start to end by using mathematical models to design the rules of the game and the characteristics of the players. Traditionally, one of the main characteristics of states in game theory models and IR theories in general is assuming they are completely strategically rational, or assuming they will always pick the best option available to them to attain their preferred outcome.

Additionally, all of the other states know that every other state shares this strategic rationality. This characteristic allows for crisis bargaining theorists to explore likely strategies and outcomes for states in conflict over an issue, and the conditions for war and peace, as closely similar to real life as possible, and thus we gain valuable explanations of state interactions. The “madman” model, developed by Avidit Acharya and Edoardo Grillo (2015), relaxes this rational assumption slightly but seeing how strategies and outcomes are affected if one state inserts the possibility that they are “crazy”, or strategically irrational and might choose a strategy that lessens the probability of its own and perhaps its opponents preferred outcomes. This can drastically affect the behavior and strategy of the other bargaining state, and in some cases can end in the “crazy” state getting a larger share of the agreement pie than it would have if it had acted normal. Put more succinctly, the “madman” model formally explains that it may in fact be rational to act “crazy”, or irrational, in certain situations (Acharya and Grillo, 2015). Considering that North Korea has often been dubbed irrational or crazy on many occasions, there is an obvious opportunity to compare whether their strategy actually aligns with the model’s expectations, thereby supporting the possibility North Korea is using its crazy reputation as a strategy to reach its nuclear goals. In order to capitalize on this opportunity, the theoretical and historical background the “madman” model is based on must be fully understood, along with the actual logic and approach of the model. Based on the model’s assumptions and ideal equilibriums, we can generate hypothesis on North Korea’s nuclear

strategy under Kim Jong-Un by studying the strategic interaction between North Korea and the United States or South Korea.

Theoretical Background and Assumptions

Before introducing the formal scholarship on crisis bargaining that helped form the “madman” theory, it is important to note that anecdotally the concept of a “madman” strategy has been discussed and utilized before by important historical figures. Acharya and Grillo (2015) note that former United States president Richard Nixon used the strategy during the Vietnam War in an attempt to coax better terms of an agreement from the North Vietnamese or Soviet Union. Additionally, they cite Iran’s former president Mahmoud Ahmedinejad as being suspected of trying to appear crazy for strategic gain along with other notable historical figures, including North Korean leaders (Acharya and Grillo, 2015). There is a clear historical precedent for the idea that country leaders in various time periods and scenarios thought it would behoove their chances of attaining what they wanted by acting crazy. Several scholarly works recognizing the “madman” strategy’s role in Nixon’s policies buttress the concept’s existence (Kaplan, 1991; Kimball, 2004; Lake, 2011).

Theoretically, the model is based on formal crisis bargaining literature of states in non-cooperative games. The models in the crisis bargaining research paradigm generally focus on the conditions for war and peace, or when conflict is likely to arise in a situation

when states are involved in a dispute of interests. As discussed previously, Acharya and Grill (2015) draw from Powell (1987), Banks (1990), and Fearon (1995) in assuming that states will not go to war “when both parties are able to locate a Pareto superior negotiated settlement”, or in other words when the benefits of an agreement do not make any state better off at the expense of making another participant worse off. Thus, the most efficient outcome of crisis bargaining situations is found in a settlement, and war is costly and inefficient (Fearon, 1995). Complications emerge when states have a reason to misrepresent information about themselves. In Robert Powell’s seminal work “Crisis Bargaining, Escalation and MAD” (1987), Powell demonstrates how game theoretic models on state conflict could be improved by introducing games of two-sided incomplete information, meaning the bargaining states do not know everything about the preferences and characteristics of their counterpart states. By creating asymmetries or gaps in information, new equilibriums, or outcomes based on the ideal strategy of each player, were found. This occurred even when assuming states were strategically rational, and subsequent works like Fearon (1995) explain conditions in which even strategically rational states will go to war despite knowing the costs and inefficiencies of war. Schultz (1999) demonstrated in an American politics context how politicians attempt to misrepresent their own characteristics as either “tough” or “lenient” as strategies to gain better payoffs from bargaining. This spark in research focusing on how misrepresenting behavioral characteristics about a bargaining actor affects equilibrium, or in other words

how uncertainty in player types in a game, is most relevantly represented in Abreu and Gul's work on how players form reputations with various effects when bargaining (2000).

The Model:

The main ideal of the model is that even a strategically rational state has incentives to act crazy because other strategically rational states may give in to a crazy state's outlandish bargain offers instead of risking escalation or conflict with a possibly crazed state (Acharya and Grillo, 2015). Additionally, even rational states will act crazy sometimes even if they are not if other states believe there is at least a chance of them being crazy. Thus, states pretend to be crazy at times depending on how crazy they are perceived to be and the payoffs they can expect from it. The determinants of how crazy a country acts are how much it is believed to be crazy prior to engagement by the other states, and two mechanisms that come in to play when two countries are guessing about the strategic rationality of the other state: reputation motive and defense motive (Acharya and Grillo, 2015). "The reputation motive describes the incentive of a country to pretend to be crazy in order to get a better settlement," while the defensive motive is "the incentive of a country to make larger demands that risk escalation, to deter its adversary from pretending to be crazy too often" (Acharya and Grillo, 2015). The model uses game theory to design a game of interactions between two states in conflict over an issue, and based on the formulas and assumptions that make the rule of the game, shows unique predictions about when states are likely to act rational or not and the costs and benefits of using the

“madman” strategy in different scenarios. For simplicity’s sake, the assumptions and comparative statistics of the model will be outlined below with accompanying diagrams of the equations, algorithms, and mathematical propositions listed in the appendix.

Game theoretic models use game trees, or a map of a bargaining game that examines all possible moves by the players, to find the optimal strategies of each player and equilibrium outcomes. In the “madman model”, the following game tree is used as a base for the rules and parameters of the game.⁴ Acharya and Grillo (2015) describe the following scenario: “Countries A and B are engaged in a dispute over an issue. Country A moves first – the choice is between attacking country B and resolving the dispute peacefully. If country A attacks, then country B chooses between surrender and retaliate. If country B retaliates, then country A can either end the war with an armistice, or it can escalate the conflict by choosing total war.” The payoffs at each progressive node that does not end in peace decrease in value for each country, with an even steeper decrease in payoffs for each country if the outcome is total war, meaning that the more the dispute escalates the more costly it is to the countries and the less ideal for the attainment of their preferences. Acharya and Grillo (2015) outline how country A, the attacking country, prefers outcomes in the following order based on payoffs: surrender, peace, armistice, total war. Country B prefers: peace, armistice, surrender, total war (Acharya and Grillo, 2015). Acharya and Grillo (2015) demonstrate that under such conditions, the pareto outcome for

⁴ See Acharya and Grillo’s figure in Figure 1 of the Appendix

each country, or “subgame perfect equilibrium” is peace, but under certain conditions states will not arrive at the most efficient outcome for both sides.

Acharya and Grillo (2015) use an example where country B’s belief that country A is strategically rational is 99%, which conversely denotes country B thinks there is 1% chance that country A could be crazy, or will always choose to attack instead of choosing peace and if country B retaliates they will always choose total war instead of armistice. Country A also thinks with 100% certainty that country B is strategically rational, and has a very slight advantage in military power, which concerns the payoffs of an armistice outcome. Both countries know the other country’s prior beliefs about their strategic rationality, and both are the only one who really know if they are strategically rational or not respectively. Assuming that both countries are actually strategically rational, country A will adjust its probability of choosing peace or attack and armistice or total war depending on the probability that country B thinks it is crazy and how strong its military is compared to country B. Similarly, country B will adjust whether it chooses retaliate or surrender based on attaining its highest preference outcome, the size of its military disadvantage, and the probability with which it thinks country A might be crazy. Acharya and Grillo (2015) calculate that in this example country A will thus attack just a bit more than 10% of the time, with country B retaliating with a probability of barely over 90%. This means war is the outcome 10% of the time in equilibrium, and demonstrates how even strategically rational states can reach a total war without either wanting to and without it being the

efficient, rational option based on the possibility that one country thinks the other is crazy. The outcomes can change based on whether country B acts strategically to choose between retaliate or surrender based on whether it thinks country A is sequentially rational, or in other words whether it think country A will choose armistice or not after attacking. This forms the basis for the logic of the model, where a country's strategic choices change based on the prior probability of its opponent thinking it is crazy; the more likely country B thinks country A is crazy, the more likely country A is to attack (Acharya and Grillo, 2015). Additionally, the stronger country A is militarily compared to country B, the less likely it is to attack and the less likely country B is to retaliate, which lowers the chances of war (Acharya and Grillo, 2015).⁵ The “madman” model is then a full mapping of the equilibrium probabilities for the various outcomes in the game tree depending on the varying characteristics of the states, such as when the probability of a crazy state is higher, both states are believed to possibly be crazy, or there are different military imbalances.

In the actual “madman” model, the game tree is modified to include a round of bargaining where country B has the ability to make an offer to country A after country A attacks. This is significant because it gives both countries two nodes to find a resolution to the conflict that avoids the costliness of total war.⁶ As Acharya and Grillo (2015) explain:

⁵ See Acharya and Grillo's paper for more in-depth explanation; see Equation 1 in Appendix

⁶ See Acharya and Grillo's figure in Figure 2 of the Appendix

“Countries A and B are engaged in a dispute. Country A begins by deciding between peace and attack. In the case of peace, the countries receive payoffs (z_A, z_B) . If country A attacks, then country B makes an offer $x_A \in X \equiv [0, 1]$, where x_A is the payoff it is offering to country A and $x_B \equiv 1 - x_A$ is the payoff that it is proposing for itself. Country A can either accept the offer or escalate the conflict by rejecting it. If it rejects, then country B either signs an armistice that leads to payoffs (y_A, y_B) , or it chooses total war, which results in payoffs $(0, 0)$. We assume that war is costly for both sides.”

Based on this game tree and the values labels for each country attached to action choice, the “madman” model splits possible state types into a crazy type and a strategic type that prefer certain actions over others. Acharya and Grillo basically make the assumption that the countries are in fact strategically rational, but both country A and B may mimic the crazy type in the face of its opponent’s probability of being crazy and in turn their opponent’s prior beliefs about them being crazy. From there, the “madman” model makes a few key assumptions about the rules of the game and the behavioral characteristics of states, and based on these assumptions makes various propositions about the equilibrium behavioral strategy profiles and associated belief system of the game (Acharya and Grillo, 2015). These propositions detail how states will act based on the prior believed probability of their opponents being crazy and vice versa at each node of the game, determining

unique equilibrium for countries A and B depending on their perceived probability of being crazy. Here are the main assumptions, and for the purpose of digestible reading, their overall meanings will be focused on rather than the mathematical algorithms that produced them.

*Assumption 1: War is Costly*⁷

*Assumption 2: Crazy Types are Greedy*⁸

These assumptions lay the foundation of the game in stating that the payoffs for country A and B start at peace and descend in value when reaching the payoffs from armistice, or in other words saying that with each round of escalation the payoffs get smaller as war becomes more costly. This means that the ideal, rational outcome for both strategic types of country A and B is peace, and that escalation and war should be avoided altogether. Also, even if country A attacks, both countries should be able to find pareto agreements that leave them better off than going to war. That being said, a crazy type in country A's role, or the attacking country, prefers a payoff that is greater than the one it could from the peace option. Additionally, crazy type of country A is open for an agreement that is better for country B than armistice (Acharya and Grillo, 2015). On the flip side, crazy type of country B makes an offer to country A after attack that gives worse payoffs for country A than if it had chosen peace or the conflict ended in armistice

⁷ See Assumption 1 in Appendix (Acharya and Grillo, 2015)

⁸ See Assumption 2 in Appendix (Acharya and Grillo, 2015)

(Acharya and Grillo, 2015).⁹ It is also important to note that while country B can technically offer an endless number of offers after country A attacks, in essence it can offer three types: a greedy offer as described above that benefits country B more than peace would have and gives less payoff to country A, an intermediate offer that gives the payoffs of what a war ended in armistice would be, or a concessional offer that is what the crazy type of A would prefer (Acharya and Grillo, 2015). Ultimately these assumptions about the game and the characteristics of crazy types allow us to find equilibrium behavioral strategy profiles for each country at each node, or strategic option choice, of the game tree. The model determines the behavioral strategy profiles by calculating the probability of the strategic type of country A to (1) attack, (2) reject country B's offer; and the probability of the strategic type of country B to (1) make any of the three types of offers, including the greedy type, (2) choose total war over armistice following a rejection of an offer by country A (Acharya and Grillo, 2015). The probability of country A and B being strategic or crazy are determined at each node using different equations using Bayes' rule of probability based on prior belief and updating information (Acharya and Grillo, 2015).¹⁰ The equilibrium behavioral strategy profiles of the game are broken up into five regions, with each region corresponding to the probability that country A and B is strategic or not. These descriptions of the equilibrium outcomes are stated in six propositions, where

⁹ See Equation 2 in Appendix (Acharya and Grillo, 2015)

¹⁰ See Equations 3 and 4 in Appendix (Acharya and Grillo, 2015)

the parameters of the game and the characteristics of the countries are tweaked to explore new behavioral results. Here is a description of the equilibria in each of the five regions of behavioral strategy profiles in the first proposition, which is the ideal base form of the game's equilibria¹¹:

Proposition 1:

- (i) If country B's is believed prior to not likely be strategic, then the strategic type of country A will not attack because it likely will have to bargain with the crazy type of country B, making for worse outcomes (Acharya and Grillo, 2015).
- (ii) If country B is believed to probably be strategic and country A is believed to likely be crazy, "then the strategic type of country A will attack for sure, and the strategic type of country B will try to settle the dispute early by making the concessional offer. This offer will be accepted by both the strategic and crazy types of country A" (Acharya and Grillo, 2015).
- (iii) If country A is believed to likely be strategic, and country B is only moderately likely to be strategic but is not believed to be crazy, then the strategic type of country A attacks with a probability that is related to the prior beliefs of country B about its likelihood of being crazy. "If it attacks, then the strategic type of country B mixes between the concessional offer, which would be accepted for sure, and the greedy offer, that the crazy type of country B would make. Therefore, the

¹¹ See Proposition 1 and Figure 3 in the Appendix (Acharya and Grillo, 2015)

strategic type of country B sometimes pretends to be crazy” (Acharya and Grillo, 2015).

- (iv) If country A is believed to be moderately strategic and country B is believed to likely be strategic, then country A attacks for sure and country B mixes between the concessional and greedy offer (Acharya and Grillo, 2015).
- (v) When both country A and B are believed prior to likely be strategic, “then country A mixes between attacking and taking the peaceful outcome” based on country B’s belief on country A’s likelihood to be crazy; “and following an attack, country B mixes between the concessional, intermediate and greedy offers” Acharya and Grillo, 2015)

These ideal regions are unique equilibrium in a game of incomplete information because of two forces that work against one another and offset each other (Acharya and Grillo, 2015). These two forces, called the reputation motive and the defense motive, render a strategic state indifferent about its action choices based on both countries knowing prior their beliefs about the other country being strategic, meaning that once one country knows the other country uses an equilibrium strategy, that country can select any choice with positive probability in their equilibrium strategy and end up with the same equilibrium payoff (Acharya and Grillo, 2015). The model shows that when uncertainty exists about two countries’ likelihood of being strategic, conflicts can arise even amongst strategic states that are hard to stop and may lead to war.

The “madman” model goes on to detail how equilibrium behavior changes along the five regions when the parameters of the game are tweaked. In the second proposition, the model details how behavioral equilibrium is changed when the probability of being crazy changes for country A and B in the five regions (Acharya and Grillo, 2015). Here are the comparative statistics results on the “prevalence of crazy types” (Acharya and Grillo, 2015):

Proposition 2:

- (i) For country A and B, nothing changes if country B thinks country A’s probability of being strategic increases and vice versa until they move into a new region. Country A does not attack.
- (ii) Nothing changes for country A and B with changed parameters within the region.
- (iii) The more country A’s probability of being crazy goes down, the less likely country A is to attack in order to maintain its equilibrium strategy and the indifference by country B to all of the possible offers in what is called the reputation motive. When country A attacks, it must be infrequent enough for country B to believe that country A might really be crazy this time. Country A will not reject country B’s greedy offer. Conversely, when country B’s probability of being crazy goes down, its odds of making the concessional offer go down and its probability of making the greedy offer goes up as it increases mimicking of the crazy type to avoid being exploited; this is what is called the defense motive.

- (iv) When country A's probability of being crazy decreases, country A always attacks as usual, but when presented with the greedy offer by country B they are more likely to reject the offer consistent with the reputation motive. If country B is more likely to be strategic, than it will become more likely to make the concessional offer as it becomes more appealing consistent with the reputation motive of country A.
- (v) The more country A's probability of being crazy goes down, the less likely country A is to attack in order to maintain its equilibrium strategy and the indifference by country B to all of the possible offers in what is called the reputation motive. When country A attacks, it must be infrequent enough for country B to believe that country A might really be crazy this time. Also, country A will reject country B's greedy offer with a constant positive probability consistent with the reputation motive. When country B's likelihood of being strategic goes up, it increases its probability of choosing the intermediate or greedy offer and decreases its chances of choosing the concessional offer consistent with the defense motive.¹²

Another important proposition details how equilibrium strategy profile changes depending on how greedy the states are, or how aggressive they are as crazy types (Acharya and Grillo, 2015). This can be summed fairly simply: the more aggressive the

¹² See Proposition 2 in the Appendix (Acharya and Grillo, 2015)

crazy type of country A or B is, or in other words how much they demand for in their greedy offers in the bargaining/offer node, the more the defense motive kicks in and the responding country's likelihood to mimic a crazy type goes up. The more concessional the offer country A seeks from country B, the less likely country B makes the offer. The more greedy the offer by country B, the more likely country A is to reject it so as not be exploited.¹³

The "madman" model also explicates how equilibrium changes when military advantages are added to one country versus the other.¹⁴ This is relevant because in the game, an armistice end to a war has its payoffs determined by the balance of military strength between country A and country B. The following is what occurs to the five regions in the first proposition on the basic strategic equilibria when the military strength of country A increases (Acharya and Grillo, 2015):

Proposition 4:

- (i) Nothing changes for country A or country B when country A's military strength increases in terms of country A's willingness to attack. Country A would still rather not negotiate with the crazy type of country B.
- (ii) Nothing changes for country A or country B when country A's military strength increases in terms of country A's willingness to attack. Country A was always

¹³ See Proposition 3 in the Appendix (Acharya and Grillo, 2015)

¹⁴ See Assumption 3 and Proposition 4 in the Appendix (Acharya and Grillo, 2015)

going to attack country B regardless, and country B was going to make the concessional offer regardless.

- (iii) As country A's military advantage increases, its likelihood of attacking increases as it gets higher payoffs from an armistice agreement. Conversely for country B, it makes the greedy offer less likely as country A's military advantage goes up since its payoffs from acting crazy decrease. However, the defense motive makes it ambiguous as to whether country B will make the concessional offer more often so as to not be exploited. Similarly, country A may increase or decrease its likelihood of rejecting a greedy offer from country B due to the defense motive's tension with increased payoffs of an agreement.
- (iv) Nothing changes for country A or country B when country A's military strength increases in terms of country A's willingness to attack. Country A was always going to attack country B regardless. Conversely for country B, it makes the greedy offer less likely as country A's military advantage goes up since its payoffs from acting crazy decrease. However, the defense motive makes it ambiguous as to whether country B will make the concessional offer more often so as to not be exploited. Similarly, country A may increase or decrease its likelihood of rejecting a greedy offer from country B due to the defense motive's tension with increased payoffs of an agreement.

- (v) As country A's military advantage increases, its likelihood of attacking increases as it gets higher payoffs from an armistice agreement. Conversely for country B, it makes the greedy offer less likely as country A's military advantage goes up since its payoffs from acting crazy decrease. However, the defense motive makes it likely that country B will offer the intermediate offer. Similarly, country A may increase or decrease its likelihood of rejecting a greedy offer from country B due to the defense motive's tension with increased payoffs of an agreement.

Finally, the last proposition that will be covered deals with the payoffs that each country can expect before the start of the game depending on the region. This is important for dealing with why states may choose to initiate a conflict and what they expect out of it, regardless of what the actual results may be. Based on the foundational assumptions 1 and 2 listed above, Acharya and Grillo (2015) determined how country A and B have higher or lower payoffs for each region, noting that country A has higher ex-ante payoffs than the peaceful option in regions two and four while country B has always expectations slightly less than the peace option.¹⁵ The fifth proposition details how country A and B expect their payoffs to change based on changes to prior beliefs about their likelihood of being the strategic type.¹⁶

Proposition 5:

¹⁵ See Equation 5 and 6 in the Appendix (Acharya and Grillo, 2015)

¹⁶ See Proposition 5 in the Appendix (Acharya and Grillo, 2015)

- (i) The more country A is likely to be crazy, the worse payoff country B can expect as it struggles to deal with the possibility of the crazy type. The more likely country A is to be strategic, the better it is for country B and its payoffs while the worse it gets for country A.
- (ii) The more country B is likely to be crazy, the worse payoff country A can expect as it struggles with dealing with the possibility of the crazy type. The more likely country B is to be strategic, the better it is for country A and its payoffs while the worse it gets for country B.
- (iii) The more country A is likely to be crazy, the worse payoff country B can expect as it struggles to deal with the possibility of the crazy type. The more likely country A is to be strategic, the better it is for country B and its payoffs while the worse it gets for country A.
- (iv) The more country B is likely to be crazy, the worse payoff country A can expect as it struggles with dealing with the possibility of the crazy type. The more likely country B is to be strategic, the better it is for country A and its payoffs while the worse it gets for country B.
- (v) The more country A is likely to be crazy, the worse payoff country B can expect as it struggles to deal with the possibility of the crazy type. The more likely country A is to be strategic, the better it is for country B and its payoffs while the worse it gets for country A.

There is one final proposition dealing with ex-ante payoffs based on the aggressiveness of crazy types, but the proposition is ambiguous based on the competing reputation and defensive motives (Acharya and Grillo, 2015). With all of these iterations considered, we can move to see how the model would predict interactions between North Korea and its rivals to unfold based on its interactions with other rival states.

Hypothesis Set for the “Madman” Model

The “madman” model points us to North Korea and its rivals’ perceptions of their own and the opponent’s likelihood to be “crazy” as the main independent variables for its choice of nuclear strategies. It treats North Korea’s goal of seeking nuclear weapons as a given, whereas its opponents prefer North Korea does not have nuclear weapons capabilities, creating a conflict of interests. Since North Korea’s opponents are the ones with the perceived issue of stopping North Korean nuclear weapons and could unilaterally choose the peaceful option, they are most likely to play the role of country A, the attacker, while North Korea plays the role of country B, the defender. However, North Korea can occasionally play the role of country A considering it disagrees on the issue of sanctions, lack of recognition, or other interests. These can be weighted in favor of one side versus the other in the model (Acharya and Grillo, 2015). This is an interesting part of the model’s application, as North Korea is generally presumed to be responding to an attack from another country in this analysis, not unlike how Victor Cha’s structural theory

explains North Korea's response to an unfriendly status quo of power distribution. For the same of simplicity in analysis, interactions between the United States and North Korea will be the main focus considering that both are the key players in any denuclearization agreement or conflict. There is reason to believe that North Korea is mainly focused on the United States in resolution of the conflict of interest considering the prestige of the nation and its influence in the region over other states like South Korea (Hymans, 2008). Thus, if we can establish qualitatively a reasonable threshold for prior belief that North Korea might be crazy by the United States, and vice versa, we can generate hypotheses on each of the four types of nuclear strategies as a part of an explanation of the equilibrium strategies found in proposition 1 above.

For this model to work, we must assume that a strategically rational North Korea would like to fulfill its nuclear weapons goal as efficiently as possible, and thus would use fast nuclear strategy as the status quo when possible.

Hypothesis Set 1: When the United States is Likely to Be Strategic (Country A), and North Korea is Moderately Likely to be Strategic (Country B) – (Region 3)

H1: Peace indicates that North Korea is on its preferred highest payoff strategy, which is the fast nuclear strategy.

H2: North Korea will use fast but slow-signaling nuclear strategy after the United States attacks when conveying an offer to the United States. This captures the idea that North Korea changes its status quo when it feels it has been attacked and it warrants a

response with an offer to cease escalation at the highest payoff point. North Korea is expected to mix between a greedy and concessional offer, where a strategic United States will accept the latter and depending on the reputation or defensive motive accept or reject the former, but rejection and armistice has positive probability.

H3: North Korea will employ a slow strategy when the United States has accepted its concessional offer. This represents the payoffs for the countries depending on the offer made by North Korea.

H4: North Korea will employ a fast nuclear strategy when the United States has agreed to a greedy offer from North Korea or rejected it. This implies North Korea can get its next highest payoff from peace and at a worse position for the United States if accepted, or use fast nuclear strategy with worse payoffs at armistice.

Hypothesis Set 2: When the United States is Likely to be Strategic (Country A), and North Korea is Likely to be Strategic (Country B) – (Region 5)

H1: North Korea will use fast nuclear strategy as the status quo

H2: North Korea will use a fast but slow-signaling nuclear strategy after being attacked by the United States and making an offer.

H3: North Korea will use a slow nuclear strategy when offering an intermediate offer and the United States accepts.

H4: North Korea will use a slow nuclear strategy when it offers the concessional offer and the United States accepts.

H5: North Korea will use a fast nuclear strategy when it offers the greedy offer to the United States and it rejects. If the United States accepts the greedy offer, North Korea will use fast but slow-signaling nuclear strategy with higher payoffs than armistice. If the United States rejects, North Korea will continue fast nuclear strategy but at the lower payoff of armistice.

Hypothesis Set 3: When North Korea is Likely to be Moderately Strategic (Country A) and the United States is Likely to be Strategic (Country B) – (Region 4)

H1: North Korea will employ slow but fast-signaling strategy when launching an attack as a limited form of force.

H2: North Korea will employ a slow nuclear strategy if it accepts the greedy offer of the United States.

H3: North Korea will employ a fast nuclear strategy if it accepts the concessional offer from the United States.

H4: North Korea will employ a fast nuclear strategy if it rejects the offer from the United States.

These two sets of hypotheses are a reflection of the logical conditions under which the United States and North Korea would bargain over a conflict in a crisis-bargaining situation. The other regions in the model beyond three and five are irrelevant, as they are not feasible calculations of how North Korea and the United States view their counterpart and their propensity to be a crazy type. Based on qualitative evidence, we can establish for

each period of transition the region of equilibrium North Korea and the United States correspond to. Next, following the hypotheses, we should assume changes from the status quo of fast nuclear strategy indicates that North Korea feels it has been attacked and is responding. We can then analyze the different nodes of the game tree and how this would affect North Korea's nuclear strategy. In short, this explanation should be able to tell us if North Korea and the United States are performing as the equilibrium strategy profiles would expect.

Application to the Case

Transition 1: Fast but Slow-Signaling to Slow, February 29th, 2012

There are several instances that could constitute what North Korea would deem an attack by the United States leading up to the Leap Day Agreement. The joint military exercises by the US-ROK military alliance forces drew the ire of North Korea, who threatened to turn Seoul into a sea of fire if the proceedings were not cancelled (Davenport, 2017). Additionally, the U.S. announced new bans on North Korean imports that could benefit nuclear weapons development, and perhaps most significantly a United States warship confronted and forced the reverse course of a North Korean trade ship near China (Davenport, 2017). Although there is no set threshold for an attack, it is reasonable to argue that the United States was using force to unilaterally impose its interest of denuclearizing North Korea and thus starting the first section of the game node. In testing

the “madman” model, we first need to reasonably establish which region of equilibrium the United States and North Korea correspond to based on their perceptions of the other country and their likelihood of being a “crazy” type.

Harvey Simon (2011) indicated that North Korea might be using the “madman” strategy, and US reports show that not much is known about North Korea’s leadership (Chanlett-Avery, 2016). It is reasonable to assume that the United States and North Korea are represented somewhere in region three, four, or five of the “madman” model, where the United States is likely to be strategic and North Korea is either moderately likely or likely to be strategic as well. The alternative regions would imply that the United States is unlikely to be strategic, which is intuitively not the reputation it holds in the international community, or that North Korea is likely to be crazy, in which case the United States would never risk provoking North Korea due to their belief that North Korea will demand too much or go to total war irrationally. Thus, North Korea must be believed to be moderately likely or likely to be strategically rational by the United States to provoke them, and that North Korea is at least moderately strategic otherwise it would always attack the United States. Based on the fact that Kim Jong-Un had just recently been declared the new leader with virtually no high-level diplomatic experience negotiating with the United States, it seems more likely that the United States would designate North Korea as moderately likely to be rational, and thus we should refer to hypothesis set one in examining the first transition (Chanlett-Avery, 2016).

Prior to February 29th, 2012, North Korea displayed a fast but slow-signaling nuclear strategy as it negotiated bilaterally with the United States while it simultaneously introduced and consolidated Kim Jong-Un's rise to leadership. North Korea had in 2011 expressed an interest in returning to the negotiating table about denuclearization after developing its uranium enrichment facilities and increased tensions with South Korea following shelling of the South Korean Yeonpyeong island in 2010 (Davenport, 2017). This coincided with the death of Kim Jong-Il, the former leader of North Korea, and the official announcement of Kim Jong-Un as the new supreme leader of the country in late 2011 (Davenport, 2017). Shortly after, North Korea and the United States announced their Leap Day Agreement, marking the transition into a period of slow North Korean nuclear strategy. For a brief period, North Korea seemed to agree to slow its nuclear weapons development with its moratorium and adhere to the deal faithfully. This change in nuclear strategy should thus indicate H2 and H3 in the first hypothesis set, where North Korea goes from offering a deal to the United States in response to the attack to the aftermath of the United States accepting their concessional offer.

The evidence seems to support the equilibrium behavior of North Korea having initially made a concessional offer to the United States. According to the model, a moderately strategic North Korea will choose between a concessional and a greedy offer based on the United States' probability of being strategic. Using the comparative statistics of the model, the military advantage and the low likelihood of the United States being a

crazy type decrease the chances of North Korea extending the greedy offer. Thus, the United States as either the strategic type or the crazy type accepted North Korea's concessional offer. The terms of the agreement seem to highly favor the United States considering that North Korea agreed to freeze its nuclear facility, allow outside inspectors into the country, and stop nuclear-related testing in exchange for a couple of hundred thousand tons of food aid (Davenport, 2017). This aligns with the equilibrium strategy profiles of the two countries and supports the idea that the game parameters seem to covary quite closely to the real situation.

Transition 2: Slow to Slow but Fast-Signaling, March 16th, 2012

The threat by North Korea to launch a weather satellite in honor of Kim Il-Sung marks the transition from slow to slow but fast-signaling nuclear strategy. This would indicate that North Korea was the aggressor acting as country A; using a threat of limited force to attack the United States over the conflicted interest of denuclearizing the peninsula. As discussed, for North Korea to attack the United States it must be operating within region four or five, otherwise North Korea would have been prone to more attacks on the United States to try and reap outlandish rewards from their crazy behavior.

The United States responded by sending a greedy offer, in which North Korea's payoff was much less than armistice or concessional offer. The equilibrium of the game indicates that country B will mix between the concessional and greedy offer, and that armistice is a positive probable outcome. Based on the comparative statistics of North

Korea likely being seen as less likely to be crazy than in the past having agreed to the Leap Day Agreement, the incentive for the United States to make the greedy offer have gone up in line with H1 from Hypothesis Set 3.

Transition 3: Slow but Fast-Signaling to Fast, April 12th, 2012

The actual launch of the satellite confirms that North Korea switched to a fast nuclear strategy, and also continues the equilibrium story in Hypothesis Set 3 with its predictions of armistice probability for region four. H4 indicates that North Korea will pursue a fast nuclear strategy when rejecting the greedy offer by the United States. This would support the expectation that North Korea was operating under a defense motive, where the North Korea is incentivized by the military strength of the United States and the likely strategic nature of the United States to not allow itself to be exploited by mimicking a crazy type. Although North Korea could not get the concessions it wanted, which was likely fast nuclear strategy and other benefits from the deal, the armistice arrangement provides an equilibrium outcome in which North Korea is able to pursue fast nuclear strategy but at the cost of the United States reenacting its sanctions and other forms of costs on them while the United States has the outcome of an armistice by not escalating the issue further to total war.

Transition 4: Fast to Fast but Slow-Signaling, January 10th, 2015

After three years of nuclear behavior consistent with a fast nuclear strategy, such as nuclear device tests and missile testing designed to increase North Korea's missile range

and ability to carry nuclear weapon payloads, North Korea conveys a rare public offer to suspend its planned fourth nuclear test in exchange for the cancellation of the annual US-ROK joint military exercises that have bothered North Korea in the past (Davenport, 2017). This public announcement through its official news agency, which claims North Korea offered the United States, was legitimate enough to illicit a rejection by the United States' State Department spokeswoman Jen Psaki, claiming that the nuclear issue and the military drills were separate issues (Davenport, 2017). Nonetheless, this indicates a gesture or offer with real audience costs that put North Korea's prestige on the line, since they would be compelled to agree or suffer reputation and possibly conflict costs if the United States had agreed and North Korea rescinded its offer. Thus, North Korea suddenly changed from fast to fast but slow-signaling strategy, dependent on negotiations with the United States. According to the model, we should look for an attack that triggered the offer by North Korea and qualitative evidence to support the identification of the United States and North Korea in a particular equilibrium region.

In early January right before the offer to the United States, North Korea had its sanctions expanded by the US as they continued to crack down on missile and nuclear-related government activity (Davenport, 2017). The timing of the sanctions and the damage of the sanctions intended for North Korea's nuclear weapons development seem to show that North Korea interpreted this as an attack of limited force over their conflicting

interests on the possession of nuclear weapons. Thus, North Korea extended an offer to the United States consistent with the game tree.

Additionally, there is reason to believe that the United States had updated its prior beliefs on North Korea to push it into the region five equilibrium, which details the behavioral profiles of two likely strategic countries. The decrease in North Korean attempts to engage with the United States over the roughly three year span since the Leap Day Agreement indicate that North Korea was adjusting its exogenous drop in likelihood of being crazy by decreasing the amount of attacks it made as a manifestation of the reputation motive. This is done in order to keep the United States indifferent on its strategic choices based on the probability that North Korea might be irrational. Likewise, the United States waited as well to attack considering its own prior knowledge that North Korea thinks of the United States as a likely strategic country. Thus, H2 and its expectations in the second hypothesis set seems to be congruent with the dependent variable in this transition.

Transition 5: Fast but Slow-Signaling to Fast, February 7th 2015

The return to a full, fast nuclear strategy was highlighted by the costly signal of launching an anti-ship missile with the leader Kim Jong-Un reported as present for the test (Davenport, 2017). Kim Jong-Un personally being present for an aggressive military test, open to the world, indicates North Korea's foreign policy plan to continue its pursuit of nuclear weapons with no intentions of backing or slowing down based on the costs this

would do to the national and personal image of North Korea and Kim Jong-Un respectively. Accordingly with transition four, we assume that this stage of the bargaining game reflects the United States' strategic choice and North Korea's response, showing equilibrium behavior for region five of the model.

The offer made by North Korea to the United States, in which they offered to suspend testing of nuclear weapons and missiles in exchange for the cancellation of the joint military exercise with South Korea, was rejected by the United States on the grounds that the two issues were not related (Davenport, 2017). Regarding the terms of the deal and the payoffs in the game tree, North Korea had clearly made a concessional offer based on the slowing down of their entire nuclear weapons testing in exchange for the cancellation of a military display. This was definitely above the armistice level of payoffs considering that conflict would not have forced North Korea to commit to that outcome. The reason for the United States' rejection can be explained through the defense motive, where despite North Korea's concessional offer, the high probability of the US being a strategic type in North Korea's eyes extends the incentive for the US to endogenously mimic crazy type behavior at this stage of the game to ensure North Korea doesn't exploit them by pretending to be crazy themselves. The United States has to keep North Korea indifferent to their strategy choices in equilibrium based on the chance that the United States might actually go to suboptimal outcomes irrationally.

As expected in equilibrium, or in H5 of hypothesis set two, North Korea chooses armistice in region five with positive probability by refusing to escalate the issue further after the rejection and living with the payoff of pursuing nuclear weapons at full pace despite the costs of the exchange inhibited upon them by the United States.

Transition 6: Fast to Fast but Slow-Signaling, July 6th, 2016

On July 6th, 2016, a high-level North Korean government spokesman released a statement declaring North Korea's openness to resume denuclearization talks provided conditions were met on the agreed definition of denuclearization and the ceasing of threatening behavior between countries (Carlin, 2016). This message constitutes a costly signal wherein North Korea would suffer audience costs if it backed down from its willingness to cooperate, yet the United States responded by adding Kim Jong-Un and other North Korean leaders to its designations of human rights abuses (Carlin, 2016). This would indicate that North Korea perceived an attack, and that North Korea and the United States are adhering to equilibrium outcomes based on their prior beliefs about the other country possibly being crazy.

At this point, there is mounting evidence to assume that North Korea and the United States consider each other likely strategic states. Kim and Cohen (2017) and Pollack (2017) outline how the United States has adapted its approach to North Korea with the expectation that North Korea is not far from deterrence and using their nuclear strategy wisely to achieve their goals. This would situate North Korea and the United States within

region five of the equilibrium profiles, but within the region the United States also seems to be making an effort to change North Korea's prior beliefs about the chance the United States will stop at nothing, including total war. The advocating of China and the use of the UN Security Council to condemn North Korean nuclear weapons development, along with the attack of increased sanctions and designations of human rights abuses on Kim Jong-Un personally constitute an attack (Davenport, 2017).

Transition 7: Fast but Slow-Signaling to Fast, August 3rd, 2016

North Korea returned to completely fast nuclear strategy through the costly signaling of arms build up, or a message through missile testing in August of 2016. This indicated North Korea was once again unwilling to back down on its nuclear ambitions and was followed by a slew of submarine-based missile tests, ICBM tests, and a fifth nuclear test (Davenport, 2017). The return to fast nuclear strategy is explained in the model's second hypothesis set H5, where North Korea's strategic choices situated within an equilibrium that can be explained through comparative statistics.

The successful testing of SLBM's, ICBM's, and a nuclear device indicate North Korea is closer than ever to second-strike capability and viable preemptive/preventive action through nuclear means (Armstrong, 2017; Kim and Cohen, 2017; Pollack, 2017). While the THAAD system was implemented in South Korea, North Korea has estimated strike range of up to Los Angeles, Denver, and Chicago (Davenport, 2017; Kim and Cohen, 2017). The improvements in nuclear capabilities indicate a relative improvement in

military strength that is used to calculate the costs of armistice between the United States and North Korea. Additionally, considering that the United States' prior beliefs about North Korea being strategic were going up, North Korea has a defense motive incentive to mimic the crazy type in order to stop exploitation from the United States when they attack. This explains why North Korea chose a greedy deal to offer the United States on denuclearization, where if denuclearization were to occur it would occur it would happen on North Korea's terms and by their definition (Carlin, 2016). These conditions are not ideal for the surplus the United States prefers, and they have a reputation motive to reject greedy offers to promote the idea that they cannot be considered 100% rational. In the end, this lead to the equilibrium outcome of armistice, where North Korea pursues fast nuclear strategy short of real war and enjoys significantly costlier payoffs than if an early agreement had been made.

Analysis:

Overall, the “madman” model compelling explains how even strategically rational states must mimic crazy behavior if there is even a small chance that their opposing country think they are irrational. It makes sense sometimes to act as if you are crazy, otherwise another strategic state will use its reputation for craziness to exploit you. Likewise, a country can try to leverage the uncertainty about their true intentions to get higher payoffs in a conflict. The model gives a roadmap to when and why states might change their nuclear strategies, and when applied to North Korea, the ideal regions of

equilibrium offer innovative explanations. The weakness in the approach used in the study is that it uses a congruence test to see whether the evidence matches the expectations of the model without eliminating the possibilities of other independent variables being responsible. In other words, it is hardly causality. Additionally, the game tree and the parameters might need to be adjusted for the North Korean case to see how subsequent interactions between the states affects their likeliness of acting crazy in the future. For instance, it is possible that North Korea initially made a concessional offer as part of the bargain on the Leap Day Agreement, but shortly after mixed in a greedy offer, which could indicate multiple bargaining opportunities or perhaps that the two should be treated as one offer. This approach also qualitatively determined the regions of equilibrium and various comparative statistics that were applied to explain North Korean nuclear strategy; this is dangerously ad-hoc but nonetheless the results seem intuitive and based on compelling anecdotal evidence from US Congressional Reports and scholars (Chanlett-Avery, 2016; Kim and Cohen, 2017; Pollack, 2017). It seems that based on these findings that North Korea actively leverages its reputation of possibly being crazy to try to find optimal strategic outcomes. As time has gone on, the United States has treated North Korea more and more like a rational, strategic state with a low chance of being actually irrational, and hence the lack of engagement with North Korea. North Korea also seems to interpret sanctions and acts against its nuclear weapons developments as threats considering the immediacy with which North Korea offers some form of bargaining to the United States

following these acts. For its part, the United States seems to be intent on its own reputation and defense motives to maintain the prospect that North Korea must respect they may not be completely rational, and if push came to shove there are scenarios where the United States would take suboptimal outcomes like total war. All in all, it seems viewing North Korean nuclear strategy as a part of the bargaining dynamics between them and the United States and other rivals is a fruitful research program.

Conclusion:

This paper was an attempt to explain the variance in North Korean nuclear strategy displayed under the rule of Kim Jong-Un. Specifically, this paper argued that we can think about North Korean nuclear strategy as either going fast or slow, and that part of its strategy is dependent on the actions of other actors. Additionally, two theoretical models that had very different analytical focuses were used to compare their effectiveness in answering the question of how North Korea tries to achieve its nuclear goals. The coercive bargaining model represents a structural neorealist approach with a twist of prospect theory, and the “madman” theory represents a crisis-bargaining approach with a basis in game theory. Both models provide compelling cases to show that North Korea is in fact rational, and that depending on the circumstances, it is sometimes rational to act irrational even, or to change course in your nuclear strategy in unexpected ways.

The coercive bargaining model hypothesized that North Korea would change its nuclear strategy based on whether it found itself in a winning or losing situation from its current status quo in the international system. Depending on whether it was winning or losing, North Korea would favorably view fast or slow nuclear strategy, and possibly hedge depending on the risks involved. Based on this study, it seems very likely that North Korea filters its nuclear strategy options to some degree because of the anarchic pressures placed on it by the power distribution of the international system. It also seems reasonable to think that North Korea bases its decision-making around risk propensities. The structure of North Korea's environment alone; however, cannot fully explain why North Korea agreed to the Leap Day Agreement at such a low payoff and subsequently broke from the deal almost immediately after agreeing to it. The "madman" theory presented unique equilibrium strategic profiles of countries with different properties that bargained with each other over a conflicted issue. The key to the model was the prior beliefs of each country about how likely they thought another country was rational or not. The model demonstrates that even a small chance of a country being crazy changes how strategically rational countries bargain, and that there are a lot of circumstances where a rational state try to act crazy even when they are not to get higher payoffs from their bargains. The model was originally used to determine when and why war will broke out when better options are available, but using the positive probabilities of the equilibriums, we can also explain how North Korea goes back and forth with the United States about its nuclear

weapons and adjusts its strategy accordingly to try and reach the best outcome. North Korea had a reputation for being moderately likely to be crazy in the eyes of the United States, while the United States was considered very likely to be strategically rational. This explains why North Korea and the United States entered an agreement and broke from it early in Kim Jong-Un's regime by performing close to equilibrium expectations. Comparative statistics showed how changes in the way the US viewed North Korea's rationality changed the strategies and outcomes of their bargains.

The study had significant limitations, most notably the use of a congruence test approach. By only comparing the hypotheses of each model with the dependent variable to see if the independent variable co-varied, we were able to reasonably assume that there is something to it, but without actually tracing the process in which North Korea chooses its nuclear strategy we cannot be confident in any claims. Unfortunately, unless more access to North Korea is opened, case study approaches like this one will be severely limited. Additionally, attempts to modify the original theoretical models to produce testable hypotheses may have stretched the models beyond their intent. For the "madman" model, a more comprehensive game theoretic approach that more systematically tests case studies against equilibrium expectations could fruitfully improve this paper, along with a modification of the game tree to include more bargaining opportunities. The findings of the paper indicate that Victor Cha's original theory still has utility in the right contexts, and that the "madman" theory deserves far more research in uncovering the strategic ways

in which actors try to misrepresent information about themselves for their own purposes and whether its effective or not.

For future implications, the coercive bargaining model would point to North Korea's improving nuclear weapons program as a basis for finding themselves in a domain of gains, and this is perhaps coming to the best time to engage North Korea diplomatically. The "madman" model has exciting research opportunities in examining how President Trump may be shifting North Korea's perceptions about the United States' rationality, and what effects this might have on equilibrium strategies and outcomes. Finally, studies on nuclear behavior can only go so far without incorporating the why question. While it provides clean analytical divisions to study just the how, the reasons why a country wants a nuclear weapons can greatly influence their decision-making and strategy on how to get them. These models mostly leave these factors unexplored, and future research could benefit from acknowledging the complexity of decision-making and strategy with the origin of nuclear preferences.

Appendix:

Table 1: North Korean Nuclear Strategy – Dependent Variable Values

	Fast Nuclear Strategy	Fast but Slow-Signaling Nuclear Strategy	Slow but Fast-Signaling Nuclear Strategy	Slow Nuclear Strategy
Speed	Fast	Fast	Slow	Slow
Signaling	Fast	Slow	Fast	Slow

Table 2: History of North Korean Nuclear Strategy Transitions from 2011-2017

	Trans. 1	Trans. 2	Trans. 3	Trans. 4	Trans. 5	Trans. 6	Trans. 7
Date	Feb. 29 th 2012	March 16 th 2012	April 13 th 2012	January 10 th , 2015	February 7 th 2015	July 6 th , 2016	August 3 rd 2016
Nuclear Strategy	Fast but Slow-Signaling to Slow	Slow to Slow but Fast-Signaling	Slow but Fast-Signaling to Fast	Fast to Fast but Slow Signaling	Fast but Slow-Signaling to Fast	Fast to Fast but Slow-Signaling	Fast but Slow-Signaling to Fast

Table 3: How Do States Frame the Status Quo? *modified from original model

	Domain of Gains	Neutral	Domain of Losses
Preemptive or Preventive Situation with Offense Having the Advantage	Slow but Fast-Signaling/Slow Nuclear Strategy (Indeterminate)	Fast but Slow-Signaling Nuclear Strategy	Fast Nuclear Strategy
Preemptive or Preventive Situation with Defense Having the Advantage	Slow Nuclear Strategy	Slow but Fast-Signaling Nuclear Strategy	Fast but Slow-Signaling Nuclear Strategy

Figure 1: Base Game Tree from Acharya and Grillo (2015)

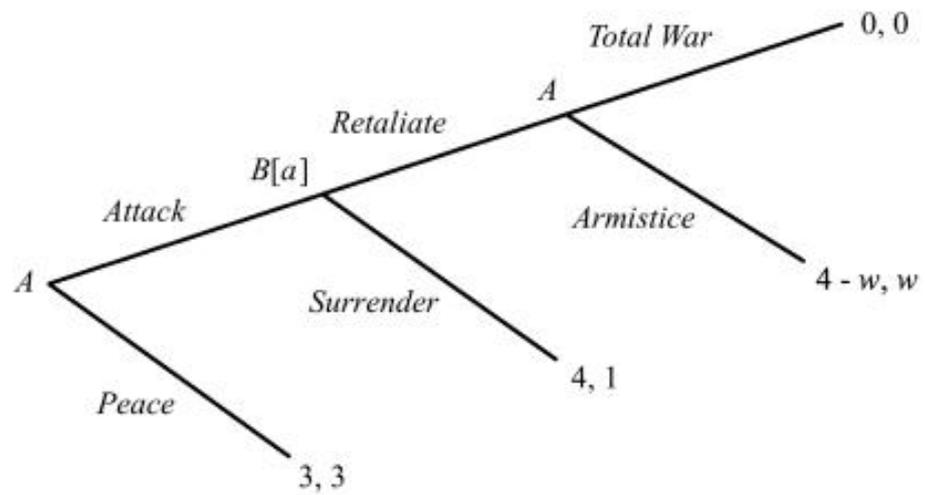


Figure 1: Game Tree for the Illustrative Example

Figure 2: "Madman" Model Game Tree

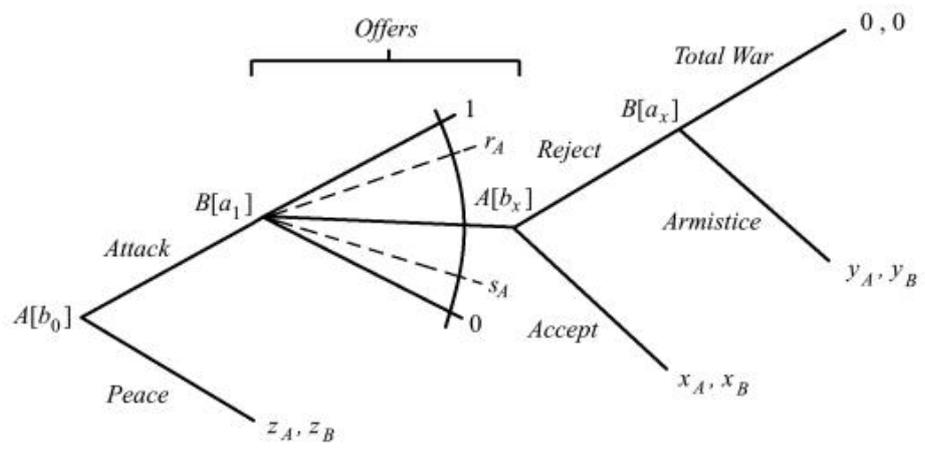
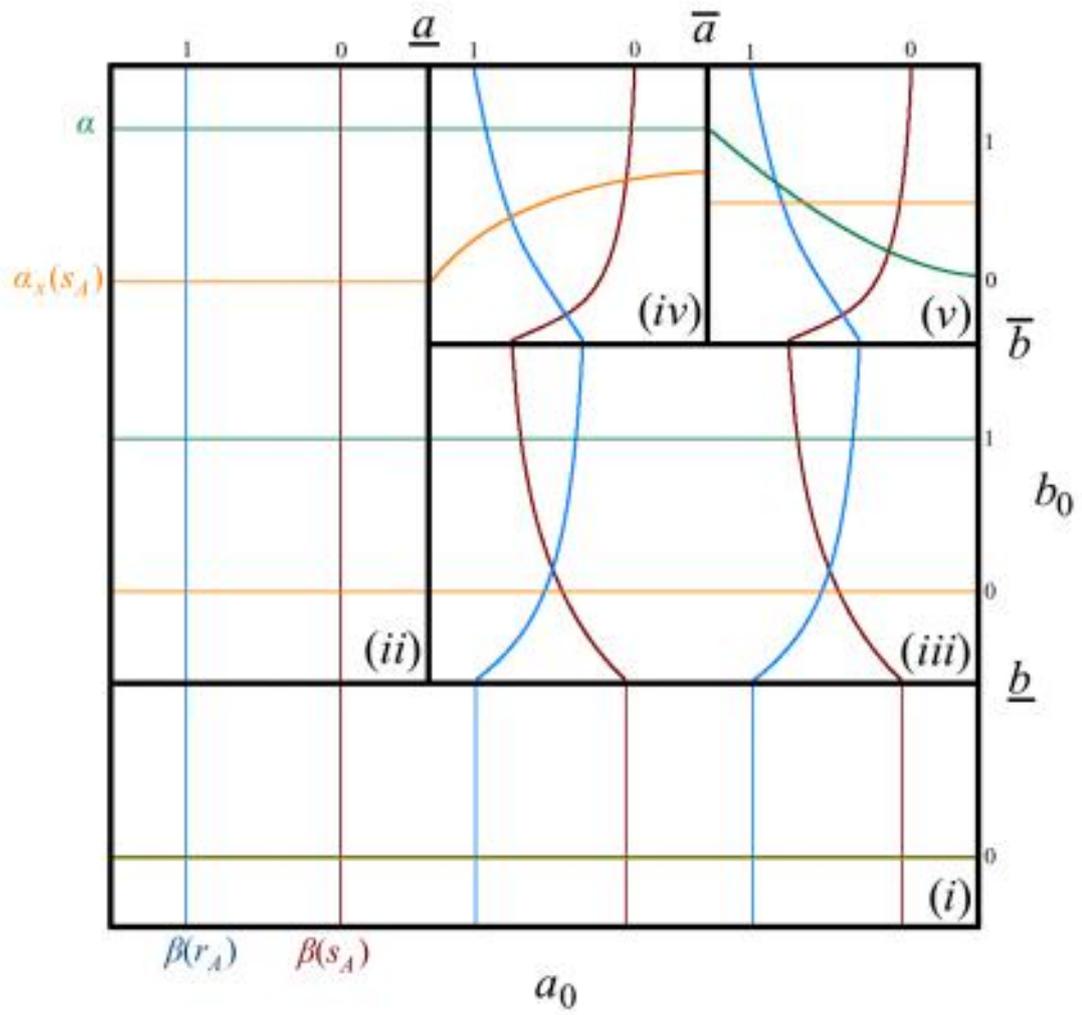


Figure 2: The Game Tree

Figure 3: Figure of Proposition 1 and Equilibrium of Five Regions (Acharya and Grillo, 2015)



Equation 1: Equilibrium Probability of War from Acharya and Grillo (2015)

$$\underbrace{\left[(1-p) + \frac{1-p}{p} \cdot \frac{1}{w-1} \right]}_{\text{Pr}[A \text{ attacks}]} \times \underbrace{\left[\frac{1}{w} \right]}_{\text{Pr}[B \text{ retaliates}]}$$

Equation 2: Assumptions 1 & 2 Combined (Acharya and Grillo, 2015)

$$1 > r_A > y_A > s_A > 0$$

Equation 3: Country B's Updated Belief that Country A is Strategic Following Attack (Acharya and Grillo, 2015)

$$a_1 \equiv \frac{\alpha a_0}{1 - a_0 + \alpha a_0}$$

Equation 4: Country A's Updated Belief that Country B is Strategic Following an Offer (Acharya and Grillo, 2015)

$$b_x(x_A) = \begin{cases} 1 & \text{if } x_A \neq s_A \\ \frac{\beta(s_A)b_0}{1-b_0+\beta(s_A)b_0} & \text{if } x_A = s_A. \end{cases}$$

Equation 5: Country A's Ex-Ante Expected Payoffs

$$V_A = \begin{cases} z_A & \text{in regions (i), (iii), (v)} \\ b_0 r_A + (1 - b_0) s_A & \text{in region (ii)} \\ b_0 \beta(r_A) r_A + [1 - b_0 \beta(r_A)] s_A & \text{in region (iv)} \end{cases}$$

Equation 6: Country B's Ex-Ante Expected Payoffs

$$\begin{aligned} V_B &= a_0[(1 - \alpha)z_B + \alpha(1 - r_A)] + (1 - a_0)(1 - r_A) \\ &= z_B - (1 - a_0 + a_0\alpha)[z_B - (1 - r_A)] \end{aligned}$$

Assumption 1: War is Costly (Acharya and Grillo, 2015)

- (i) $z_A > y_A > 0$, (ii) $z_B > y_B > 0$, and (iii) $z_A + z_B > 1 > y_A + y_B$.

Assumption 2: Crazy Types are Greedy (Acharya and Grillo, 2015)

- (i) $1 - y_B > r_A > z_A$ and (ii) $\min\{1 - z_B, y_A\} > s_A > 0$.

Assumption 3: Payoff Profiles where Country A is Stronger than Country B (Acharya and Grillo, 2015)

$$y_A + y_B = y'_A + y'_B = \bar{y}, \text{ and } y'_A > y_A.$$

Proposition 1: Equilibria of the Game for the Five Behavioral Strategic Profile Regions (Acharya and Grillo, 2015)

(i) If $b_0 < \underline{b}$ then in every equilibrium, we have $\alpha = 0$,

$$\alpha_x(x_A) \in \begin{cases} \{0\} & \text{if } x_A = s_A \text{ or } x_A > y_A \\ [0, 1] & \text{if } x_A = y_A \\ \{1\} & \text{if } x_A < y_A \text{ and } x_A \neq s_A, \end{cases} \quad (*)$$

$\beta(r_A) = 1$ and $\beta(x_A) = 0$ for all $x_A \neq r_A$.

(ii) If $b_0 > \underline{b}$ and $a_0 < \underline{a}$, then in every equilibrium, we have $\alpha = 1$, α_x is given by (*) above, $\beta(r_A) = 1$ and $\beta(x_A) = 0$ for all $x_A \neq r_A$.

(iii) If $\bar{b} > b_0 > \underline{b}$ and $a_0 > \underline{a}$, then in every equilibrium, we have $\alpha = \frac{1-a_0}{a_0} \cdot \frac{1-r_A-y_B}{r_A-s_A}$, α_x is given by (*) above, $\beta(s_A) = 1 - \frac{z_A-s_A}{b_0(r_A-s_A)}$, $\beta(r_A) = \frac{z_A-s_A}{b_0(r_A-s_A)}$, and $\beta(x_A) = 0$ for all $x_A \neq s_A, r_A$.

(iv) If $b_0 > \bar{b}$ and $\bar{a} > a_0 > \underline{a}$, then in every equilibrium we have $\alpha = 1$,

$$\alpha_x(x_A) \in \begin{cases} \{1\} & \text{if } x_A < y_A \text{ and } x_A \neq s_A \\ \left\{1 - \frac{1-r_A-y_B}{a_0(1-s_A-y_B)}\right\} & \text{if } x_A = s_A \\ [0, 1] & \text{if } x_A = y_A \\ \{0\} & \text{if } x_A > y_A \end{cases}$$

$\beta(s_A) = \frac{1-b_0}{b_0} \frac{s_A}{y_A-s_A}$, $\beta(r_A) = 1 - \frac{1-b_0}{b_0} \frac{s_A}{y_A-s_A}$ and $\beta(x_A) = 0$ for all $r_A \neq s_A, r_A$.

(v) If $a_0 > \bar{a}$ and $b_0 > \bar{b}$ then in the unique equilibrium $\alpha = \frac{1-a_0}{a_0} \frac{1-r_A-y_B}{r_A-y_A}$,

$$\alpha_x(x_A) = \begin{cases} 1 & \text{if } x_A < y_A \text{ and } x_A \neq s_A \\ \frac{y_A-s_A}{1-s_A-y_B} & \text{if } x_A = s_A \\ 0 & \text{if } x_A \geq y_A, \end{cases}$$

$\beta(s_A) = \frac{1-b_0}{b_0} \frac{s_A}{y_A-s_A}$, $\beta(y_A) = 1 - \frac{z_A-b_0y_A}{b_0(r_A-y_A)} - \frac{(1-b_0)s_A}{b_0(y_A-s_A)}$, $\beta(r_A) = \frac{z_A-b_0y_A}{b_0(r_A-y_A)}$ and $\beta(x_A) = 0$ for all $x_A \neq s_A, y_A, r_A$.

Proposition 2:

- (1) α is continuous and weakly decreasing in a_0 for all b_0 . Furthermore, it is strictly decreasing in a_0 only in regions (iii) and (v).
- (2) $\beta(r_A)$ is constant (and equal to 1) in regions (i) and (ii), strictly decreasing in b_0 in region (iii) and (v) and strictly increasing in b_0 in regions (iv).
- (3) $\beta(s_A)$ is constant (and equal to 0) in regions (i) and (ii), strictly increasing in b_0 in region (iii) and strictly decreasing in b_0 in regions (iv) and (v).
- (4) $\beta(y_A)$ is constant (and equal to 0) in regions (i) to (iv) and strictly increasing in region (v).
- (5) $\alpha_x(s_A)$ is constant (and equal to 0) in regions (i) to (iii), strictly increasing in a_0 in region (iv) and constant (strictly between 0 and 1) in region (v).

Proposition 3:

- (1) The probability $\beta(r_A)$ with which country B makes the concessional offer is weakly decreasing in r_A .
- (2) The probability $\alpha_x(s_A)$ with which country A rejects the greedy offer of country B is weakly decreasing in s_A .

Proposition 4:

- (1) If $a_0 < \underline{a}$ or if $b_0 < \underline{b}$ or if $a_0 \in (\underline{a}, \bar{a})$ and $b_0 > \bar{b}$, then $\alpha' = \alpha$; otherwise, $\alpha' > \alpha$.
- (2) If $a_0 < \underline{a}$ or if $b_0 < \underline{b}$ or if $a_0 > \underline{a}'$ and $b_0 \in (\underline{b}, \bar{b}')$, then $\beta'(s_A) = \beta(s_A)$; otherwise, $\beta'(s_A) < \beta(s_A)$.
- (3) If $a_0 < \underline{a}$ or if $b_0 < \underline{b}$ or if $a_0 > \underline{a}$ and $b_0 \in (\underline{b}, \bar{b}')$, then $\beta'(r_A) = \beta(r_A)$; if $a_0 > \bar{a}'$ and $b_0 > \bar{b}'$ then $\beta'(r_A) < \beta(r_A)$; otherwise, $\beta'(r_A) > \beta(r_A)$.
- (4) If $a_0 > \bar{a}'$ and $b_0 > \bar{b}'$ then $\beta'(y_A) > \beta(y_A)$; if $a_0 \in (\bar{a}, \bar{a}')$ and $b_0 > \bar{b}$ then $\beta'(y_A) < \beta(y_A)$; otherwise, $\beta'(y_A) = \beta(y_A)$.
- (5) If $a_0 > \underline{a}'$ and $b_0 \in (\bar{b}', \bar{b})$ or if $a_0 \in (\bar{a}, 1)$ and $b_0 > \bar{b}$ then $\alpha'_x(s_A) > \alpha_x(s_A)$; if $a_0 \in (\underline{a}, \bar{a})$ and $b_0 > \bar{b}$ then $\alpha'_x(s_A) < \alpha_x(s_A)$; otherwise, $\alpha'_x(s_A) = \alpha_x(s_A)$.

Proposition 5:

- (1) Country A 's *ex ante* expected payoff V_A is:
 - (i) weakly increasing in b_0 and strictly increasing in regions (ii) and (iv).
 - (ii) piecewise constant in a_0 with downward jumps at the boundaries of regions (ii), (iii), (iv) and (v).
- (2) Country B 's *ex ante* expected payoff V_B is:
 - (i) weakly increasing in a_0 and strictly increasing in regions (i), (iii) and (v).
 - (ii) piecewise constant in b_0 with downward jumps at the region boundaries.

References:

- Abreu, D. and Gul, F. 2000. "Bargaining and Reputation" *Econometrica* 68:1: 85-117
- Acharya, A., & Grillo, E. 2015. "War with crazy types." *Political Science Research and Methods* 3.2: 281-307.
- Allison, Graham T., and Philip Zelikow. 1999. *Essence of Decision: Explaining the Cuban Missile Crisis*. Vol. 2. New York: Longman.
- Armstrong, C. K. 2017. "North Korea in 2016." *Asian Survey*, 57(1): 119-127.
- Axelrod, Robert, and Robert O. Keohane. 1985. "Achieving cooperation under anarchy: Strategies and institutions." *World Politics* 38.01: 226-254.
- Banks, J. 1990. "Equilibrium Behavior in Crisis Bargaining Games." *American Journal of Political Science* 33:3: 599-614.
- Carlin, Robert. 2016. "North Korea Said it is Willing to Talk about Denuclearization... But No One Noticed." *38 North*. US-Korea Institute at John Hopkins SAIS.
- Cha, Victor D., and David Chan-oong Kang. 2003. *Nuclear North Korea: A Debate on Engagement Strategies*. Columbia University Press.
- Cha, Victor D., & David Chan-oong Kang. 2004. "The debate over north korea." *Political Science Quarterly* 119.2: 229-254.
- Chanlett-Avery, E., Rinehart, I. E., and Nikitin, M. B. D. 2016. "North Korea: US relations, nuclear diplomacy, and internal situation." *Congressional Research Service*.
- Davenport, Kelsey. 2017. "Chronology of US-North Korean Nuclear and Missile Diplomacy Fact Sheet." *Arms Control Association*. December.
- DiCicco, Jonathan M. and Levy, Jack S. 1999. "Power Shifts and Problem Shifts: The Evolution of the Power Transition Research Program." *Journal of Conflict Resolution*, 43(6), 675-704..
- Fearon, J. 1995. "Rationalist Explanations for War." *International Organization* 49:3: 379-414.

- Fearon, J. 1997. "Signaling foreign policy interests: Tying hands versus sinking costs." *Journal of Conflict Resolution*. 41.1: 68-90.
- Fearon, J. 1994. "Signaling versus the balance of power and interests: An empirical test of a crisis bargaining model." *Journal of Conflict Resolution* 38.2: 236-269.
- Goo, Y., & Kim, S. 2009. "A study on the military alliance of south korea-united states with the existence of threat from north korea: A public good demand approach." *Empirical Economics* 36.3: 597-610.
- Goo, Y., & Lee, S. 2014. "Military alliances and reality of regional integration: Japan, south korea, the US vs. china, north korea." *Journal of Economic Integration* 29.2: 329-342.
- Hart, B. L. 1967. *Strategy*. New York: Frederick A.
- He, Kai, and Huiyun Feng. 2013. *Prospect Theory and Foreign Policy Analysis in the Asia Pacific: Rational Leaders and Risky Behavior*. Routledge.
- Hee-gwan, C., Kwang-seo, K., Gwang-oon, K., Keun-sik, K., Yeon-chul, K., Gee-dong, L., ... & Bo-hyuk, S. 2013. *Understanding North Korea: Indigenous Perspectives*. Lexington Books.
- Hymans, Jacques. 2004. *The Psychology of Nuclear Proliferation: Identity, Emotions, and Foreign Policy*. New York: Cambridge University Press.
- Hymans, Jacques. 2008. "Assessing North Korean nuclear intentions and capacities: A new approach." *Journal of East Asian Studies* 8.2: 259-292.
- Jervis, Robert. 1978. "Cooperation under the security dilemma." *World Politics* 30.02:167-214.
- Jervis, Robert. 1982. "Deterrence and Perception." *International Security* 7(3): 3-30.
- Jervis, Robert. 1998. *System Effects: Complexity in Political and Social Life*. Princeton University Press.
- Kaplan, F. 1991. *Wizards of Armageddon*. Palo Alto: Stanford University Press.

- Kahneman, Daniel, and Amos Tversky. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica: Journal of the Econometric Society* 263-291.
- Kahneman, Daniel, and Amos Tversky, eds. 2000. *Choices, Values, and Frames*. Cambridge University Press.
- Keohane, Robert O. 2005. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton University Press.
- Keohane, Robert O., and Joseph S. Nye. 1977. *Power and Interdependence: World Politics in Transition*. 2nd ed. Boston: Little, Brown.
- Kim, S. C., and Cohen, M. D. (Eds.). 2017. *North Korea and Nuclear Weapons: Entering the New Era of Deterrence*. Georgetown University Press.
- Kimball, J. 2004. *The Vietnam War Files: Uncovering the Secret History of Nixon Era Strategy*. Lawrence: University Press of Kansas.
- Lake, D. 2011. "Two Cheers for Bargaining Theory: Assessing Rationalist Explanations of the Iraq War". *International Security* 35:3:7-52
- Levy, Jack S. 1992. "Prospect theory and international relations: Theoretical applications and analytical problems." *Political Psychology* 283-310.
- Luttwak, E. 2001. *Strategy: the logic of war and peace*. Harvard University Press.
- Luttwak, E. 2009. "The grand strategy of the Byzantine Empire." *Center for Strategic and International Studies in Washington*, 600, 3. Chicago
- Mearsheimer, John J. 2001. *The Tragedy of Great Power Politics*. WW Norton & Company.
- Morganthau, Hans J. 1960. *Politics Among Nations: The Struggle for Power and Peace*. AA Knopf.
- Organski, A.F.K. 1958. *World Politics*. New York: Knopf

- Pollack, J. D. 2017. *No exit: North Korea, nuclear weapons, and international security*. Routledge.
- Powell, R. 1987. "Crisis Bargaining, Escalation and MAD." *American Political Science Review* 81:3:717-36.
- Powell, R. 2002. "Bargaining theory and international conflict." *Annual Review of Political Science* 5(1): 1-30.
- Schultz, K. 1999. "Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War," *International Organization* 53:2: 233-66.
- Shimko, Keith I. 1992. "Realism, neorealism, and American liberalism." *The Review of Politics* 54.02: 281-301.
- Simon, Harvey. 2013. "Kim Jong-un and the "Madman Theory" of Diplomacy". *Origins*. Published by the History Departments at The Ohio State University and Miami University.
- Snyder, G. H., & Diesing, P. 2015. *Conflict among nations: Bargaining, decision making, and system structure in international crises*. Princeton University Press.
- Tammen, R., Kugler, J., Lemke, D., Stam, A., Abdollahian, M., Alsharabati, C., & Organski, A. F. K. 2000. *Power transitions*. Chatham House, New York.
- Van Evera, S. 1984. "The cult of the offensive and the origins of the First World War." *International Security*, 9(1): 58-107.
- Von Clausewitz, C. 1832. *On War*, ed. and trans. Michael Howard and Peter Paret. Princeton, NJ: Princeton Univ. Press, 1976.
- Von Neumann, J., & Morgenstern, O. 1945. *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.
- Walt, Stephen M. 1987. *The Origins of Alliance*. Cornell University Press.
- Waltz, Kenneth N. 2001. *Man, the State, and War: a Theoretical Analysis*. Columbia University Press.

Waltz, Kenneth N. 2010. *Theory of International Politics*. Waveland Press.

Weisiger, A., and Yarhi-Milo, K. 2015. "Revisiting reputation: How past actions matter in international politics." *International Organization*, 69(2): 473-495.

Yarhi-Milo, K., Kertzer, J. D., and Renshon, J. 2017. "Tying Hands, Sinking Costs, and Leader Attributes."

국문초록:

정책 입안자들과 정치학자들은 북한이 핵무기 국가가 되려고 한다고 동의하지만, 북한이 그 목표를 실현시키려고 하는 방법에 대해서는 의견이 분분하다. 과거 수십년동안, 북한은 예상치 못하게, 또는 겉보기에는 비이성적인 방법으로 핵무기를 달성하는 것을 종종 늦추거나 가속화 시켰다. 북한의 다양한 핵무기 전략을 설명하는 것은 김정은의 집권 이후, 그가 매일 대륙간, 제 2 격의 핵무기 능력을 성장시키면서 더욱 더 중요해졌다. 이러한 연구 주제에 대한 몇몇 이론적 접근을 평가하기 위해, 본 연구는 북한의 핵전략에 적용되는 두 가지 이론적 모델의 설명력을 검토하고 비교하고자 한다: Victor Cha 에 의해 고안된 강압적 거래 모델과 Avidit Acharya 과 Edoardo Grillo 에 의해 고안된 "매드맨" 모델이다. 권력 이행 이론과 전망 이론에 기반한 강압적 거래 모델은, 북한의 위험한 핵전략 선택 수용을 북한의 국제 사회에서의 자신의 위치에 대한 인식을 통해 합리화한다. 양쪽의 불완전한 정보 게임 이론에 기반한 "매드맨" 모델은, 그 자신을 위해 더 큰 부당 전략적 결과에 대해 북한이 "미치게 행동"하는 논리를 극찬한다. 일치성 검사가 모델들의 예측이 북한의 신속한 것에서부터 느린것까지, 또는 그 반대이거나, 2011 년부터 2017 년까지의 핵전략에서의 이행과 서로 달라지는지 사용되었다.