ASAN
REPORT

Implications of Perfect Deterrence Theory for South Korea

Kim Chong Woo

September 2020

Implications of Perfect Deterrence Theory for South Korea

KIM CHONG WOO

SEPTEMBER 2020

About

Acknowledgement

Disclaimer

Author

KIM Chong Woo

Dr. KIM Chong Woo is a Senior Research Fellow and oversees the Quantitative Methods Program at the Asan Institute for Policy Studies. Previously, he was an analyst working on choice modelling and valuation at RAND Europe. He was also a Senior TCAD Engineer at the Samsung Semiconductor Research and Development Center and a Java Application Developer at PCMS-Datafit, UK. Dr. Kim's research focuses on the estimation and application of Discrete Choice Modelling; Stated Preference Analysis, valuing public services and non-market goods; SP model development in transport, health, communication and utilities sector. He has published widely including "Security at What Cost? Quantifying Individuals' Trade-offs between Privacy, Liberty and Security," RAND Report (2010) and "Modelling Demand for Long-Distance Travellers in Great Britain: Stated preference surveys to support the modelling of demand for high speed rail," RAND Report (2011). He received his B.Sc. in Mathematics from the University of London and his Ph.D. in Mathematical Physics from Imperial College of Science, Technology and Medicine, London. He also holds a post-graduate Diploma in Computer Science from the University of Cambridge.

Table of Contents

List of Figures

List of Tables

Executive Summary

Despite the Singapore and Hanoi summit meetings between Trump and Kim Jong-un and inter-Korean summits between Moon Jae-in and Kim Jong-un, there has not been any real progress towards *'final, fully-verified denuclearization'* over two years. If anything, all signs point to the contrary. At a military meeting in May, 2020, Kim Jong-un was reported to have called for stronger "nuclear war deterrence" by the state media.[1] In the following month, North Korea blew up an inter-Korean liaison office in the North Korean city of Kaesong. With the prospect of North Korea's denuclearization diminishing, it is imperative that *'deterrence,'* our subject of interest, is robustly maintained against the North Korea's growing nuclear threat. In particular, we have become interested in Perfect Deterrence Theory developed by Zagare & Kilgour as an alternative to classical deterrence theory. Above all, its predictions are more in agreement with empirical findings and free of irrationality and, therefore, the theory has been chosen to be the basis of our study.[2] It has basically provided a framework for exploring South Korea's deterrence against North Korea. Our role was to understand its solutions[3] (i.e., *'Perfect Bayesian equilibria'*) found by Zagare & Kilgour, and to draw out valuable implications in the context of North-South Korea. Our study was limited to *Direct Deterrence*[4] which includes the *Generalized Mutual Deterrence Game* and the *Unilateral Deterrence Game*. Our study has drawn extensively on their book *"Perfect Deterrence."*[3]

Chapter 1 is a brief overview of Perfect Deterrence Theory. Some advantages of this theory over classical deterrence theory are mentioned. One advantage is that Zagare & Kilgour's imposition of the *'Perfectness'* condition prevents us from encountering irrational solutions later on (e.g., classical deterrence theory cannot explain the *paradox of mutual deterrence*). The theory also explores two differentiated players (i.e., challenger and

---

1. "Kim Jong-un calls for greater 'nuclear war deterrence,'" The Korea Herald, 24th May 2020. http://www.koreaherald.com/view.php?ud=20200524000204.

2. This author is solely responsible for any errors of fact or misinterpretation of Perfect Deterrence Theory.

3. Frank C. Zagare & D. Marc Kilgour, "Perfect Deterrence," Cambridge University Press, 2000.

4. *'Direct Deterrence'* essentially describes a (strategic) situation in which two players face off against each other whereas *'Extended Deterrence'* describes a (strategic) situation in which one player (*defender*) is trying to protect its protégé against the other player (*challenger*).

---

defender), and gives due consideration to the *'status quo'* which was much neglected in favor of the cost of conflict. It also provides rationale for minimum deterrence.

Chapter 2 reviews basic concepts in Perfect Deterrence Theory that are required for understanding the later chapters. *'Capability'* and *'credibility'* are two basic components which play critical roles in the theory. Capability is one's ability to hurt one's adversary while credibility is one's willingness to fight rather than capitulate. Some examples are provided in the context of Korea. The Korean War (1950-1953) is an example of *incapable* South being invaded by *capable* North. The Admiral Yi Sun-sin's famous saying, *"Those who seek death shall live. Those who seek life shall die,"* before defeating Japanese fleet at the battle of Myeongnyang in 1597 epitomizes the highest level of credibility. By contrast, the sinking of Cheonan, a South Korean navy corvette, and the subsequent shelling on the island of Yeonpyeong in 2010 could be partly attributed to low credibility. Finally, the definitions of type *'Soft'* and type *'Hard'* are given and the incomplete information game is explained using these types.

Chapter 3 introduces the *Generalized Mutual Deterrence Game*. Our primary interest is in the *Sure-Thing Deterrence Equilibrium* among all solutions found by Zagare & Kilgour.[5] Its existence conditions contain North Korea's utilities for *the 'Status Quo* (N_{SQ}),*' 'Conflict* (N_{DD_+})*'* and *'North Korea wins* (N_{DC}).*'* We have examined each utility variable and suggested ways to strengthen deterrence (i.e., robustly fulfilling the existence conditions). We caution against, for example, blindly increasing the value of N_{SQ} in the absence of genuine progress in North Korea's denuclearization. The only sure way to bring about deterrence is to decrease N_{DD_+}. This necessitates a show of force including, for example, a display of new high-tech F35A Joint Strike Fighters and the establishment of *'Decapitation Unit.'* Also, South Korea's credibility (as perceived by North Korea), P_S, is also a critical factor in deterrence calculations. Some noticeable failures in the South Korean military in recent times, which contribute to the lowering of P_S, are noted. We have reviewed Zagare & Kilgour's other equilibria (i.e., the *Attack Equilibria* and the *Bluff Equilibrium*).[5] Deterrence always fails in the *Attack Equilibria*. In the case of the

---

5. See Appendix 4 of 3 for detailed derivations of these equilibria which are grouped into Class 1, Class 2A, Class 2B and Class 3. Class 1 includes the *Sure-Thing Deterrence Equilibrium* which preserves the status quo. Besides this, it has the *Separating Equilibrium* and the *Hybrid Equilibrium*. See also Table 1 in Appendix at the end of this report. Note that Class 2A and Class 2B correspond to Class 2N and Class 2S respectively.

Bluff Equilibrium, with both Koreas lacking credibility, particular attention should be drawn to the inherent danger of an unwanted war/conflict arising out of mutual bluffing and misjudgments of each other. A scenario in which North Korea's credibility, P_N, relating to its use of nuclear weapons could be genuinely high is also mentioned.

Chapter 4 introduces the *Unilateral Deterrence Game* in which South Korea, a defender, continues to preserve peace and stability while North Korea, a challenger, seeks to defeat South Korea. This is an asymmetric game. Once again, our main focus is on deterrence equilibrium and, in particular, on the *Certain Deterrence Equilibrium* among all solutions found by Zagare & Kilgour.[6] As a way of increasing North Korea's utility for the status quo, N_{SQ}, South Korea must avoid funding, in effect, Kim Jong-un and his trusted super-elites' luxurious lifestyles in Pyongyang. Money can be easily funneled to support North Korea's nuclear weapons and ballistic missile programs, which undermines deterrence. N_{SQ} is not to be over-trusted as it can change overnight. Instead, our efforts should be focused on decreasing N_{DD+}. Strengthening both active and passive defense lowers N_{DD+} and, in particular, the U.S. THAAD anti-missile defense system must be upgraded and integrated with the Patriot systems in operation without delay. Despite recent disputes between South Korea and Japan, both countries share the same core values—freedom, democracy, respect for human rights and the rule of law. They must work closely to face up to North Korea's challenges and beyond. As for increasing P_S, South Korean military's new rules of engagement, empowering frontline commanders to order retaliation swiftly, serves as a good example of strengthening deterrence.

Among other solutions (the *Attack Equilibrium*, the *Bluff Equilibrium* and the *Separating Equilibrium*), the *Bluff Equilibrium* is particularly noted for its inherent advantage for the defender (i.e., South Korea). The incident of August 20, 2015, in which North Korea exchanged fire with South Korea over loudspeaker, is viewed from this perspective.

In the concluding chapter, the importance of maintaining high credibility, P_S, is stressed again as many deterrence failures can be traced to low P_S. In Perfect Deterrence Theory, being capable is necessary (but not sufficient) for deterrence to hold. South Korea is not capable of competing against North Korea in the nuclear arena. As the country totally depends on U.S. extended (nuclear) deterrence to defend against the growing North Korean nuclear threat, the U.S. needs to provide concrete assurance to its allies in this region.

---

6. See Appendix 5 of 3 for detailed derivations of these equilibria which are grouped into Certain Deterrence & Steadfast Deterrence, Separating Equilibrium, Bluff Equilibrium and Attack Equilibrium. See also Table 2 in Appendix at the end of this report.

1. Overview of Perfect Deterrence Theory

Perfect Deterrence Theory has been developed as an alternative to classical deterrence theory. We have become interested in the theory as it seems to be aesthetically more appealing than classical deterrence theory. It provides another way of looking at deterrence but, this time, without its internal structure plagued by logical inconsistency. As Frank Zagare & Kilgour, the creators of Perfect Deterrence Theory, explain in their co-authored book 'Perfect Deterrence,'[3] classical deterrence theory suffers from what is known as the *paradox of mutual deterrence*. Classical deterrence theory hinges on the fact that the status quo is reinforced as the cost of conflict becomes higher. However, logic dictates that the status quo is not a rational choice when both States (A & B) prefer capitulation to conflict when challenged. The status quo is not a Nash equilibrium[7] in the standard 2 x 2 game of Chicken in which there are four outcomes (i.e., A wins, B wins, status quo and conflict). The game of Chicken was a tool favored by classical deterrence theorists as it captured essential aspects of deterrence during the age of nuclear confrontation between the United States and the Soviet Union, as it was then. In this game, conflict is understandably assumed to be the worst choice and, hence, the contradiction ensues. The proponents of classical deterrence theory are yet to provide a satisfactory resolution to this logical inconsistency although various attempts have been made.[8] Perfect Deterrence Theory was also developed to overcome the empirical deficiencies of classical deterrence theory. Zagare & Kilgour point out that the 'power imbalance' is a poor indicator for measuring the likelihood of war as there have been many instances of war fought between two states with roughly equal power, World War II between Britain and Germany for one. There is little empirical support for associating the power imbalance with the likelihood of war. It is a prerequisite that a good theory must stand up reasonably well to empirical evidence. Overall, Perfect Deterrence Theory's predictions are more in agreement with empirical findings.

Here are some distinguishing features of Perfect Deterrence Theory. As explained, in

---

7. In Game Theory, a Nash Equilibrium is a collection of strategies, one for each player in a game, where there is no incentive for any player to switch his/her strategy. The game is at equilibrium as all the players in the game are satisfied with their choices at the same time.

8. Zagare & Kilgour explain why the paradox of mutual deterrence still remains despite these attempts. For details, see Chapter 2 of 3.

classical deterrence theory, the cost of conflict plays a *'central'* role in deterring an adversary. Hence, it makes a perfect sense for a state to stock up on more powerful weapons to deter the adversary. These weapons will unleash more destructive power, thereby significantly raising the cost of conflict which, in turn, reduces the possibility of conflict.[9] Any adversary would think twice before challenging to upset the status quo. However, as Zagare & Kilgour point out, this model of deterrence is deficient in the sense that there is no maximum limit on the cost of conflict (i.e., that is, a state should just keep on accumulating stockpiles of ever more destructive weapons as long as it can monetarily afford them). There is no built-in mechanism within the model that provides the maximum limit. Such a limit can be shown to exist inherently in some models of Perfect Deterrence Theory, beyond which further strengthening of deterrence becomes totally redundant (i.e., making a case for minimum deterrence). This is a desirable feature to have in the model of deterrence. The theory also gives due consideration to the importance of maintaining the '*status quo*' which is somehow neglected in the past. The status quo did not receive much attention, as the focus was heavily on the cost of conflict. The emphasis was on the punishment side. Moreover, States A and B no longer have to be undifferentiated in the theory. If needed, one state can be specifically designated as challenger determined to upset the status quo while the other state, as defender, is determined to keep the status quo. This will better approximate a real-world situation.

'Capability' and 'credibility' are two critical variables in the analysis of a deterrence situation. In Perfect Deterrence Theory, capability is defined as one's ability to hurt the other (i.e., adversary) while credibility is defined in terms of one's willingness to fight rather than capitulate. These variables together with other utility variables (e.g., a utility for A wins, B wins, Status Quo and Conflict) are used to determine rational choices and the conditions under which those choices exist. These conditions usually take the form of mathematical inequalities and many insights are gained through analyzing these inequalities. At some decision points in the game (or tree),[10] it may be necessary that States A and B update their belief about the other's determination to fight rather than capitulate based on conditional probability.[11]

---

9. Of course, the paradox is still lurking in the background.

10. This refers to some nodes in extensive-form representation of the game.

11. In statistics, the conditional probability is an update of probability in light of new evidence.

Zagare & Kilgour provide all mathematical details in the appendix section of their book.[3, 12] Their work is firmly founded on logic and, hence, model predictions are the direct consequences of a rational decision-making process. In practice, rational choices are nothing but Nash equilibria. Determining these choices simply amounts to finding Nash equilibria. There can be more than one Nash equilibrium under a given condition, which is perhaps a less desirable feature of the theory as it loses predictability. It is simply not possible to know which course of action will be taken in a situation with multiple Nash equilibria.[13] Hence, multiple equilibria may be able to explain the success or failure of a deterrence situation that has occurred in the past. One can only hypothesize which one might have been in play. In the theory, only a Nash equilibrium that fulfills Selten's '*Perfectness*' condition (i.e., being subgame-perfect) is thought to be a *'rational'* choice. This condition is used to eliminate all choices which have at least one incredible (a technical term as explained later) decision made along the decision path. Only subgame-perfect Nash equilibria are of true interest as they are free of irrationalities. Accordingly, Zagare & Kilgour have named their deterrence theory with this game theory term '*Perfect*' in front. It does not mean *'Perfect Deterrence'* in the ordinary sense of *'perfect,'* as such deterrence has never existed and never will.

Perfect Deterrence Theory has been specifically chosen to be the basis of our study for these reasons. We sought to draw out valuable implications for South Korea based on Zagare & Kilgour's findings.[3] Basically, what can this deterrence theory tell us about strengthening the security of South Korea and continuing to preserve peace and stability in the face of North Korea's growing nuclear threat? In the next chapter, basic concepts in Perfect Deterrence Theory are reviewed.

---

12. For those with technical background, the appendix section contains the mathematical derivations of the solutions. In places, their proofs are rather compact.

13. For certain Nash equilibria, an improbable sequence of decisions is required. It is thought that these equilibria won't find their counterparts in real-world situations. Zagare & Kilgour also note that multiple Nash equilibria can be looked upon positively as a state has freedom to steer towards one outcome over the others.

2. Basic concepts in Perfect Deterrence Theory [14]

The tools behind Perfect Deterrence Theory basically consist of utilities (i.e., payoff), strategic variables and a probability measuring credibility of one's opponent to retaliate. The utilities are specified in order of preference (e.g., State A: conflict > capitulation) satisfying the completeness and transitivity conditions of game theory. Strategic variables are probabilities connected to the States' choices, contingent on the type (i.e., Hard or Soft as explained in Section 2.2). There is also a probability which measures the perceived credibility of one's opponent to retaliate. This can be updated when required before one decides on the next move. A subgame-perfectness ensures that any threat to retaliate, which seems less than convincing to oneself or one's opponent, will never be included in a solution. Zagare & Kilgour have found and classified all perfect Bayesian equilibria based on their characteristics.[15] Here, we are primarily interested in the solution(s) in which deterrence holds so that the status quo is maintained in the end. A game can be played either with complete[16] or incomplete information. In a game of incomplete information, one does not have all the information concerning the utilities of one's opponent (i.e., is one's opponent Hard or Soft?).[17, 18] In a real world situation, it is more likely that States A and B will encounter this type of game as it is difficult to be fully aware of the mind (or type) of one's opponent and, hence, what action or reaction one's opponent will take.

2.1 Capability and Credibility

Capability and credibility play critical roles in success or failure of deterrence. From our everyday use of language, these are quite familiar concepts. In Perfect Deterrence Theory, the exact definitions of these concepts are given by imposing certain conditions.

---

14. The mathematical notation and symbols are those in 3.

15. These equilibria are also subgame-perfect.

16. In finite games of complete information, one can use backward induction to identify a subgame-perfect Nash equilibrium (or equilibria).

17. This is not to be confused with imperfect information. In a game of imperfect information, there exists at least one non-singleton information set. A player with this information set does not know all the moves leading up to that point (or node).

18. The approach based on the Bayes theorem was originally developed by John Harsanyi in the 1960s.

Zagare & Kilgour (Perfect Deterrence, 2000, p.82) stated that *"… we define threat capability in terms of the relationship between the outcome that results when no action is taken (i.e., the status quo) and the outcome that results when an untoward action is taken and the threat is executed (i.e., conflict). A threat will be said to be capable, then, if and only if the threatened player prefers the status quo to conflict; when this relationship is reversed, the threat will be said to lack capability."*

- State A is said to be *'capable'* if State B prefers the status quo to conflict
  ⇒ State B: *Status Quo* $>_B$ *Conflict* (i.e., State A is capable)[19]
  ⇒ State B: *Conflict* $>_B$ *Status Quo* (i.e., State A is not capable)

The same can be said of B if we swap A and B above. It turns out that capability is an *'absolutely necessary condition'* for deterrence success, but this alone does not guarantee a successful deterrence.[20] Although some theorists may hold the view that it should suffice, it is not true in Perfect Deterrence Theory. However, a lack of capability does not necessarily lead to deterrence failure. The Korean War (1950-1953) is one good example in which capable North Korea invaded not capable South Korea. South Koreans only found out about the full invasion in the early hours of Sunday, 25 June 1950. Even if they had known about North Korea's intentions earlier, there was perhaps nothing much they could have done to deter North Korea from executing its invasion plan. North Korea preferred conflict to the status quo. It was very clear that South Korea lacked military hardware and manpower necessary to defend itself. It simply lacked capability, a necessary but not sufficient condition required for deterrence success. Therefore, it only becomes meaningful to speak of devising a deterrence strategy when this condition holds for both A and B (i.e., both capable). For this reason, only deterrence relationships in which both sides can make capable retaliatory threats are studied in 3. This brings up the question: which of North and South Korea has capable retaliatory threat at present?[21] Can one truly say both sides can make capable retaliatory threat? One is a self-declared nuclear weapon state and the other, a non-nuclear weapon state. There will be more to say on this matter later on.

---

19. "X $>_B$ Y" is interpreted as "X is preferred to Y by B."
20. Zagare & Kilgour mention that a second-strike capability remains necessary for deterrence success. A capability that can be removed by an enemy first strike cannot be capable.
21. Here we are referring to the military strength of South Korea alone and not the combined military strength of South Korea and the United States.

Credibility is another factor that contributes to the success or failure of deterrence. A threat is useful only to the extent that it can be believed by one's opponent. For a threat to be credible, it must be rational to carry out that threat. It must be convincing and believable to one's opponent. In game theory terms, this is translated to requiring that Nash equilibria satisfy the perfectness condition. Now, threat credibility is defined (Perfect Deterrence, 2000, p.68) as *"the extent to which a threatener is seen to prefer to execute the threat (should appropriate contingency arise) … we assume that an actor prefers to execute a threat when anticipated worth of doing so exceeds the anticipated worth of failing to do so. Otherwise, the threat is irrational and, hence, incredible."* In Perfect Deterrence Theory, credibility is defined as follows:

- State A is said to be *'credible'* if A prefers *'Conflict'* (retaliation) to *'B wins'* when challenged
  ⇒ State A: *Conflict* $>_A$ *B wins* (i.e., State A is credible)

One of the most important consequences of this theory is that a defending state's threat to retaliate be both '*capable*' and '*credible*' for deterrence to succeed. The defending state must leave no doubt that it is willing to retaliate if challenged by another state. A game is structured such that there is an opportunity to retaliate should the defending state wish to. If the two conditions are fulfilled throughout, then the theory shows that deterrence will hold for certain and the status quo can be maintained. In the famous 2 x 2 Prisoner's Dilemma game[22], the order of preference for State A (or B) is as follows: A (or B) wins, Status Quo, Conflict and B (or A) wins. So, the conditions for capability and credibility hold for both states, and an imperfect information game played under such conditions will engender deterrence although this is never completely guaranteed. Specifically, the status quo and conflict are the only Nash Equilibria that are also perfect and, hence, either choice is rational. However, the status quo is favored by at least one state while the other state remains indifferent.[23] Note that the order of preference is reversed between Conflict and B (or A) wins in the game of Chicken.

---

22. See, for example, Prisoner's Dilemma, Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/prisoner-dilemma/.
23. This is demonstrated in the *Generalized Mutual Deterrence Game* (imperfect) with both players having a credible retaliatory threat (Section 3.2 of 3). See also Zagare & Kilgour's explanations of '*Pareto-superior*' and '*Pareto-optimal.*'

Besides capability, it basically comes down to showing its resolve to defend itself, come what may. Yi Sun-sin, the most respected Korean admiral, said to his generals, *"Those who seek death shall live. Those who seek life shall die"* before defeating an armada of 133 or more Japanese warships with only 13 warships of his own at the battle of Myeongnyang in 1597.[24] His brilliant naval strategy based on local geographical knowledge and sheer determination won through in the end. His meager number of warships against the armada of more than one hundred Japanese warships wouldn't even have fulfilled the necessary capability condition let alone winning the battle. Admiral Yi was determined to fight to the last man in the Joseon Dynasty's hour of need. His famous saying epitomizes the highest level of *'credibility.'*[25] Had Yi's naval strength appeared capable also, deterrence would have stood a better chance of success.

In March 2010, the South Korean warship Cheonan was sunk at night by a torpedo near Baengnyeong Island in the West Sea, killing 46 sailors. The evidence strongly suggests that the warship was attacked by a North Korean midget submarine. The South Korean warship was caught totally off guard during the night with disastrous consequences, and the midget submarine presumably slipped back to North Korean waters unhindered. For the South Korean military, it failed to hunt down perpetrators. Despite all the subsequent measures brought in by the South Korean government to punish North Korea, the most credible and effective response would have been to force the suspected submarine to surface and then seize it or sink it altogether. Unfortunately, there was not even a clear target to retaliate against in the midst of confusion, and it would seem that no effective counter-attack could have been launched in time to hunt down the submarine.

---

24. The battle of Myeongnyang and Yi Sun-sin, http://dh.aks.ac.kr/Korea100/wiki/index.php/The_Battle_of_Myeongnyang_and_Yi_Sun-sin.

25. Paik Sun-yup, a famous Korean War hero and South Korea's first four-star general, is quoted as telling his soldiers during the battle, "There is nowhere to retreat now. … If we are pushed back, the U.S. troops will withdraw. Then the Republic of Korea will be over. I will be at the forefront as your commander. If I retreat in fear, shoot me first." The battle ended in victory and saved the Republic of Korea from being overrun by the communist North. This serves as another good example of high credibility in action. "Paik Sun-yup, Korea's first four-star general, dies at 99," Korea JoongAng Daily, 12th July 2020. https://koreajoongangdaily.joins.com/2020/07/12/national/defense/obituary-Paik-Sunyup-general/20200712182800378.html.

Eight months later, North Korea fired around 170 artillery shells onto the small South Korean island of Yeonpyeong, killing 2 marines and 2 civilians and injuring scores more. The shelling damaged military infrastructure and started widespread fires on the island destroying houses and public buildings. The attack came in two waves catching the South Korean military off guard. It retaliated by firing back 80 artillery shells targeting the North Korean military. However, the South Korean military may have failed to deliver an immediate and '*befitting*' response. Satellite photos released by STRATFOR cast doubt on effectiveness of its counter-attack on a North Korean multiple rocket launcher (MRL) battery contrary to the claims by JCS and NIS.[26] The

Figure 1. A burnt out house in the island of Yeonpyeong after the North Korean Shelling

Source: Yonhap News.

---

26. See the satellite photo showing the *probable aftermath of South Korean counter-attack against North Korean MRL battery.* https://www.stratfor.com/sites/default/files/DigitalGlobe_North_Korean_Attack_11-.23.pdf.

island of Yeonpyeong was not properly equipped to carry out a deadly counter-attack.

With hindsight, the Cheonan sinking coupled with failure to carry out an immediate and effective counter-attack would have eroded South Korea's threat credibility. This failure would undoubtedly have given North Korea strong incentives to mount a more daring attack at a later time as manifested by the shelling of Yeonpyeong. From the perspective of Perfect Deterrence Theory, maintaining threat credibility becomes all the more important.[27]

2.2 Type *Soft*, Type *Hard* and Incomplete Information

State A is said to be *'Hard'* if the credibility condition below holds.

- State A: *Conflict* $>_A$ *B wins* (i.e., State A is credible)

If not, then it is said to be *'Soft'* with an inequality sign reversed as below.

- State A: *Conflict* $_A<$ *B wins* (i.e., State A is incredible)

In incomplete information games, neither state knows for certain whether its opponent is Soft or Hard. Hence, for example, State A can only assign a probability P_B which is defined as the probability of State B being Hard as observed by A based on all information available up to that point - if this applies. Similarly, $(1 - P_B)$ is defined as the probability of State B being Soft. Likewise, the same rule applies to State B and its associated probabilities P_A & $(1 - P_A)$.

In this chapter, '*capability*' and '*credibility*' are defined and explained using the Korean War (1950), the famous battle of Myeongnyang (1597) and the two volatile incidents (2010) that occurred between North and South Korea as illustrative examples. In the next two chapters, we focus on Zagare & Kilgour's models of direct deterrence, specifically, the '*Generalized Mutual Deterrence Game*' and the '*Unilateral Deterrence Game.*'

---

27. In both the Cheonan and Yeonpyeong Island incidents, had South Korean military responded effectively, the ball would then have been firmly in North Korea's court to risk further escalation.

3. Implications of the *Generalized Mutual Deterrence Game*

The *Generalized Mutual Deterrence Game*[28] describes the simplest situation in which two states can simultaneously choose an option between '*cooperate/capitulate*' and '*defect,*' and after initial moves by both, each state has an opportunity to retaliate if the other state has chosen to defect. Figure 2 is an extensive-form representation of the *Generalized Mutual Deterrence Game* between North and South Korea. Unless stated otherwise, South Korea means South Korea alone and does not refer to the combined strength of South Korea and the United States.

Figure 2. Generalized Mutual Deterrence Game[29]
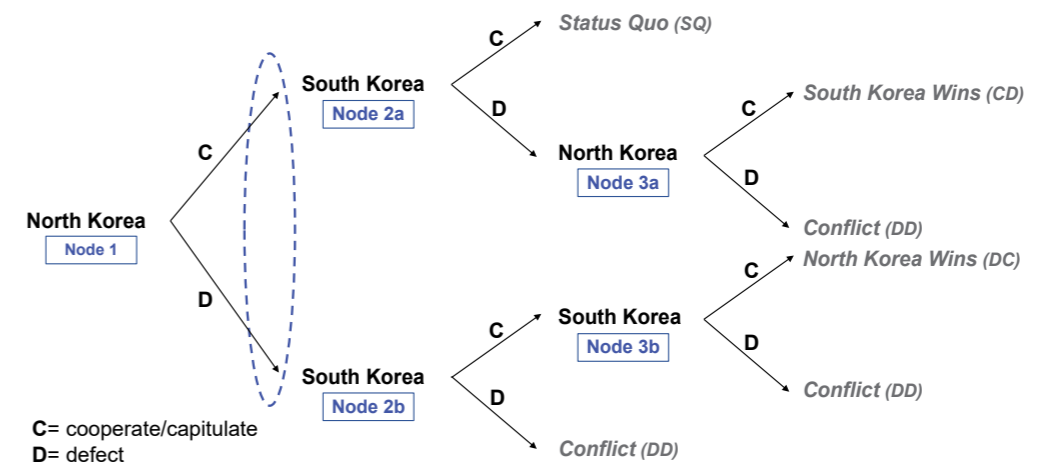
Source: Frank C. Zagare, D. Marc Kilgour, "*Perfect Deterrence*," Cambridge University Press, 2000.

The oval-shaped dotted line represents a non-singleton information set and, hence, it is a game of imperfect information. South Korea has no knowledge of whether North

---

28. See Chapter 4 of 3 for Zagare & Kilgour's detailed exposition of the *Generalized Mutual Deterrence Game.* This chapter and the next chapter are largely based on 3.

29. The figure has been adopted and relabeled from Zagare & Kilgour's original version to suit our needs.

Korea has initially chosen to cooperate or to defect and vice versa. If North Korea has opted to cooperate, then South Korea will find itself at node 2a. Otherwise, it will be at node 2b. Furthermore, if South Korea has initially cooperated but not North Korea, it can either retaliate by choosing defect or capitulate at node 3b. If the former, the result will be conflict, but if the latter, then North Korea wins. North Korea faces a similar situation at 3a. The *Generalized Mutual Deterrence Game* is, as well as being imperfect, an incomplete information game. Hence, each Korea could be either Hard or Soft. It knows about what type it is, but only has probabilistic knowledge of its opponent's type. As shown in Figure 2, there are four outcomes of this game. For South Korea, the order of preference is *South Korea Wins* >$_S$ *Status Quo* >$_S$ [*Conflict*, *North Korea Wins*] and analogously for North Korea, it is *North Korea Wins* >$_N$ *Status Quo* >$_N$ [*Conflict*, *South Korea Wins*]. The preference between the two outcomes in the square bracket depends on the type.

Here, the '*capability*' condition is assumed throughout as it is necessary, though not sufficient, for deterrence success. However, it is not convincing to assume that both Koreas are indeed '*capable*' considering South Korea's military can only compete in the conventional weapons arena while North Korea is a de facto nuclear weapon state. Considering the destructive power of nuclear weapons, it is difficult to imagine how conventional forces could deploy a militarily meaningful defense. One cannot think of a war or a conflict between a nuclear-weapon and a non-nuclear weapon state in which the non-nuclear weapon state has prevailed.[30] Without U.S. extended (nuclear) deterrence, South Korea cannot fulfill the '*capability*' condition on its own.[31] It will be hard put just to maintain the status quo. The only way for its own military to fulfill this condition is to acquire nuclear weapons. This will restore the balance of terror which has tilted heavily in favor of North Korea at the present time. With this fact in mind, let's continue with the *Generalized Mutual Deterrence Game*. Zagare & Kilgour have found and

---

30. In the Vietnam and the Soviet-Afghan Wars, for example, the territories of the United States and the former Soviet Union were never in any danger of being attacked or encroached upon. They were fighting outside their territories to support their allies and friends. In Falklands War, the United Kingdom, a nuclear weapon state, has re-taken the Falkland Islands from Argentina. This was a localized war which took place thousands of miles from mainland United Kingdom. In Persian Gulf Wars, Iraq stood no chance against the United States-led coalition forces.

31. This is an over-simplification as there are many areas in which U.S. support is vital besides nuclear deterrence.

grouped ten perfect Bayesian equilibria into four distinct classes.[5] Our task is to draw out valuable implications for South Korea based on these equilibria. As the connection between deterrence, preferences and threat credibility takes the form of a mathematical equation, it is quite clear how they interact with one another. Zagare & Kilgour have determined that a four-tuple of probabilities [x$_H$, x$_S$; y$_H$, y$_S$] is required to specify a perfect Bayesian equilibrium. The definitions are given in a footnote below.[32] The *Sure-Thing Deterrence Equilibrium* is the name they have given to describe a situation in which neither side (i.e., North and South Korea in our context) wants to upset the status quo. Therefore, this equilibrium is of particular interest to us as deterrence holds and the peace is maintained.

3.1 The *Sure-Thing Deterrence Equilibrium*[33]

The *Sure-Thing Deterrence Equilibrium* is the only equilibrium in which the preservation of the status quo is the logical outcome. Its four-tuple takes the form [0, 0; 0, 0] so that the probability of defection by either Korea, irrespective of its type, is zero. The existence conditions are as follow[34]:

$$\mathrm{P_S} \geq \mathrm{N_2} = \frac{N_{DC} - N_{SQ}}{N_{DC} - N_{DD+}}, \quad \mathrm{P_N} \geq \mathrm{S_2} = \frac{S_{CD} - S_{SQ}}{S_{CD} - S_{DD+}}$$

Here, P$_S$ denotes the probability that South Korea is Hard as perceived by North Korea and analogously, P$_N$ is the probability that North Korea is Hard as perceived by South

---

32. The definitions below are adopted and relabeled from Zagare & Kilgour's original version to suit our needs.

x$_H$ = the probability that North Korea chooses D, given that it is Hard

x$_S$ = the probability that North Korea chooses D, given that it is Soft

y$_H$ = the probability that South Korea chooses D, given that it is Hard

y$_S$ = the probability that South Korea chooses D, given that it is Soft

A four-tuple belonging to Class 1, for instance, is of the form [· , 0; · , 0] where · denotes any positive value between 0 and 1.

33. Zagare & Kilgour's full account of this equilibrium, which forms the basis of this section, can be found in Section 4.3.1.1 of 3.

34. The notations are adopted and relabeled from Zagare & Kilgour's original version to suit our needs.

Korea. Also, N_{DC}, N_{SQ} and N_{DD+} denote the North Korea's utilities for *North Korea Wins*, *Status Quo* and *Conflict* respectively (see Figure 2 above). One can clearly see how P_S, South Korea's credibility to fight rather than capitulate, is mathematically related to these utilities.[35] As this is a symmetric game, a similar relationship mirroring P_S exists for P_N. Let's elaborate on how South Korea can strengthen its deterrence and defense posture based on the existence conditions above so that the status quo is preserved.

Mathematically, the best thing is to make P_S much greater than N_2 (i.e., $P_S \gg N_2$) so that North Korea does not even dare to think about defecting (i.e., *'D')*.[36] The overall value of N_2 needs to be decreased by considering North Korea's utility variables. Our task is to decrease the numerator and to increase the denominator of N_2. P_S also needs increasing.

<N_{SQ}: North Korea's utility for *the Status Quo*>

As North Korea's evaluation of the status quo increases, the numerator of N_2 decreases bringing down the value of N_2. For a given value of P_S, this means that the existence conditions are more likely to be satisfied. The Kaesong Industrial Complex and Mt. Geumgang tourism are two examples in which N_{SQ} could be increased. These projects symbolize the inter-Korean cooperation initiated by Kim Dae-jung's[37] Sunshine policy toward North Korea. His immediate successor, Roh Moo-hyun,[38] picked up the baton and continued the Sunshine policy. It should be borne in mind that North Korea had been running a clandestine nuclear weapons program all along. It had not adhered to the Joint Declaration on the Denuclearization of the Korean Peninsula signed between North and South Korea in January 1992. Both Koreas agreed not *"to test, manufacture, produce, receive, possess, store, deploy or use nuclear weapons,"* and not to *"possess nuclear reprocessing and uranium enrichment facilities."*[39] Not only has this declaration gone down the drain, but subsequent agreements such as the 1994 US-DPRK Agreed Framework

---

35. Note that DD+ denotes *Conflict* when it is Hard. DD- is for Soft.

36. It is assumed that a value of P_S is common knowledge to both North and South Korea, but only South Korea knows its own type.

37. Kim Dae-jung served as President of South Korea from Feb, 1998 to Feb, 2003.

38. Roh Moo-hyun served as President of South Korea from Feb, 2003 to Feb, 2008.

39. Nuclear Threat Initiative, South Korea Overview, https://www.nti.org/learn/countries/south-korea/.

and the 2005 Six-Party Joint Statement have also followed the same path. North Korea has simply violated all its agreements.[40] These are valuable lessons to be learned for South Korean policymakers especially today in their dealings with North Korea. Blindly increasing the value of N_{SQ} may not always prove to be a prudent policy. The Munich Agreement[41] signed on 29th September 1938 testifies to this statement.[42] It shows potential pitfalls of pursuing a policy of appeasement. Great Britain and France allowed German annexation of Sudetenland, part of Czechoslovakia, in exchange for peace. Czechoslovakia was not *'capable'* as evidenced by the fact that the Agreement was only signed by Great Britain, France, Germany and Italy. It sealed the fate of Czechoslovakia. Neville Chamberlain, British Prime Minister at the time, gleefully boasted of *"Peace for Our time"* upon his return from Munich on 30th September 1938. The rest is history. The Munich Agreement was not worth the paper it was written on.

It is necessary for the incumbent South Korean government to implement a *carrot-and-stick* policy appropriately to force North Korea to fully comply with the UN demands. The *carrot-and-stick* policy is fully in line with what Perfect Deterrence Theory advocates. South Korea must make it absolutely clear that without concrete signs of progress on the nuclear front, North Korea should not expect the lifting of economic sanctions and extensive economic cooperation with the South. It must hold that line firm. Continuing with its nuclear weapons development will only worsen and prolong economic hardship for the North Korean people. Kim Jong-un has a clear choice. South Korea must not backslide to one-sided cooperation and reward North Korea when no real progress has been made.[43] This could potentially be interpreted as the policy of a weak state and also of appeasement and send the wrong signal to Kim Jong-un and his military advisers. North Korea may well ratchet up the tension a notch higher on the Korean peninsula to force concessions while proudly menacing with its nuclear weapons.[44] For this reason, South Korea must always maintain military readiness against provocations by the North as witnessed lately by the blowing-up of an inter-Korean liaison office in the

---

40. In early May 2020, North Korea fired several shots toward a South Korean guard post inside the DMZ in violation of the 2018 inter-Korean Comprehensive Military Agreement.

41. Munich Agreement, Encyclopedia Britannica, https://www.britannica.com/event/Munich-Agreement

42. This is undoubtedly the perfect example of appeasement and a warning from history.

43. Even if real progress has been made, one cannot deny the fact that rewarding North Korea still amounts to basically paying off its illegal activities.

44. By a notch higher, we mean going beyond typical missiles launches or rocket firings.

North Korean city of Kaesong in June, 2020. Its overall pattern of warnings, threats and provocations is fairly clear even though the exact manner in which it will carry out a provocation is difficult to predict.

<N_{DD+}: North Korea's utility for *Conflict*>

The main objective of the South Korean military is to deter an attack by North Korea. As mentioned, reducing North Korea's utility for *Conflict*, N_{DD+}, increases the denominator of N_2 In turn, this reduces the overall value of N_2. For a given South Korean credibility, P_S, there is a better chance of satisfying the existence conditions. There is more room for deterrence success. When North Korea's expected cost of conflict increases, N_{DD+} is reduced accordingly. Therefore, the status quo is likely to remain intact provided that initiating a conflict imposes high costs on North Korea. Undoubtedly, no weapon does better than nuclear weapons when it comes to 'imposing high costs on an adversary. In fact, North Korea, an illegal nuclear weapons holder, has done a much better job of reducing South Korea's utility for *Conflict*, S_{DD+}. The United States provides a nuclear

Figure 3. South Korea-U.S. joint military exercises

Source: Yonhap News.

umbrella to plug a significant gap in South Korea's military capability, ensuring its security. Its successful testing of a new B61-12 scalable (low-yield) nuclear bomb with the F-15E Strike Eagle fighter aircraft at the Tonopah Test Range in Nevada will strengthen the credibility of U.S. extended deterrence. This does not make a nuclear war more likely. As a matter of fact, it should be to the contrary.[45, 46]

One way of reducing N_{DD+} conventionally is to set up an elite military unit dedicated to hunt down the leadership of North Korea. It is a hermit kingdom ruled by three successive generations of the Kim dynasty. What would be in it for Kim Jong-un if he weren't *physically* there to enjoy his kingship and to rule his people like his father and grandfather had done? The very survival of his regime would be of utmost importance to him. For this reason, it is highly improbable that he will ever relinquish nuclear weapons. There is no self-enforcing strategy to induce North Korea's denuclearization. Perhaps all our efforts to find such a strategy may prove vain and fruitless in the end. As a way to reduce N_{DD+}, it would make perfect sense to have such a military unit ready for full operation. In fact, Song Young-moo, former South Korean defense minister, announced the establishment of *'Decapitation Unit'* before the end of 2017.[47]

Also, the ROKAF (Republic of Korea Air Force) has taken delivery of new high-tech F35A Joint Strike Fighters which can instill much fear into Kim Jong-un as these fighters can evade detection, and support the *'Decapitation Unit'* on the ground. The ceremony marking the arrival of first two F35As at the 17th Fighter Wing Air Base in the South Korean city of Cheongju was deliberately low key so as not to provoke North Korea. However, it is necessary from time to time to put up a display of state-of the-art military weapons given the need to reduce N_{DD+}. By the same token, North Korea has been developing and test-firing new types of missiles and rockets. So far, the Moon Jae-in administration policies have prominently centered on increasing N_{SQ} and the importance of reducing N_{DD+} (and increasing P_S) has been relegated to the background. There has been much emphasis on N_{SQ} at the expense of N_{DD+} in his dealings with the North. The bottom line is that the factual assessment of North Korea's *'capability'* should

---

45. Nuclear Posture Review, 2018.

46. A low-yield nuclear weapon can overcome the credibility problem.

47. "South Korea plans *'Decapitation Unit'* to try to scare North's leaders," The New York Times, 12th September 2017. https://www.nytimes.com/2017/09/12/world/asia/north-south-korea-decapitation-.html.

form the basis of the Moon administration's future strategic policies and not North Korea's goodwill gestures or good intentions which can change on Kim Jong-un's whim. If this happens without proper preparation on South Korea's part, South Korea will face a real difficult situation. In fact, we are facing one now. The importance of reducing N_{DD+} irrespective of N_{SQ} cannot be stressed enough. This is what mathematics tells us about strengthening deterrence.

<N_{DC}: North Korea's Utility for *North Korea Wins*>

As N_{DC} increases, N_2 also increases requiring a higher value of P_S to satisfy the inequality condition. This is true as long as $N_{SQ} > N_{DD+}$ and this capability condition is already assumed to hold in the *Generalized Mutual Deterrence Game*. *"Not surprisingly, the probability of deterrence success is inversely proportional to the value the players attach to winning,"* Zagare & Kilgour (Perfect Deterrence, 2000, p.121) say. *"For the DPRK government, the reunification of Korea—on the DPRK's own terms—has been an overriding policy objective since its very inception. The urgent priority accorded to the goal of unconditional unification has been fused into the fundamental documents of both party and state,"* wrote Everstadt.[48] Clearly, South Korea needs to keep its credibility very high to deter North Korea from realizing its main policy objective. It has already failed once in its attempt to unify Korea under communist rule in 1950.

<P_S: South Korea's *Credibility* as perceived by North Korea>

Let's focus on the L.H.S. of the inequality. P_S, South Korea's credibility, measures its determination to fight rather than capitulate. This term is independent of utility variables on the R.H.S. of the inequality. The theory requires a high value of P_S (and P_N) to bring about the *Sure-Thing Deterrence Equilibrium*. In general, it is far more difficult to estimate even its rough value than, say, counting the number of tanks which is an aspect of capability. This does not mean that its estimation is impossible. There are overt clues and signs which can be easily gathered by the enemy, not to mention information obtained from computer hacking. In the military, a high level of discipline

is a critical factor in maintaining combat readiness. However, there have been numerous incidents in recent times which undermine confidence South Koreans have in their military. It raises a serious question whether the South Korean military today is really combat ready.[49] Here are some unfortunate incidents reported in the local media.

In May 2020, North Korea fired several shots at a South Korean guard post inside the DMZ.[40] The South Korean troops tried to return fire with a remotely controlled K-6 heavy machine gun, but it malfunctioned. Only after an interval of half an hour were troops able to respond with two bursts of machine gun fire. The faulty heavy machine gun turned out not to have been properly maintained for 3 months. This unfortunate incident plainly shows the troops were far from ready for combat. If North Korea had kept on firing, the South Korean troops could have found themselves in real trouble.

In June 2019, a small boat carrying four North Koreans docked at the port of Samcheok, 130km south of the Northern Limit Line, on South Korea's east coast. It crossed the maritime border and sailed freely for days in South Korean waters undetected either by the military or the Coast Guard. In fact, it was a civilian resident who first reported the boat to the local police. There were other security breaches at several military bases in 2020 including an air defense outpost in Siheung, a naval base on Jeju Island and the Jinhae Naval Command in Changwon.[50] In all three incidents, the intruders (a drunken man, civic activists and a mentally ill man) were able to roam freely inside these military bases for over an hour before being apprehended.

There are also incidents of a soldier assaulting his superior officer, fighter pilots consuming alcohol while on standby duty, and a group of soldiers including field officers violating the rules in order to go out clubbing during the Covid-19 pandemic. These unfortunate incidents keep occurring still.

Each incident undoubtedly helps to lower P_S, benefitting the enemy. This coupled with lack of live military training exercises with U.S. ally,[51] the South Korean troops will gradually become undisciplined, untrained and not combat-ready. South Korea's credibility

---

48. The End of North Korea, Nicholas Eberstadt, American Enterprise Institute, The New York Times on the Web. https://archive.nytimes.com/www.nytimes.com/books/first/e/eberstadt-korea.html.

49. The United States Forces-Korea's motto is "Be Ready to Fight tonight."
50. Defense minister: No excuse possible for security breaches at military bases, Yonhap News Agency, 17th March 2020. https://en.yna.co.kr/view/AEN20200317001952325.

will, in effect, resemble that of Chicken rather than the Prisoner's Dilemma. As a consequence, it becomes increasingly difficult to maintain the status quo. President Moon Jae-in, a long-time advocate of the liberal Sunshine policy of engagement with North Korea, and his administration have failed to keep up vigilance when there is, in effect, no real progress made toward denuclearization or genuine threat reduction (e.g., dismantling of a nuclear warhead). The premature signing of the 2018 inter-Korean Comprehensive Military Agreement has likely contributed to lax discipline within the South Korean military. After all, South Korea's 2018 Defense White Paper has dropped North Korea as its *'main enemy'* altogether.[52] The importance of maintaining high credibility cannot be overemphasized.

3.2 The *Attack Equilibria* and the *Bluff Equilibrium*[53]

The *Sure-Thing Deterrence Equilibrium* has been our primary focus as deterrence holds in this case. But there are other classes of perfect Bayesian equilibria in which the breakdown of deterrence is either a certainty (the *Attack Equilibria*[54]) or cannot be ruled out (the *Bluff Equilibrium*[55]). Zagare & Kilgour's *Attack Equilibria* consists of three equilibria each for North and South Korea (see Appendix). One four-tuple, for instance, takes the form $[1, 1; \cdot, 0]$ where $\cdot$ takes on a value between 0 and 1. This means that North Korea will always attack irrespective of its type. If South Korea is Soft, it will capitulate. If the equilibrium is of the form $[1, 1; v, 0]$ where $0 < v < 1$, then South Korea, if Hard, will defect probabilistically. They have determined its existence condition as follows[34]:

---

51. "USFK commander calls for major combined drills against evolving North Korean threats," Yonhap News Agency, 1st July 2020. https://en.yna.co.kr/view/AEN20200701008651325?section=national/defense.

52. 2018 Defense White Paper, Ministry of National Defense, Republic of Korea. http://www.mnd.go.kr/user/mndEN/upload/pblictn/PBLICTNEBOOK_201908070153390840.pdf.

53. Zagare & Kilgour's full account of these equilibria, which forms the basis of this section, can be found in Sections 4.3.2 and 4.3.3 of 3.

54. These are grouped into Class 2A and Class 2B which correspond to Class 2N and Class 2S respectively in Appendix.

55. This is grouped into Class 3.

$$P_S \leq N_u = \frac{N_{DC} - N_{SQ}}{N_{DC} - N_{SQ} + N_{CD} - N_{DD-}}$$

The breakdown of deterrence occurs when P_S is low. It occurs because North Korea calculates that South Korea is likely to capitulate rather than fight. South Korea is deterred from putting up a fight as it believes P_N, North Korea's credibility, to be quite high. North Korea strongly believes that it may succeed in achieving its goal by upsetting the status quo (i.e., defect). Here, the *'goal'* is highly context-dependent. For example, it could range from weakening South Korea's stance on sanctions regimes imposed on North Korea, redrawing the maritime boundary line known as the Northern Limit Line (i.e., NLL) to unifying Korea under its own terms. The inequality sign here is reversed compared to the one in the *Sure-Thing Deterrence Equilibrium*. The motivations for maintaining high credibility (i.e., a high P_S) are self-evident. It provides rationale behind why South Korea must adopt, at minimum, a proportional retaliation when provoked. Now, $(N_{DC} - N_{SQ})$ appears both in the numerator and the denominator of N_u. If no other terms exist in the denominator, then they will simply cancel out to give one. However, there is also $(N_{CD} - N_{DD-})$ in the denominator with $N_{CD} > N_{DD-}$, which makes N_u less than one. The greater the difference in $(N_{CD} - N_{DD-})$, the lower N_u becomes, forcing the existence of this *Attack Equilibrium* to be less likely. Mathematically, this can be achieved by either increasing N_{CD} or decreasing N_{DD-} or both. N_{CD} and N_{DD-} respectively denote North Korea's utility for *'South Korea Wins'* and *'Conflict'* it wants to avoid (i.e., *DD-*) due to high cost. As this is a symmetric game, the exact same reasoning also applies to P_N.

Finally, when both P_S and P_N are low, the *Bluff Equilibrium* can arise. Its four-tuple is of the form $[1, \mu; 1, v]$. The probabilities that North and South Korea choose to defect, given that they are both Soft, are μ and v where $0 < \mu$ & $v < 1$. This equilibrium occurs under the existence conditions $(P_S < N_u)$ and $(P_N < S_u)$.[56] At first sight, it's surprising that they would still opt to defect probabilistically even when both have rather low credibility. This happens, as Zagare & Kilgour explain, because each side fears that it may be perceived to be Soft (a chicken) by the other side. To conceal its weakness, it pretends to be credible by defecting sometimes. They (Perfect Deterrence, 2000, p.127)

---

56. Here, S_u is the counterpart to N_u. See Table A4.1 in Appendix 4 of 3 for the exact mathematical expressions. It shows μ as a function of P_N & S_u and v as a function of P_S & N_u. These notations are relabeled from Zagare & Kilgour's original version to suit our needs.

say, *"as a player becomes more likely to be Soft, i.e., as p_A or p_B decreases, the player compensates by increasing the probability of bluffing when Soft, and conversely. By sometimes defecting when Soft, a player conceals its type and avoids exploitation"* and also warns that *"a crisis could be instigated as a purely defensive measure, that is, as a way of fending off pressure for concession."* The possibility that conflict can still occur as a logical consequence even when neither side truly wants it should be borne in mind. It is imperative for South Korea to maintain a high level of credibility (i.e., a high value of P_S) to avoid getting into this situation in the first place. Once again, the critical role credibility plays in Perfect Deterrence Theory cannot be over-emphasized. However, a strong disparity in capability (i.e., 'nuclear' vs 'without-nuclear') undeniably exists between the two Koreas, raising South Korea's credibility, P_S, alone will be insufficient. Zagare & Kilgour note that a nuclear weapon is so destructive that a threat to use such a weapon may actually reduce credibility (i.e., P_N). For North Korea, however, P_N could be genuinely high when Kim Jong-un thinks he has absolutely nothing to lose by using nuclear weapons. He will act like a 'water ghost,' as they say in Korea, trying to drag someone (e.g., South Korea) to a watery grave with him.

In Perfect Deterrence Theory, multiple Nash equilibria can co-exist simultaneously under the same existence conditions. Therefore, there is no guarantee that the *Sure-Thing Deterrence Equilibrium* will be chosen over other equilibria (e.g., the Separating Equilibrium[57]) in the *Generalized Mutual Deterrence Game*. The theory does not provide a definite answer here (see footnote 13). As conflict cannot be ruled out, the South Korean military must be combat-ready at all times and prepare for all eventualities.

In the *Generalized Mutual Deterrence Game*, each side has taken the role of a challenger trying to take a step towards defeating the other side when logic dictates it. In the following chapter, Zagare & Kilgour's *Unilateral Deterrence Game* in which one side takes the role of a challenger and the other side a defender is considered. This is an asymmetric game.

---

57. The *Separating Equilibrium* takes the form [1, 0; 1, 0] so deterrence breaks down when North Korea is Hard.

4. Implications of the *Unilateral Deterrence Game*

The *Unilateral Deterrence Game*[58] describes the situation in which one state, a challenger, seeks to upset the status quo while the other state, a defender, tries to keep it. Unlike the previous game, the two states are no longer undifferentiated, dropping one core assumption of classical deterrence theory. This asymmetric game can be considered to better reflect the situation surrounding the Korean Peninsula today.

Figure 4. Unilateral Deterrence Game[59]
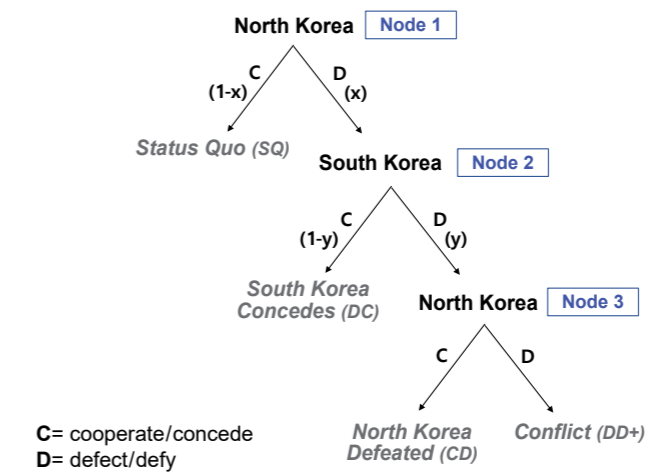
Source: Frank C. Zagare, D. Marc Kilgour, "*Perfect Deterrence*," Cambridge University Press, 2000.

North Korea takes the role of the challenger attempting to unify Korea under communist rule while South Korea takes the role of the defender protecting freedom and democracy (in a region ruled by dictators to a large extent). Figure 4 depicts an extensive-form representation of the *Unilateral Deterrence Game* between North and South Korea. As before, '*South Korea*' refers to South Korea by itself and not the combined strength of South Korea and the United States. At node 1, North Korea can

---

58. See Chapter 5 of 3 for Zagare & Kilgour's detailed exposition of the *Unilateral Deterrence Game.*

59. The figure has been adopted and relabeled from Zagare & Kilgour's original version to suit our needs.

either cooperate or defect to keep the status quo or upset it. If the latter, South Korea can choose either concede or defy at node 2. If defy is chosen, then the game continues and North Korea has to choose between concede and defy in the final node 3. If 'concede,' North Korea's challenge is defeated but if not, then conflict results. There can be many variations of this game. However, Zagare & Kilgour have designed the game relatively simply while retaining the essential features of the strategic situation. It is a game with four outcomes. For South Korea, the order of preference is [*North Korea Defeated*, *Status Quo*] $>_S$ [*Conflict*, *South Korea Concedes*]. There is no restriction on the first pair of outcomes and the order of preference on the second pair of outcomes depends on South Korea's type (e.g., Hard or Soft). In the case of North Korea, the order of preference is *South Korea Concedes* $>_N$ *Status Quo* $>_N$ [*Conflict*, *North Korea Defeated*]. The order of preference on the last pair of outcomes depends on North Korea's type. Therefore, this is an incomplete information game. Once again, the capability condition is assumed to hold by both sides (i.e., they both prefer the status quo to conflict). Zagare & Kilgour have found and grouped five perfect Bayesian equilibria into four distinct categoriesand[6] and have determined that a five-tuple of probabilities [x_H, x_S; y_H, y_S, p] is required to specify a perfect Bayesian equilibrium. However, only the first four-tuple is included for brevity in their discussions, and the specific details of the remaining p are provided in the appendix section of 3. The definitions of these strategic variables are given in a footnote below.[60] In the *Unilateral Deterrence Game*, the '*Certain Deterrence*' and the '*Steadfast Deterrence*' have the status quo as the rational outcome. Only in these equilibria can South Korea continue to maintain deterrence into the future.

4.1 Deterrence Equilibrium (*Certain Deterrence* and *Steadfast Deterrence*)[61]

The *Certain Deterrence Equilibrium* is the most robust deterrence equilibrium with [0,

---

60. The definitions below are adopted and relabeled from Zagare & Kilgour's original version to suit our needs.

x_H = the probability that Hard North Korea will choose D at node 1.

x_S = the probability that Soft North Korea will choose D at node 1.

y_H = the probability that Hard South Korea will choose D at node 2 (this is always 1).

y_S = the probability that Soft South Korea will choose D at node 2.

P = South Korea's conditional probability that North Korea is Hard, given that it chooses D at node 1.

61. Zagare & Kilgour's full account of deterrence equilibrium, which forms the basis of this section, can be found in Sections 5.4.1.1~5.4.1.2 of 3.

0; 1, unrestricted]. Neither type (Hard or Soft) of North Korea tries to upset the status quo. Zagare & Kilgour have given its existence condition as follows[34]:

$$P_S \geq N_t = \frac{N_{DC} - N_{SQ}}{N_{DC} - N_{DD+}}$$

This condition appears quite similar to the one encountered in the *Sure-Thing Deterrence Equilibrium* earlier. Here, P_S is a priori probability that South Korea is Hard as perceived by North Korea.[62] Also, N_{DC}, N_{SQ} and N_{DD+} denote the North Korea's utilities for South Korea Concedes, Status Quo and Conflict respectively (see Figure 4 above). In this game, it is only South Korea trying to keep the peace with no intention of upsetting the status quo. North Korea, on the other hand, always seeks to defeat South Korea at an opportune time. As mentioned, the exact meaning of '*defeat*' here is context-dependent. But, ultimately, it is to turn South Korea into a communist country. For the *Certain Deterrence Equilibrium* to exist, South Korea's credibility, P_S, must take a high value. Ideally, P_S must be much greater than N_t (i.e., $P_S \gg N_t$). Now, much of what we have said about what South Korea '*ought to*' and '*ought not*' do with regard to deterrence in Section 3.1 also applies here. Let's elaborate a little further here.

<N_{SQ}: North Korea's utility for *the Status Quo*>

N_{SQ} basically measures how satisfied North Korea is with the situation it is currently in. A higher value of N_{SQ} indicates higher satisfaction and hence the *Certain Deterrence Equilibrium* becomes more likely. However, Kim Jong-un may one day find himself no longer satisfied with the status quo and decides to forcefully demand or challenge South Korea for more concessions. In other words, N_{SQ} is very fluid and can easily change on Kim Jong-un's whim. South Korea must avoid implementing misguided policies towards North Korea at all costs with regard to increasing N_{SQ}. For example, the wages of Korean workers at the Kaesong Industrial Complex could have funded North Korea's nuclear weapons and ballistic missile programs. They could have been spent on strengthening air defense and anti-aircraft warfare capabilities around Pyongyang or on building submarines. These activities contribute to the raising of N_{DD+}, which weakens deterrence. Besides, the money was likely used to finance Kim Jong-un and his

---

62. Its complement, 1 - P_S, denotes a priori probability that South Korea is Soft.

trusted super-elites' luxurious lifestyles. This may temporarily increase N_{SQ}, but helping North Korea in this way could become a bottomless pit prolonging his dictatorship while ordinary North Koreans suffer in silence without getting any tangible benefits. Such policies could potentially incur heavy costs later on. For instance, we witnessed scenes of South Korean businessmen in their desperate attempts to bring out goods overloaded on the roofs of regular cars as the suspension of the Kaesong Industrial Complex came into force. South Korea's policies on N_{SQ} should be for the betterment of *'ordinary'* North Korean people and must avoid funding, in effect, the privileged few living a life of luxury in Pyongyang and the North Korean military. Its policies could be made conditional upon improvements in North Korea's human rights situation or upon returning South Korean prisoners of war. The South Korean government must stop being tongue-tied about such issues as if all is well when it is not.[63] Medical aid to North Korea on humanitarian grounds could also increase N_{SQ}.

Unlike N_{SQ}, decreasing N_{DD+} can bring about the *Certain Deterrence Equilibrium* in a more concrete and immediate way. N_{DD+} is less prone to changes than N_{SQ} in the sense that it cannot change overnight. Therefore, this should be our ultimate line of defense against North Korea's threats.

<N_{DD+}: North Korea's utility for Conflict>

Not surprisingly, strengthening South Korea's defense capabilities will decrease N_{DD+}. It is quite difficult to challenge a well-defended state whose retaliation can incur high costs. Despite North Korea's rapidly expanding arsenal of ballistic missiles and super-large rockets with increasing accuracy, South Korea's missile defense against these types of weapons is far from adequate. It will become technically much harder to discriminate a ballistic missile with a potential nuclear warhead, the highest priority target, from a super-large rocket with conventional explosives, a low to medium priority target.[64] Also, the high technological edge the South Korean military has held over the North Koreans could be disappearing as they continue to develop and refine their weapons. For instance, the KN-23 missile can follow an irregular trajectory to avoid being hit by an anti-missile missile.

Figure 5. North Korea's KN-23 missile

Source: Yonhap News.

Against this backdrop, it is imperative to ensure the U.S. THAAD anti-missile defense system, the only one of its kind in South Korea, should be upgraded and integrated with the Patriot systems in operation without delay. Once the South Korean Defense Ministry flatly denied the report that there had been an upgrade to the THAAD system in Seongju presumably for fear of offending China after a military convoy was seen entering the U.S. base there.[65] Taking such a stance undermines public confidence, opens the door to more Chinese interference and only bolsters North Korea's military. In July 2019, Russia and China tried to drive a wedge between South Korea and Japan by cleverly flying their military aircraft into the Korea Air Defense Identification Zone (KADIZ). The timing of the incident perfectly coincided with deteriorating political

---

63. In July 2020, there are six South Koreans forcibly detained in North Korea.
64. Of course, the order of priority also depends on the missile's destination.

---

65. "Surprise transport onto THAAD base sparks suspicions over upgrade or additional deployment," The Korea Herald, 1st June 2020. http://www.koreaherald.com/view.php?ud=20200601000637.

and trade relations between South Korea and Japan which hit an all-time low. This incident serves as a stark reminder that any gaping weakness will be ruthlessly exploited.

Once again, South Korea and Japan's collective readiness in facing up to North Korea's challenges and beyond should be strengthened as both countries share the same core values—freedom, democracy, respect for human rights and the rule of law.

Last but not least, South Korea also needs to substantially strengthen its *'passive defense.'* The role of passive defense is to minimize civilian casualties from a barrage of North Korean long-range artillery and rocket attacks while that of *'active defense'* is to minimize damages to critical infrastructures and military bases.[66] Strengthening both *'active defense'* and *'passive defense'* with a robust, multi-tiered missile defense system and providing clear instructions to people in an emergency situation will certainly decrease North Korea's utility for conflict, N_{DD+}.

As for N_{DC}, North Korea's utility for South Korea Concedes, it takes on a similar role to that of N_{DC}, North Korean Wins, in the *Generalized Mutual Deterrence Game.* A high value of P_S is a requisite to deter North Korea's challenges. After the shelling of Yeonpyeong Island, the South Korean military changed its rules of engagement, empowering frontline commanders to order retaliation swiftly and report it up the chain of command afterwards. This simple rule change will have a deterring effect by increasing P_S and minimizing potential South Korean casualties when deterrence breaks down.

Finally, Zagare & Kilgour have found an additional deterrence equilibrium which they called *"Steadfast Deterrence."*[61] This takes the form [0, 0; 1, u] where $0 < u < 1$, and it is subject to the condition $P_S < N_t$ for its existence. They have shown that the *Steadfast Deterrence Equilibrium* is not as robust as the *Certain Deterrence Equilibrium* in the sense that another possibility (i.e., equilibrium) also co-exists under the same condition. Following their line of reasoning, South Korea's credibility, P_S, falls below what is required for the *Certain Deterrence Equilibrium* here and to counteract this shortfall, it needs to defy probabilistically even when it is Soft. North Korea anticipates encountering retaliation even from Soft South Korea once it upsets the status quo. It is not difficult

---

to see that Hard South Korea will always retaliate in the *Unilateral Deterrence Game* (i.e., $y_H = 1$) and not just in this equilibrium. South Korea appears to doubt strongly North Korea's willingness to go down the path to conflict. This very possibility deters North Korea from choosing 'defect' at node 1. Deterrence holds in the end, as the status quo is better than defeat for North Korea. Even intermittent (i.e., u) but commensurate retaliation against North Korean aggression is likely to sow the seeds of doubt about challenging in the minds of North Koreans. To put it another way, the worst possible response would be not to respond/retaliate at all when challenged. From the South Korean perspective, it is much better to have relatively high credibility, P_S, so that North Korea is not mistakenly led to believe that there is a shortfall in South Korea's credibility in the first place.

4.2 The *Attack Equilibrium* and the *Bluff Equilibrium*[67]

The *Attack Equilibrium* is a rational possibility when South Korea's credibility, P_S, is low while that of North Korea, P_N, is high. There is an asymmetry in credibility between the two Koreas. Zagare & Kilgour have determined its four-tuple to be of the form [1, 1; 1, 0] with the following existence conditions.[34]

$$P_S \ < \ N_s = \frac{N_{DC} - N_{SQ}}{N_{DC} - N_{CD}} \ , \quad P_N \ \geq \ S_n = \frac{S_{CD} - S_{DC}}{S_{CD} - S_{DD-}}$$

As x_H and x_S are equal to one, deterrence always breaks down. The fact that $N_{SQ} > N_{CD}$ ensures that N_s is always less than one. As N_{CD} approaches N_{SQ}, N_s becomes close to 1. North Korea's utilities between *the status quo* and *North Korea Concedes* become indistinguishable. As North Korea finds the status quo increasingly difficult to endure, it will take a chance on upsetting the status quo. After all, South Korea's credibility is perceived to be too low to retaliate. Due to North Korea's high credibility, it calculates that South Korea will concede. The blowing-up of an inter-Korean liaison office in the North Korean city of Kaesong could be viewed in this way. This was a blatant attack on South Korean property for which it must be held accountable. Again, this clearly demonstrates the need to maintain a high level of credibility so that $P_S < N_s$ never

---

66. Uzi Rubin and Chong Woo Kim, "Israel's Missile Defense and Some Implications for South Korea," Asan Issue Brief, 16th November 2017.

67. Zagare & Kilgour's full account of these equilibria, which forms the basis of this section, can be found in Sections 5.4.2.2~5.4.2.3 of 3.

comes to pass and ideally $P_S \gg N_s$. It is imperative that North Korea is never tempted to embark on a dangerous adventure in desperate times. The North Korean military must be firmly convinced that South Korea prefers Conflict to South Korea Concedes (i.e., type Hard). North Korea faces increasingly difficult times ahead as international sanctions gradually bite. The current situation, as captured by N_{SQ}, is entirely of its own making and it has only itself to blame. It knows quite well that the only way out of hardship is through implementing FFVD (*'Final, Fully Verified Denuclearization'*).

South Korea cannot meet the capability condition in the nuclear arena. This weakness has to be fully addressed by the USFK (United States Forces Korea). In view of this, the Moon Jae-in government's policy towards the dismantling of South Korea's once thriving civil nuclear industry which has taken several decades to build up with huge efforts seem short-sighted and imprudent. Aside from providing a '*dependable*' carbon-free energy with export potential, its retention also has strategic implications in the face of an existential threat. It makes very little sense to forgo all the benefits that come along with it.

The *Bluff Equilibrium* occurs when both North and South Korea are perceived to have relatively low credibility. Zagare & Kilgour's four-tuple takes the form $[1, v; 1, u]$ where $0 < v, u < 1$ with the existence condition, $P_N < S_n$. Their explanations are as follows.[68] Soft North Korea tries to exploit potential gains by defecting with probability v. After all, the chances are that South Korea is likely to concede as it is perceived to be low in credibility. Likewise, Soft South Korea is also likely to defy with probability u. Both need to bluff as though they are Hard to counteract their low credibility. There is an inverse relationship between u and P_S.[69] Once South Korea has defied at node 2, North Korea is then faced with two stark choices, North Korea Defeated and Conflict. As P_N is relatively low, its rational choice would be North Korea Defeated. Here, it is highly improbable that Hard North Korea has initiated the chain of events. However, this doesn't mean that this possibility can be completely ruled out. Conflict will inevitably result if this turns out to be true. The *Unilateral Deterrence Game* provides an inherent advantage for South Korea (i.e., defender). Zagare & Kilgour (Perfect Deterrence, 2000, p.156) explain that *"But if Defender resists, a rational Soft Challenger will back*

---

68. It is rewritten in the context of North-South Korea.

69. $y_S = u = (N_s - P_S) / (1 - P_S)$, see footnote 6.

down, so that Defender will win. In neither case, though, will a conflict result unless Challenger is prepared for it. Thus, given the postulated sequence of choices, Challenger will not necessarily win, even if it defects initially."

In August 2015, North Korea fired a shell at a South Korean loudspeaker which was broadcasting anti-Pyongyang propaganda. Several more shells landed in South Korean territory near the border on the same day. South Korea retaliated by firing back dozens of artillery rounds at North Korea. In turn, North Korea threatened to start military action in 48 hours unless South Korea stops the propaganda broadcasts. The South Korean military was put on highest readiness to respond to any North Korean military action. However, North Korea's threatened military action '*never*' came after 48 hours. The tensions were eventually diffused by holding high-level talks at the border village of Panmunjom. As this incident clearly illustrates, North Korea's initial choice of defect[70] could lead to a situation in which North Korea finds itself boxed into a corner and forced to choose between North Korea Defeated and Conflict. Neither choice is preferable to the status quo. Conflict could lead to the end of Kim Jong-un's regime while North Korea Defeated would be a humiliating climb-down.[71] Of course, there will be a face-saving measure for North Korea. Therefore, it is a rational choice not to defect in the first place given the setup shown in Figure 4. North Korea must think that it faces a reasonable chance of facing retaliation even from Soft South Korea. This provides South Korea with a clear rationale for holding its ground during times of high tension.

Finally, the *Separating Equilibrium*[72] is possible when South Korea's credibility, P_S, falls short of the threshold required for the *Certain Deterrence Equilibrium*, but it is still above what is required for the *Attack Equilibrium* and the *Bluff Equilibrium* (i.e., $N_s \leq P_S \leq N_t$). Zagare & Kilgour have determined its four-tuple to be $[1, 0; 1, 0]$. It shows that the status quo is maintained only when North Korea is Soft. When it is Hard, it will always defect with the final outcome determined by South Korea's type. If South Korea also turns out to be Hard, then conflict will occur. Otherwise, South Korea will concede. In this equilibrium, each side seeks to influence the other side to be Soft for

---

70. It is committed to start military action in 48 hours.

71. Nothing came out of its threat to start military action after 48 hours.

72. See Sections 5.4.2.1 of 3 for Zagare & Kilgour's full account of this equilibrium.

its benefit. Recently, Kim Yo-jong, the only sister of Kim Jong-un and once his envoy to the 2018 Winter Olympics in South Korea, has progressively taken up a leading role in attacking South Korea and issuing a number of threats. Perhaps these developments could be viewed as North Korea's latest attempts at forcing South Korea to adopt policies favorable to its interests. These policies, advocated by Soft South Korea, include the easing of sanctions imposed against North Korea. North Korea has accused South Korea of a lack of progress in inter-Korean relations. Any effort to ease sanctions on North Korea at this stage would play into the hands of Kim Jong-un and his sister. Once again, South Korea must hold its ground and not waver if it really wants to keep the ever-diminishing prospect of North Korea's denuclearization alive. South Korean policymakers must make it absolutely clear that inter-Korean relations cannot move an inch forward unless there is *'genuine'* progress towards the denuclearization of North Korea (i.e., FFVD).[73]

---

73. In Snyder and Diesing's book "*Conflict Among Nations*," the authors talk about various ways in which a message sent by A to B can be affected during transmission and interpretation. In the end, it is quite possible that the message sent by A need not necessarily be the message received by B. South Korea must not give a false impression/promise as this will only exacerbate the situation later on.

5. Conclusion

Perfect Deterrence Theory has been developed by Zagare & Kilgour as an alternative theory to classical deterrence theory. Their deterrence theory has several desirable features which are lacking in classical deterrence theory. One such feature is to exclude solutions with incredible threats. Furthermore, the theory's predictions are more in agreement with empirical findings and, therefore, it has been chosen to be the basis of our study. In *Direct Deterrence*, Zagare & Kilgour have found solutions (i.e., *'Perfect Bayesian equilibria'*) to the *Generalized Mutual Deterrence Game* and the *Unilateral Deterrence Game* and classified them according to their characteristics. We sought to draw out valuable implications for South Korea by reviewing these solutions. Our primary interest is in the nature of *'Deterrence Equilibrium'* in which the status quo is robustly maintained (i.e., $x_H = x_S = 0$). The connection between deterrence, preferences and threat credibility takes the form of a mathematical equation and, hence, it is quite clear how they interact with one another. *'Capability'* and *'credibility'* are two basic components of Perfect Deterrence Theory. As a non-nuclear-weapon state, South Korea by itself cannot fulfill the *'capability'* condition which is necessary but not sufficient for deterrence success. As the country totally depends on U.S. extended (nuclear) deterrence to defend against the growing North Korean nuclear threat, the U.S. must do more to assure its allies in this region. Washington's continuous downplaying of short-range North Korean missile launches does little to assure the South Korean public of U.S. security commitment[74] and could, in fact, encourage more launches helping North Korea perfect its ballistic missile technology. The theory also requires South Korea's credibility to be quite high to maintain deterrence and reaffirms the importance of having strong determination to retaliate/fight (i.e., type Hard) when challenged. Some examples of failures (and the famous Admiral Yi's success) both past and present with regard to capability and credibility are given in the context of Korea. Each utility variable in a *perfect Bayesian equilibrium* is assessed in the context of North-South Korea and any significant weaknesses relating to this variable are identified. Some suggestions are made to strengthen deterrence. Perfect Deterrence Theory (and its solutions) have basically provided a framework for exploring South Korea's deterrence against North Korea. It

---

74. In fact, there were hardly any messages condemning these missiles launches from the Blue House, the official residence of the President of South Korea, either.

helps to weigh up whether a certain policy is likely to strengthen or weaken deterrence.

At a news briefing on February 12, 2002, a former U.S. Secretary of Defense Donald Rumsfeld gave the following response regarding the limitations of intelligence reports:

"There are known knowns. There are things we know we know. We also know there are known unknowns. That is to say, we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know."[75]

In a documentary film[76] about his political career, he completed this list by saying that a major purpose of Department of Defense is to evaluate *"unknown knowns,"* or *"the things you think you know, that it turns out you did not,"* to anticipate hostile actions before they take place.[77] This author thinks one can perhaps add its complement to this same group, in other words, *'the things you should definitely know (as they actually happened), but you have either forgotten or failed to learn (from the incidents).'* This category is worse than *"unknown unknowns"* and, without a doubt, the worst of all things.

Our failure to learn from our past failures/mistakes will inevitably lead to more failures, and one day we may not have the luxury of failing once more.

---

75. "Rumsfeld's Wisdom, Where the known meets the unknown is where science begins," Michael Shermer, Scientific American, 1st September 2005. https://www.scientificamerican.com/article/rumsfelds-wisdom/.

76. "The Unknown Knowns: The Life and Times of Donald Rumsfeld," a 2013 American documentary directed by Academy Award winning documentarian and filmmaker Errol Morris.

77. "The Unknown Known," Wikipedia, https://en.wikipedia.org/wiki/The_Unknown_Known.

Appendix

Table 1. *Perfect Bayesian equilibria and existence conditions for Generalized Mutual Deterrence Game with incomplete information*[78]

Class	Equilibrium		Strategic variables				Existence conditions		
			North Korea		South Korea				
			x_H	x_S	y_H	y_S	on P_N		on P_S
1	Sure-Thing	STDE	0	0	0	0	$\geq S_2$	and	$\geq N_2$
	Separating	SE	1	0	1	0	$\geq S_u$	and	$\geq N_u$
	Hybrid	HE	u	0	v	0	$\geq S_2$	and	$\geq N_2$
2N	Attack$_{1N}$	AE$_{1N}$	1	1	0	0			$\leq N_1$
	Attack$_{2N}$	AE$_{2N}$	1	1	v	0			$\leq N_u$
	Attack$_{3N}$	AE$_{3N}$	1	1	1	0			$\leq N_1$
2S	Attack$_{1S}$	AE$_{1S}$	0	0	1	1	$\leq S_1$		
	Attack$_{2S}$	AE$_{2S}$	u	0	1	1	$\leq S_u$		
	Attack$_{3S}$	AE$_{3S}$	1	0	1	1	$\leq S_u$		
3	Bluff	BE	1	u	1	v	$\leq S_u$	and	$\leq N_u$

Source: Frank C. Zagare, D. Marc Kilgour, "*Perfect Deterrence*," Cambridge University Press, 2000.

---

78. Tables 1 and 2 are relabeled from Zagare & Kilgour's original version to suit our needs. In Table 1, the terms in P_N are exact counterparts of P_S as the game is symmetric.

Table 2. *Perfect Bayesian equilibria and existence conditions for Unilateral Deterrence Game with incomplete information*

Equilibrium	Strategic variables				Existence conditions
	North Korea		South Korea		
	x_H	x_S	y_H	y_S	
Certain Deterrence	0	0	1	unrestricted	$P_S \geq N_t$
Steadfast Deterrence	0	0	1	u	$P_S < N_t$
Separating Equilibrium	1	0	1	0	$N_s \leq P_S \leq N_t$
Bluff Equilibrium	1	v	1	u	$P_S < N_s$ and $P_N < S_n$
Attack Equilibrium	1	1	1	0	$P_S < N_s$ and $P_N \geq S_n$

Source: Frank C. Zagare, D. Marc Kilgour, "*Perfect Deterrence*," Cambridge University Press, 2000.

A S A N
R E P O R T

Implications of Perfect Deterrence Theory for South Korea

Kim Chong Woo

THE ASAN INSTITUTE
for POLICY STUDIES