The Secret Failure of Nuclear Containment

Word Count: 13841

September 1, 2019

Abstract

I develop a theory of nuclear bargaining where a great power can use both containment and war to thwart nuclear aspirants. Consistent with existing literature, I show that containment deters nuclear aspirants from starting nuclear research when research is expensive and both states' preferences are aligned. Inconsistent with the existing literature, I show that containment can also undermine the great power's credible threat of preventive war and cause deterrence to fail. When containment is sufficiently cheap relative to the cost of war, great powers choose to contain a nuclear aspirant. This hinders, but does not stop, the aspirant's nuclear program. Nuclear aspirants that care intensely about shifting the status quo still invest in their nuclear program knowing that they will face containment. These aspirants would have been deterred from starting nuclear research programs by the credible threat of preventive war. However, they are willing to start nuclear research if the worst they face is containment. The perverse effects of containment arise under the conditions we want deterrence to hold the most: when aspirants are highly industrialized and hold policy preferences that diverge a lot from the great power. An extension shows that the perverse effects are worse in dynamic settings where the risk of proliferation increases over time. Containment delays the chance of proliferation in the short-term but raises the long-term probability that the aspirant will discover nuclear weapons.

1 Introduction

How do nuclear powers stop non-nuclear powers from developing nuclear weapons? One option is preventive war—military actions designed to deny nuclear aspirants the option to pursue nuclear weapons (Bas and Coe, 2016; Reiter, 2005; Debs and Monteiro, 2014). The advantage of preventive war is that it is effective. Removing those that want to build nuclear weapons eliminates the risk of proliferation. But when nuclear aspirants command a strong military, or are protected by powerful allies, the cost of military intervention can be too large to bare (Monteiro and Debs, 2014; Volpe, 2017). Fortunately, even when preventive war is too costly to be credible, powerful states can rely on containment—policies that raise the cost and difficulty for nuclear aspirants to successfully develop weapons (Paul, Morgan, and Wirtz, 2009; Spaniel and Smith, 2015). For example, the United States has sanctioned nuclear aspirants to hinder their nuclear weapons research.

It is widely assumed that these two forms of competition are complimentary (Feaver and Niou, 1996; Jo and Gartzke, 2007; Gartzke and Kroenig, 2014; Jackson, 2009). Powerful states turn to containment and war under different conditions to thwart nuclear aspirants. Better yet, the threat of containment and war have complimentary deterrent effects. Aspirants that believe they will be toppled for starting a nuclear research program do not start one (Debs and Monteiro, 2014). But even when the threat of war is too costly to be credible, the threat of containment can raise the cost of a successful nuclear program by enough to deter many would-be aspirants (Hersman and Peters, 2006; Baldwin, 2000; Miller, 2014). Although deterrence is difficult to observe, clever statistical work suggests that containment has deterred would-be aspirants from starting nuclear programs, demonstrating that it produces "secret success (Miller, 2014)."

In this paper, I show that the threat of containment also has a dark-side: containment undermines the credible threat of war causing deterrence to fail. If containment was not an option the great power's threat of war would deter nuclear aspirants from starting nuclear weapons research programs. But because containment is an option, and powerful states prefer it to war, aspirants pursue nuclear weapons research knowing they will face containment.

To flesh out the logic I develop a model of nuclear bargaining between a great power and a nuclear aspirant. The nuclear aspirant starts without nuclear weapons and then chooses whether or not to invest in nuclear weapons research to acquire them (at a cost). The great power chooses between making peaceful offers, hoping that the aspirant's nuclear research fails, or one of two forms of competition: major war and nuclear containment.

My logic follows from combining three previously studied effects of containment and war in the context of bargaining in the shadow of power and weapons development. Debs and Monteiro (2014) study how the threat of war can deter rising powers from military investments that produce delayed shifts in the balance of power. The threat of war prevents an aspirant from realizing the benefits of their military investment because the aspirant anticipates their military investment will trigger war before it can realize the benefits. Like Debs and Monteiro (2014), I assume nuclear weapons research produces a delayed shift in power if the great power does not respond with preventive war. Unlike Debs and Monteiro (2014), and similar to Bas and Coe (2016), I assume that nuclear weapons research only produces a shift in power with some probability. I make this assumptions because developing the technologies required to build nuclear weapons is a difficult and uncertain process. Neither the aspirant nor the great power knows when or whether the aspirant's research will succeed (Smith and Spaniel, 2018). However, when there is a high probability that nuclear research will produce nuclear weapons, great powers can credibly threaten war (Bas and Coe, 2016). When the threat of preventive war is credible, it deters aspirants from starting nuclear programs in the first place (Debs and Monteiro, 2014).

The threat of containment can also deter nuclear aspirants from starting nuclear research (Hersman and Peters, 2006). Unlike war, containment does not deny the aspirant the choice to conduct nuclear research. Rather, containment makes nuclear research more difficult and expensive, which only reduces the aspirant's expected benefit (Spaniel and Smith, 2015). For some aspirants, containment shifts the cost-benefit calculus enough to induce them to give up their program (Miller, 2014). But other aspirants are not deterred. Countries like Iran and North Korea care enough about acquiring nuclear weapons because they have a strong interest in shifting the status quo. These states will continue nuclear weapons research even in the face of containment.

In my theory, great powers choose between containment, war, and making peaceful offers hoping the aspirant's nuclear research fails. When the aspirant's nuclear research is likely to produce nuclear weapons the great power prefers both forms of competition (containment and war) to a peaceful offer. Whether the great power selects containment or war depends on how much each costs, and how effective containment is at stalling the aspirant's nuclear program. When preventive war is sufficiently cheap, the great power selects it over containment.

However, when containment is sufficiently cheap and effective the great power selects it over war to thwart nuclear aspirants.[1] This is when the trouble starts. Since nuclear aspirants anticipate containment, they no longer fear war. Aspirants who care intensely about the bargaining advantage that nuclear weapons will give them prefer to keep investing in their nuclear program even in the face of containment. These types would have been deterred by a credible threat of war. But they are willing to start nuclear programs if the worst thing they face is containment. For these aspirants, containment causes deterrence to fail because it undermines the credible threat of war.

To be clear, I find that containment's affects on deterrence cut two ways. Consistent with existing literature, the threat of containment can deter aspirants from starting nuclear research programs. Inconsistent with existing literature, the promise of containment can also undermine the threat of war and cause deterrence to fail. I use game theoretic analysis to isolate the conditions under which containment helps and hurts deterrence. In particular, I analyze containment's effects as a function of preference divergence between the great power and the aspirant (Spaniel and Bils, 2017), and the aspirant's cost of conducting nuclear

[1] The mechanism is similar to McCormack and Pascoe (2015) who shows containment can reduce the chance of preventive war when conventional weapons produce power shifts.

research.[2] These variables are of great interest to U.S. policy-makers: If containment hurts deterrence success against states with preferences close to the United States, but facilitates deterrence against America's worst enemies then we need not worry about its adverse affects too much. There is some cause for optimism: past studies show that great powers credibly threaten preventive war when their foreign policy preferences significantly diverge from the aspirant's preferences (Monteiro and Debs, 2014). Given that war is most attractive when preferences diverge the most, containment may only undermine deterrence against friendly states.

Unfortunately, I find that the perverse effects of containment arise when the nuclear aspirant is: (1) well industrialized; and (2) holds preferences that diverge a lot from the great power's preferences. The reason is that the stakes must be high for the great power to choose containment and the nuclear aspirant to choose nuclear investment in the face of containment. In contrast, containment deters aspirants who do not care that much about what they are bargaining over. It follows that containment undermines deterrence in the cases that policymakers want deterrence to work the most: When containment is possible, only the great powers worst (in terms of preference divergence) and most capable adversaries are undeterred from starting nuclear weapons research.

I then turn my attention to how containment affects the risk of proliferation for ongoing nuclear research programs where the risk of proliferation is increasing over time. I model proliferation risk using the increasing risk function common in labor economics models that study employment choices with shifting vacancy rates (Rogerson, Shimer, and Wright, 2005; Diamond, 1982). This setting matches ongoing policy debates about the effects of sanctions against states like Iran and North Korea: the risk of proliferation increases as aspirants continues to invest in their nuclear program, the United States has repeated opportunities to fight wars and enact containment, and aspirants have repeated opportunities to end their nuclear program.

[2] Which I think about as industrial capacity.

I show that containment has positive short-term effects but perverse long-term effects. In the short-term, containment delays an aspirant's nuclear research. Some are deterred by the prospect of this delay. But others care so much about the benefits of nuclear weapons that they invest in the face of containment. For these aspirants, containment only "kicks the can down the road." It delays but does not stop proliferation. If the threat of war does not stop them, then nothing does.

Unfortunately, whenever we observe containment on the equilibrium path, it means that the long-run risk of proliferation is larger than it would have been if the great power never chose containment. The reason is that containment delays the great power's credible threat of war more than it delays the time it takes for the aspirant to discover nuclear weapons. Thus, once a great power enacts containment, they push their credible threat of war down the road to such an extent that the aspirant has a larger overall probability of discovering nuclear weapons.

My findings have important implications for how states choose to develop instruments for coercion. Researchers and policy-makers are constantly searching for innovative ways great powers can control their rivals (Spaniel and Smith, 2015; Miller, 2014; Drezner, 2011; Volpe, 2017; Bas and Coe, 2018; Carnegie and Carson, 2018). There is enormous value from isolating these effects and understanding their benefits. But sometimes these instruments undermine each other when deployed simultaneously (Mehta and Whitlark, 2017; Feaver and Niou, 1996; Narang and Mehta, 2017). In rational theory, containment is widely thought to improve (or at least not worsen) American national security interests, and more broadly promote peace. I show that making containment an option creates opportunities for our worst adversaries to pursue nuclear weapons.

1.1 War, containment and state-motives in nuclear bargaining

I focus on two forms of competition great powers can use to deal with nuclear aspirants: major war and containment. By major war I mean the deployment of military forces to topple the government of the nuclear aspirant, totally dismantle their nuclear program, or install a puppet regime that will not pursue nuclear weapons (Powell, 2006, 1999). In short, war seeks to eliminate the aspirant's choice to pursue nuclear research.[3] In practice, great powers often consider but rarely use major war to terminate an aspirant's nuclear program. The United States used war against Iraq, and considered it against the Soviets (1949), North Korea (1994, 2001) and Iran (2006). The Soviets also considered using preventive war against China. The fact that war is often considered but rarely used matches an important insight about selection into nuclear programs from past research: nuclear aspirants will not start nuclear programs if they believe doing so will trigger major war.[4]

By containment, I mean any policy that reduces the probability that the aspirant's nuclear research program will succeed by making it more difficult or expensive for the aspirant to develop nuclear weapons. Containment changes the relationship between the amount the aspirant invests in its nuclear program, and the aspirant's expectation that the program will produce a nuclear weapon but does not seek to overthrow the aspirant's regime, or deny them the decision to pursue nuclear weapons research. I do not include the development of conventional forces (a-la Coe, 2018), or the forward deployment of forces (a-la Gartzke and Kroenig, 2014) as types of nuclear containment. I also do not include exclusively punitive sanctions that do not effect the probability a nuclear program will succeed (Drezner, 2011).[5] In practice, the United States and other great powers restrict the transfer of key technologies to all other states through non-proliferation cartels (Coe and Vaynman, 2015). I consider these part of the overall landscape of world politics and not a specific containment policy. Rather, containment policies must target the proliferation efforts of a specific nuclear aspirant.

[3] To be clear, I consider military strikes against an aspirant's nuclear facilities as a form of containment, not war (a-la Kreps and Fuhrmann, 2011).

[4] War in the Iraq case may have resulted from information asymmetries (Debs and Monteiro, 2014; Bas and Coe, 2016).

[5] In practice, many sanctions regimes are designed to both make nuclear proliferation more difficult and place economic pressure on a regime. So long as the purpose is in part to make the development of nuclear weapons more difficult my theory should apply.

To be clear I distinguish between types of competition based on their consequences[6] and not their form.[7] Doing this helps policy-focused researchers turn the assumptions of formal models into policy insights. For example, in making the case for US military intervention against Iran, Kroenig (2012) argues that the "United States should attack Iran and attempt to eliminate its nuclear facilities." Kroenig's logic explicitly draws from theories of preventive war that assume military strikes will deny an aspirant the opportunity to seek nuclear weapons. As Kahl (2012) points out, striking Iran's nuclear facility would only delay Iran's program a few years and may even induce rally round the flag effects just like sanctions would. In effect, Kroenig's recommendations are grounded in a formal logic of preventive war as a game-ending move. But since the war he recommends are designed to delay and not end Iran's program, it is unclear that the logic would still hold.

I present two models that capture the deterrent effects of containment at different phases of nuclear research. First, I study a two-period model to understand how the opportunity for containment effects an aspirant's choice to start a nuclear program. Second, I study an infinite horizon model where the aspirant has repeated opportunities to invest and the great power has repeated opportunities to compete. I use this model to understand how the opportunity for containment can deter aspirants as their program develops, and the implications that containment has for the long-run probability of proliferation success.

To show that my results do not follow from secrecy or other information asymmetries, I assume both players have complete information about the aspirant's nuclear research choices. In practice, nuclear aspirants often start their nuclear programs in secret. However, history shows us that even the most careful nuclear aspirants have their programs exposed before they successfully develop nuclear weapons and convert that research into military power. In this way, I think of my two-period model as starting once a nuclear aspirant thinks about investing in a nuclear weapons program, and ends in the years after their nuclear weapons

[6] Whether competition denies the aspirant from seeking nuclear weapons altogether, or makes proliferation more difficult.

[7] Whether it is military, clandestine, economic or based on social mobilization

program is discovered by the great power's intelligence community. This period carries large incentives for preventive war (Bas and Coe, 2016). As a result, this is exactly the time that war should have the largest deterrent effect.

2 Containment, war and deterrence at the onset of nuclear programs

I model an interaction between two states, a nuclear aspirant (A, male) and major power (B, female) that bargain over setting a policy q. B's ideal policy is $q = 0$. A's ideal policy is $q = \pi$, where $\pi \in (0, 1]$. High values of π reflect A and B have divergent policy preferences. Both players have linear preferences and so their utility from any bargain is 1 less the linear distance from their ideal policy. The model lasts two-periods: $t \in \{1, 2\}$ and players discount the future by $\delta \in (0, 1)$. In each period, players bargain to set the policy in that period q_t with an outside options of war (for both), containment (for B) and nuclear research (for A).

Each period starts with A choosing to invest or not in his nuclear program. A's investment in period $t - 1$ has two effects. First, A pays a costs R (for research) in that period and only that period. I think about high values for R as representing nuclear aspirants with low industrial and scientific capacity. These states need to dedicate more resources to yield the same expected success of a nuclear program.

Second, at the beginning of the next period (t), there is a $(1 - \lambda)r_{t-1}$ probability A discovers nuclear weapons, and a $\lambda r_{t-1} + (1 - r_{t-1})$ probability that A does not discover nuclear weapons. In this equation, r_{t-1} is an indicator function equal to 1 if A invested in nuclear research at $t - 1$ and 0 otherwise. Thus, when A does not invest in nuclear research at $t - 1$, he has no chance of discovering nuclear weapons in the next period ($r_{t-1} = 0$). $1 - \lambda$ is the probability that A's active research program will produce nuclear weapons.

Nuclear weapons alter the distribution of power between A and B.[8] At the beginning

[8] As we shall see in a moment, relative power influences the outcome of bargaining and war.

of the game A starts with p_0 power. This reflects A's conventional military capabilities. However, if A discovers nuclear weapons in period t, her power shifts from $p_{t-1} = p_0$ to $p_t = p_0 + \Delta$.

After B observes A's choice to invest in his nuclear program, B chooses between (1) major war, (2) making a peaceful offer, or (3) making an offer under nuclear containment.

If B selects (1) major war, bargain ends immediately and players enter into a terminal sub-game war. War forces a costly lottery which A wins with probability p_t and B wins with probability $1 - p_t$. War costs both players w in that period and every subsequent period. A's one period expected value from war is:

$$U_t^A(War) : p_t + (1 - p_t)(1 - \pi) - w - Rr_t \equiv 1 - \pi(1 - p_t) - w - Rr_t. \tag{1}$$

The first term, is A's probability of winning (p_t) multiplied by A's value from setting the policy π (which is 1). The second term is A's probability of losing the war $(1-p_t)$ multiplied by A's expectation that B will set her ideal policy (0), which leaves A with $1 - \pi$. The final term is A's cost from choosing to invest in research in that period (R), which A only pays if she invested in research ($r_t = 1$).[9] If B chooses war in the first period, then A's total two-period expected utility is:

$$EU_1^A(War) : (1 - \pi(1 - p_1) - w)(1 + \delta) - R(r_1 + \delta r_2). \tag{2}$$

Similarly, B's one period expected value from war is: $U_t^B(War) : p_t(1-\pi)+(1-p_t)-w \equiv 1 - p_t\pi - w$ and B's total expected utility from fighting in the first period is:

$$EU_1^B(War) : (1 - p_1\pi - w)(1 + \delta). \tag{3}$$

If B selects either (2) a peaceful offer or (3) containment, B must offer A some policy

[9] I use notation U_t^{player} to describe a player's one period pay-off in period t. I use EU_t^{player} to describe a player's total expected utility starting in period t and extending to all future periods (and discounted).

compromise q_t. I assume that B's offer q_t is restricted by a status quo bias such that B's offer must be at least $q_0 = 1 - \pi(1 - p_0) - w$. I set q_0 at A's minimum demand from war under the assumption that A never invests in her nuclear program. Substantively, this minimum offer reflects a long-standing status quo between A and B over the policy in dispute before the aspirant considered nuclear weapons research. Including a status quo bias into the model creates a commitment problem such that B prefers fighting a war rather than risking nuclear proliferation when power shifts sufficiently fast and A has a good chance of acquiring nuclear weapons in the next period (λ is high).[10]

If B selects a peaceful offer, B makes a policy proposal q_t and the game moves to A's decision. However, if B makes an offer under containment, B also pays a cost c (for containment). In return, containment reduces the probability that A will discover nuclear weapons in the next period and all future periods. A's chance of success changes from $(1 - \lambda)r_{t-1}$ to $(1 - \lambda_c)r_{t-1}$ where $1 - \lambda_c < 1 - \lambda$.[11]

Finally, A chooses between accepting B's offer or war. A's one period pay-off from accepting an offer is $U_t^A(accept) : 1 - |q_t - \pi| - Rr_t$. B's one period payoff from accepting an offer is $U_t^A(accept) : 1 - q_t - Cc_t$. Here c_t is an indicator function equal to 1 in the period that B enacts containment.

In summary, the sequence of moves for one period of the game (period t) is as follows:

- A discovers nuclear weapons with probability $(1 - \lambda)r_{t-1}$.
- A chooses to either: invest in nuclear research, or not invest.
- B chooses to either: bargain peacefully, bargain under containment, fight a war.
  - If B chooses war, enter sub-game war where bargaining stops.
  - If B chooses containment, enter sub-game containment where bargaining continues, but $\lambda \to \lambda_c$ in future periods.
  - If B chooses a peaceful offer, bargaining continues.

[10] I reach the same conclusions if I induce commitment problems through manipulating bargaining protocols across rounds (a-la Coe, 2018), or restricting side-payments and exploiting the limited size of the policy space (a-la Krainin, 2017), or assuming that the issue is indivisible.

[11] This well models nuclear containment because it only effects A's chances of nuclear discovery and not the underlying balance of power. Further, it only has an effect on A if A has chosen nuclear investment. To be clear, if A discovers nuclear weapons under containment, power still shifts by Δ.

- If bargaining continues, A chooses to accept B's offer or major war.
- If A chooses to accept, period payoffs are realized and the next period starts. If A chooses war, enter sub-game war.

Before moving to my analysis of how containment influences proliferation choices, I want to emphasize two features of any sub-game perfect equilibrium that are similar to Debs and Monteiro (2014)'s complete information result. Highlighting these similarities demonstrates that the core of our models are similar, making it easier for readers to understand how including containment builds on existing work. The results also rule out most second period strategies in any SPNE and so I can focus my analysis on B's first period strategic choices.

Lemma 2.1 *War cannot appear on the path in any sub-game perfect Nash Equilibrium.*

B's incentives for war come from preventing A from discovering nuclear weapons. If A does not invest in nuclear research, B has no incentive to fight a major war. If A does invest in nuclear research, B can have an incentive to fight. However, A knows that war is only triggered by her nuclear research. A's value for war today is always less than not investing in nuclear research and accepting the status quo because nuclear research is costly (R).[12]

Define q_2^* as the offer that leaves A indifferent to war in the second period. A's second period utility from war is: $U_2^A(war) : 1 - \pi(1-p_2) - w$. The offer that leaves A indifferent with this value satisfies: $1 - |\pi - q_2^*| = 1 - \pi(1-p) - w \implies q_2^* = p_2\pi - w$. Then,

Lemma 2.2 *In the second period of any SPNE, A does not invest in nuclear research, B makes a peaceful offer q_2^* yielding:*

$$U_2^A(q_2^*) : 1 - \pi(1-p_2) - w \tag{4}$$

$$U_2^B(q_2^*) : 1 - \pi p_2 + w \tag{5}$$

See Appendix A.1. Lemma 2.2 states that B will always make a second period offer that leaves A indifferent with fighting a war in the second period. The exact size of the offer

[12] To be clear, this does not mean that B's credible threat of war cannot deter A, or that A is always deterred. Both of these outcomes are possible.

depends on whether or not A discovered nuclear weapons in the first period. If A discovered nuclear weapons then $q_2^* = \pi(p_0 + \Delta) - w$. If A did not discover nuclear weapons then $q_2^* = \pi p_0 - w = q_0$.

I can exploit the result in 2.2 to focus on each player's first period strategic choices (s^A, s^B). Each player's total expected utility in the first period can be summarized as:

$$EU_1^A : \overbrace{U_1^A(s_1^A, s_1^B)}^{\text{A's period 1 pay-off}} + \overbrace{(1 - (\lambda|c_1))r_1}^{\text{pr. A gets nukes at period 2 onset}|c_1, r_1} \delta \overbrace{(1 - \pi(1 - p_0 - \Delta) - w)}^{\text{A's period 2 pay-off (discounted) if A gets nukes}} + \overbrace{(\lambda|c_1)r_1}^{\text{pr. A doesn't get nukes at period 2 onset}|c_1, r_1} \delta \overbrace{(1 - \pi(1 - p_0) - w))}^{\text{A's period 2 pay-off (discounted) if A doesn't gets nukes}} \tag{6}$$

$$EU_1^B : U_1^B(s_1^A, s_1^B) + \delta((1 - (\lambda|c_1))r_1(1 - \pi(p_0 + \Delta) - w) + (\lambda|c_1)r_1(\pi p_0 - w)) \tag{7}$$

A's total expected utility emphasizes that both player's first period decisions (s^A, s^B) impact their second period pay-offs because they alter the probability that A acquires nuclear weapons in the second period. A gets nuclear weapons at the beginning of the second period as the result of a random component (Nature's choice λ) and three strategic components: whether A invests $(r_1 = 1)$, whether B chooses containment $(\lambda \to \lambda_c)$ or whether B chooses to fight a war before A's research can be implemented. If A does get nuclear weapons, she can demand $\pi\Delta$ more in the second period.

In Appendix A.2, I write out each player's first period expected utilities for every possible first-period strategic choice they can make. There are several possible outcomes to consider leading to many different equilibrium. My goal is to focus exclusively on how different forms of competition influence B's ability to deter A, and how that influences the risk of nuclear proliferation.

2.1 Analysis of deterrence success and failure

The major claim of this paper is that containment undermines the credible threat of war and causes deterrence to fail. To highlight this result, I structure the analysis in three sections. First I define deterrence success and failure. Second, I identify the conditions under which deterrence succeeds. Finally, I identify the conditions under which deterrence fails.

I say deterrence succeeds when:

1. **A wants to invest:** If A expects no competition in period 1 even if A invests in nuclear research, A prefers to invest than not invest.
2. **B's threat is credible:** B can credibly promise that if A invests in his nuclear program, B will compete (either through containment or war).
3. **B's threat deters A from investment:** Anticipating his nuclear investment will trigger competition, A forgoes nuclear investment.

Under this definition two conditions arise where deterrence neither succeeds nor fails: A never intends on investing in his nuclear program; B always intends to make peaceful offers.

Lemma 2.3 ***A does not want to invest:*** *If*

$$R > \pi\delta\Delta(1 - \lambda) \tag{8}$$

is satisfied, then there is a unique equilibrium where A forgoes investment in period 1 and period 2. B makes peaceful offers in both periods.

Inequality 8 defines the types of aspirants that are not deterred because they never intended on investing in their military. Intuitively, A does not want to invest in nuclear research when it is expensive given A's industrial capacity (R is high) relative to either the chance of success (λ), or the effects of nuclear proliferation (Δ). Further, A will not invest if his preferences are closely aligned with B's (π is small). The reason is that when A and B have similar policy goals, there is not much of a concession to make. Thus, A has little use for additional bargaining leverage.

Lemma 2.4 ***B never wants to compete:*** *If B prefers a peaceful offer to war,*

$$w > \frac{\Delta\pi\delta(1-\lambda)}{2(1+\delta)}, \tag{9}$$

and a peaceful offer to containment,

$$c > (\lambda_c - \lambda)\Delta\pi\delta \tag{10}$$

then in any equilibrium where A invests in nuclear research B will make peaceful offers in both periods.

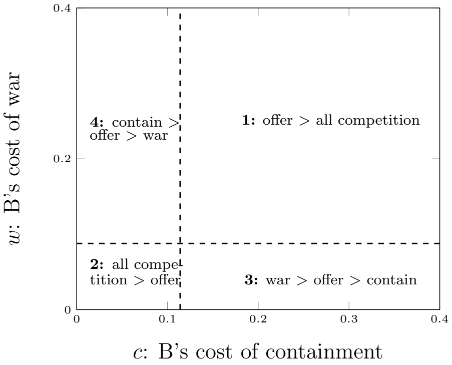
When inequalities 9 and 10 both hold deterrence cannot fail because B prefers peaceful offers to any form of competition. I prove this in Appendix A.4. Figure 1 summarizes the results of Lemma 2.4 by showing the ranges of the parameters where deterrence does and does not apply. To be clear, this is not an equilibrium plot. Rather it marks the conditions where B is and is not willing to compete assuming A invested in her nuclear program. In region **1**, all forms of competition are so expensive that B makes peaceful offers no matter what A does. Here deterrence does not apply.

In region **2**, B prefers both forms of competition to a peaceful offer. I call this region the two-competition region. In this region the threat of competition is remarkably strong. Policy-makers that think containment and war are complimentary might think about these conditions as creating redundant threats that a great power could use to deter A from investment. After all, if there was a sudden shock to the cost of war, B could still credibly threaten containment.

Although redundant forms of competition are nice, they are not necessary for deterrence

Figure 1: When does B prefer competition?

The following plots depicts the spaces where B can credibly threaten different forms of competition as a function of the cost of war and the cost of competition. The plot holds constant $\pi = 1$, $\Delta = .4$, $p = .3$, $\lambda = .4$, $\lambda_c = .6$. All competition refers to both containment and war.



to succeed. I now describe the conditions under which war and containment successfully deter A from investment.

Lemma 2.5 ***Containment deters:*** *When B's best reply to A's nuclear investment is containment, and:*

$$\Delta\pi\delta(1-\lambda) > R > \Delta\pi\delta(1-\lambda_c), \tag{11}$$

is satisfied, then B's credible threat of containment deters A from nuclear research. In equilibrium, A does not invest and B makes a peaceful offer.

Proof of Lemma 2.5 is in Appendix A.5. The intuition is identical to what Miller (2014) described: B can credibly threaten to contain any Aspirant that invests. A anticipates this and weighs the benefits of investing under containment against not investing at all. For aspirants who satisfy $R > \Delta\pi\delta(1-\lambda_c)$, the cost of investment under containment is not worth the expected benefit. These types do not invest. Within that set, those who satisfy $R < \Delta\pi\delta(1-\lambda)$ would have invested if they thought that B would make a peaceful offer. These types are deterred from investment because B's credible threat of containment prevents them from investing. The effect of containment on A's chance of success ($1-\lambda \rightarrow 1-\lambda_c$) drives the result because it reduces A's expectation that his investment will payoff.

Lemma 2.6 ***War deters:*** *When B's best reply to A's nuclear investment is war and:*

$$\Delta\pi\delta(1-\lambda) > R, \tag{12}$$

is satisfied, then B's credible threat of war deters A from nuclear research. In equilibrium, A does not invest and B makes a peaceful offer.

Proof of Lemma 2.6 is in Appendix A.6. The intuition is similar to Lemma 2.5. But there are important differences between the results. First, there is only one condition on R that must be satisfied for war to deter A. War differs from containment because war terminates bargaining and eliminates A's benefits from acquiring nuclear weapons in expectation. When A anticipates that his investment will trigger war, A never invests because investment is costly and he cannot profit from it.

Second, the conditions under which each type of competition is credible depends on the costs of each and these costs vary independently (w, c). As the cost of war (containment) increases, B's benefit from war (containment) decreases but it has no effect on B's value from selecting containment (war).

The differences in the conditions between war and containment are consistent with existing intuition. When the threat of war is credible, it provides the most powerful deterrent effect. Yet war is so costly it is credible under fewer conditions. Containment does not always deter, but great powers can threaten it under many more conditions. In this way, different types of threats can have complimentary deterrent effects. Even when powerful states cannot rely on threats of major war, they may be able to rely on the threat of containment to deter possible nuclear aspirants. Thus, having diverse threats can expand the conditions that deterrence works.

We might expect that nuclear proliferation is incredibly unlikely when the cost of all types of competition are low. After all, both types of competition have independent deterrent effects. If both are credible, A must face some form of competition if he invests. However,

Proposition 2.7 ***Deterrence can fail in the two-competition region:*** *Suppose B* *prefers both forms of competition to making a peaceful offer, but B prefers containment to war:*

$$2w(1+\delta) > \Delta\pi\delta(1-\lambda_c) + c. \tag{13}$$

Then if A prefers to invest and face containment rather than not invest at all,

$$\Delta\pi\delta(1-\lambda_c) > R, \tag{14}$$

deterrence fails. In equilibrium, A invests in his nuclear program in the first period and B contains. There is a positive probability that A discovers nuclear weapons at the beginning of the second period.

Corollary 2.8 ***Containment causes deterrence to fail:*** *For any set of parameters that satisfy equilibrium conditions in Proposition 2.7 consider a counter-factual game where containment was not an option. In any SPNE for this counter-factual, A is deterred by B's credible threat of war. A has no chance of discovering nuclear weapons at the beginning of the second period.*

The proof of propositions 2.7 and corollary 2.8 are in Appendix A.7. The intuition follows from the differential effects of containment and war. From A's perspective, these different types of competition lead to very different consequences. Following war there is no more bargaining. As a result, when B can credibly threaten to engage in war, A is always deterred. In contrast, containment undermines B's nuclear program but does not end it entirely. When A cares intensely about acquiring nuclear weapons relative to the cost, A is willing to invest in his nuclear program even in the face of containment.

One might wonder why B would even pick containment over war? The reason is that containment can be much cheaper than war ($c < w$) but still dramatically reduce A's chance of successful proliferation when $\lambda_c >> \lambda$. When containment is both cheap and effective B prefers it to war. Anticipating containment not war, A prefers to invest and suffer the costs of containment if A's cost of nuclear research is sufficiently low.

2.2 When does containment help or hurt deterrence?

The promise of containment can undermine the threat of war and cause deterrence to fail. But the threat of containment can also deter nuclear programs by raising the risk that the research will fail. If containment only hurts deterrence under a small number of conditions, or only against aspirants that are benign, then we need not worry about it.

In this section I study the conditions under which containment helps and hurts deterrence as a function of the aspirant's industrial capacity (R) and preference divergence from the great power (π). I chose these variables because they are the most important to great powers. Great powers do not care equally about all nuclear aspirants. They worry the most about the most capable adversaries who want to undermine the existing world order. In practice, the US would care a great deal more if Iran acquired nuclear weapons than if Fiji did.

Unfortunately, my theory predicts bad news:

Prediction 1 *Containment causes deterrence to fail only for highly industrialized and it is expensive for them to conduct nuclear research (R is high). It causes deterrence to succeed when nuclear aspirants are weakly industrialized and it is cheap for them to conduct nuclear research (R is low).*

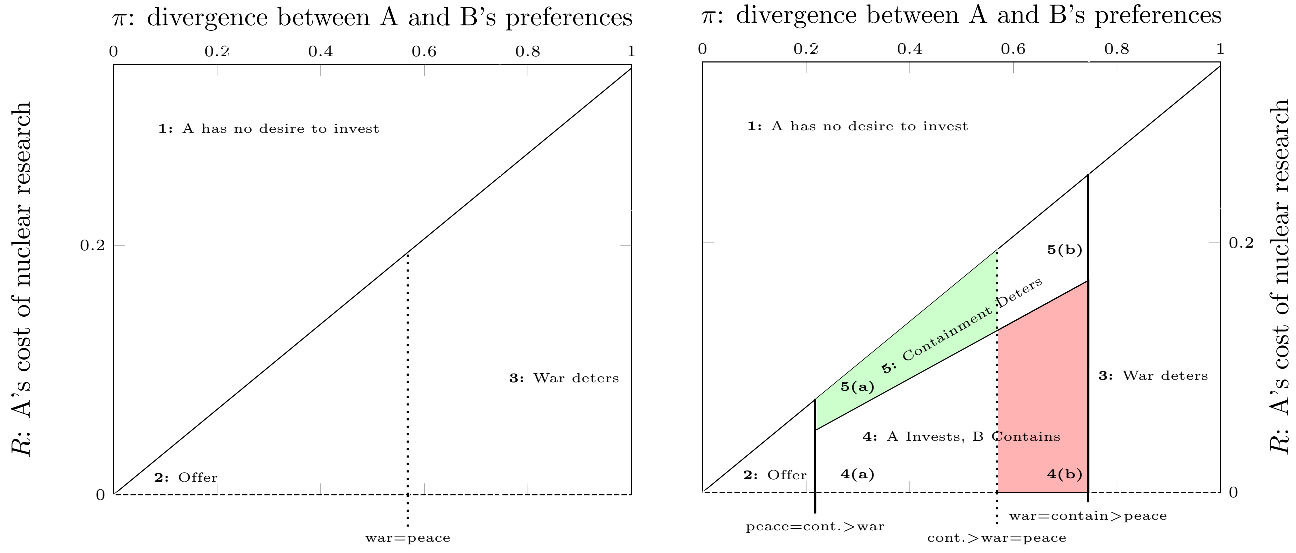
Prediction 2 *Containment causes deterrence to fail when nuclear aspirants preferences diverge significantly from the great powers (π is high). It causes deterrence to succeed when nuclear aspirants preferences are close to the great powers (π is low).*

I visualize these expectations in Figure 2. The Figure plots the model's equilibria as a function of π and R. In panel (a) the cost of containment is set so high that B never prefers containment over either war or a peaceful offer. In panel (b) the cost of containment is set low enough that B prefers containment over both options at middle values of π. All other parameters are held constant. The two plots allow me to compare two worlds: One where containment is allowed and the other when it is not (or is cost prohibitive). When containment is allowed two new equilibrium emerge. One where containment deters A from investment (**5**), and another where A invests and B contains (**4**).

In both plots, the dotted vertical line marks the point where B is indifferent between war and a making a peaceful offer following A's investment. To the right of this line in panel (a) war deters A from investment. But in panel (b) B prefers containment to both war and a peaceful offer. In this case, the dotted line does not delineate different equilibria. Rather,

Figure 2: When does containment help or hurt deterrence?

The plots depicts different equilibrium results as a function of A's cost of investment (R) and the divergent preferences of both states (π). Panel (a) sets $c = w$ so that B always prefers war over containment. Panel (b) sets $c < w$ so that B can prefer containment over war.
Inequalities below the x-axis emphasize that vertical lines mark B's preference for war, peace or containment.



(a) Cost of containment is high

The dotted line marks B's point of indifference between war and making a peaceful offer if A invests.

(b) Cost of containment is low

In **5(a)** (shaded green) containment creates deterrence. In **4(b)** (shaded red) containment undermines deterrence. Only solid lines mark equilibrim spaces. Dotted line emphasizes what B would have done if containment was not an option.

it marks alternative counter-factuals: what B would have done if containment was not an option.

I emphasize the implications of B's different counter-factual strategies through shading parts of the equilibrium plot. **5(a)**, shaded green, marks the secret success of containment. In this space, war is too costly to be credible. If B could not credibly threaten containment, A would invest in his nuclear research and B would make peaceful offers (see panel (a)). In this region, containment facilitates deterrence because reduces A's expected benefit from nuclear research. Region **4(b)**, shaded red, marks the secret failure of nuclear containment. If containment was not an option, B's credible threat of war would deter A from nuclear research. But B prefers to enact containment rather than fight. Knowing this, A invests and faces containment.

Comparing the red and green areas, containment undermines deterrence when R is low and π is high. When preferences diverge significantly the stakes are higher for both players. This means B prefers both forms of competition to a peaceful offer. This also means that A is willing to invest under the threat of sanction. When the stakes are lower, B is less willing to pay the cost of war, and A is less willing to invest under containment. Similarly, A is only undeterred by containment when the cost of investment is sufficiently low. The perverse effects only arise for nuclear aspirants that are willing to bare the cost of research even when they face containment.

2.3 Re-interpreting Existing Evidence

Predictions 1 and 2 rely on counter-factual reasoning: A would not start nuclear weapons research if B could not credibly threaten containment because A worried that B would choose war instead. It is difficult to observe this counter-factual. Leadership deliberations about starting nuclear programs are usually classified. Even if they were available, I cannot observe the decisions aspirants would have made if they thought they would face war instead of containment. To overcome these empirical challenges, I exploit a sudden shock in the cost of containment for the United States identified by Miller (2014). Like Miller, I exploit differences in the patterns of proliferation at each side of the shock to show that the properties of aspirants are different in periods when the US could credibly threaten containment (because the cost was low) and when the US could not credibly threaten containment (because the cost was high).

Miller (2014) argued that the 1973 US-led sanctions effort against South Africa altered the ease in which the United States could sanction nuclear aspirants going forward. Before 1973, the United States had not previously developed legal instruments for sanctions, nor coordinated with international partners to impose sanctions on nuclear aspirants. Completing these tasks for the first time cost the Nixon Administration considerable time and attention. But Nixon paid these up-front costs to sanction South Africa's nuclear program. Once Nixon developed these tools, all future Administrations could rely on them. Therefore, the cost of implementing sanctions after 1973 was lower. Miller exploited 1973 as a cut-point. He reasoned that pre-1973 sanctions were less credible because aspirants believed US sanctions were very costly (c was high). Post-1973 aspirants believed US sanctions were credible because sanctions were cheap (c was low).[13]

I borrow Miller's approach to provide support for my theory. I break out cases of nuclear proliferation into the periods before and after 1973. I then plot the year that aspirants started their programs as a function of their industrial capacity (R), and preference divergence from the United States (π).[14] I measure preference divergence from the United States using Bailey, Strezhnev, and Voeten (2017) measure based on UN roll-call votes. The measure produces a score that is the distance in some state's policy preferences from the United States.[15] I measure the aspirant's industrial capacity using Banks and Wilson (2014)'s logged industrial production measure.

The results are reported in Figure 3 and are organized in the same way as the comparative static result reported in Figure 2.[16] The countries colored blue started their nuclear programs before 1973 when containment was costly for the United States. The countries colored red started their nuclear programs after 1973 when containment was less costly for the United States. The year in which a country started their nuclear program is reported in parentheses. We can think of the case plot in Figure 3 as overlaying the two counter-factual worlds reported in Figure 2. Pre-1973 cases of proliferation onset should conform to Figure 2(a) and post-1973 cases should conform to Figure 2(b).

For emphasis, I drew lines on the plot to match the comparative static results in Figure

[13] Since I make non-linear predictions about the effect of containment, that interact with other variables, and also given that there are so few cases of new nuclear research programs, regression analysis is not appropriate.

[14] Since there are only 16 country-year cases of nuclear programs, and my results are broken out into two time periods, with two predictor variables, I do not attempt a cross-national time series regression.

[15] The years Miller records North and South Korea as starting their nuclear programs are missing from Bailey et al. (2017). As a result, I take the closest years for these states.

[16] Notice that high values of industrial capacity are inverted to match the plotting of R.

2(b). The exact position of these lines is arbitrary. But if my theory well explains the data, I should be able to partition cases into 4 distinct regions consistent with regions marked 4(a), 4(b), 5(a), 5(b) in Figure 2(b).

Consistent with my argument that containment can undermine the threat of war and cause deterrence to fail (4(b)), I can isolate a region where I only observe nuclear research programs starting after 1973. The aspirants in this space are Iran, Iraq and North Korea. As predicted, all three of these aspirants started nuclear weapons research programs and faced sanction from the United States. Thus, the threat of containment did not deter them. Instead, the were willing to start nuclear weapons programs and face containment.

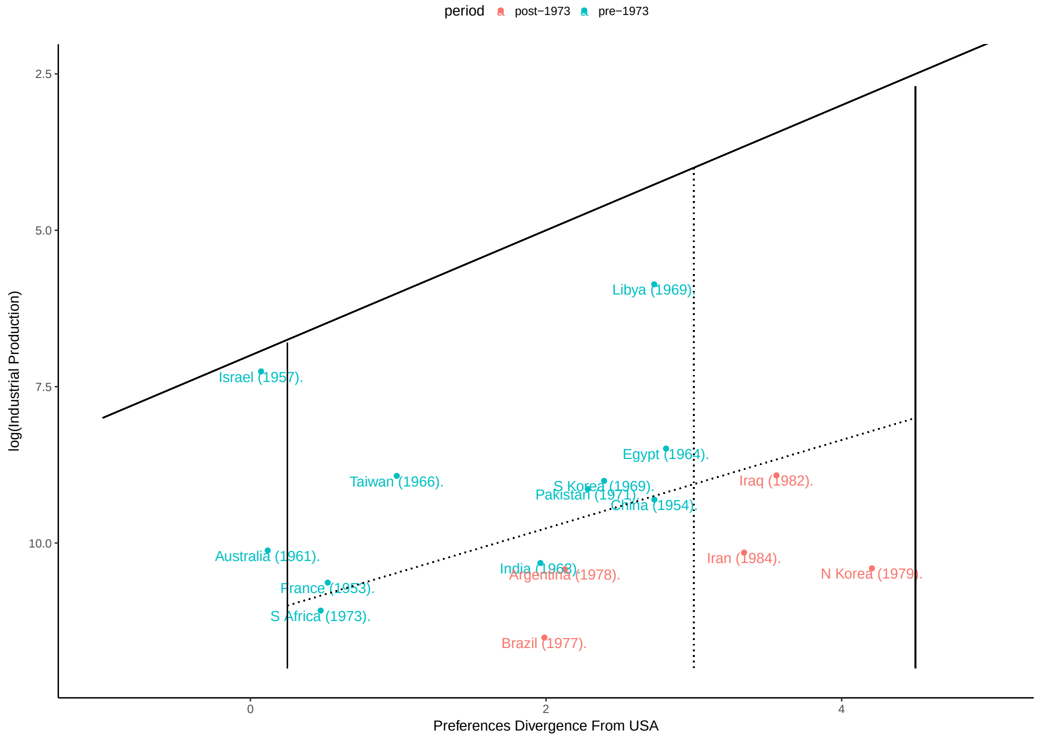
Consistent with my argument that the threat of containment can facilitate deterrence (5(a)), I can isolate a region where I only observe nuclear research programs starting before 1973. The aspirants in this space include Egypt, South Korea, Taiwan and France. Consistent with Miller's logic, we observe no cases from the post-1973 period because they were deterred by the credible threat of containment.

Consistent with my argument that containment and war can both deter (5(b)), I can isolate a region where there are no aspirants from either period. Furthermore, I can isolate a region with aspirants from both periods (4(a)).

Figure 3 provide a new interpretation of Miller's result. I find that once sanctions became a plausible response to nuclear aspirants the types of states that started nuclear weapons programs changed. Consistent with Miller's logic, I find that states with limited industrial capacity and foreign policy interests similar to the United States were deterred by the credible threat of sanction. Although we may never know which states were successfully deterred by the threat of sanctions post-1973, the evidence and theory suggests that these aspirants were similar to South Korea (1969), Taiwan (1963) and France (1953).

Inconsistent with existing thinking, I also find that once nuclear sanctions became credible industrialized aspirants with preferences far away from the United States started nuclear programs. It appears that sanctions made way for states like Iran, North Korea and Iraq to

Figure 3: Cases of nuclear research onset by periods of high/low cost of containment.



pursue nuclear programs.

These results are the exact opposite of what American policy-makers would want. Policy-makers hope that threats work the best against highly capable adversaries who want to impose radically different policies. It seems that this is precisely the condition under which containment undermines the deterrent threat of a military strike causing deterrence to fail.

3 Containment as the risk of proliferation increases.

In the baseline model I studied how containment affected an aspirant's choice to start a nuclear program. But nuclear weapons take decades to develop under the best conditions. The expected probability that an aspirant will discover the bomb increases with time; and with it the great power's incentives for competition. Aspirants have many opportunities to negotiate with their programs (Volpe, 2017), and great powers have many opportunities to fight (Bas and Coe, 2016). How does containment influence the likelihood of proliferation as the aspirant's research program advances?

Indeed, policy-makers fiercely debate the prospect of competition as nuclear aspirants develop their programs. For example, when Iran started its nuclear program in 1982, US policy-makers took no action against it. In the decades that followed the United States altered its policy to match Iran's progress. In 1996 the US imposed sanctions against Iran's energy sector and key scientific technologies. Through the 2000s, the US applied tighter sanctions and used targeted military strikes and covert operations against Iran's nuclear facilities to delay Iran's progress.

At each of these junctures, American policy-makers debated the merits of containment. Critics thought containment only "kicked the can down the road"[17] because containment delayed one of two permanent outcomes: Iran would either acquire nuclear weapons, or the US would invade to stop them. Critics argued that delaying the inevitable created additional risk that Iran's program would succeed and did not escape the fact that invasion was the only permanent solution. Supporters argued that, at minimum, containment delayed Iran's program for years. At maximum, containment raised the cost of proliferation by so much that Iran may just give up.

While these two positions have radically different policy implications, in the context of my theory, they emphasize two different parts of one rational response to containment's effects. As the supporters of containment observe, containment increases the time it takes for an aspirant to develop nuclear weapons by making each phase of the research process more difficult. Thus, the short-term risk of proliferation is lower if powerful states contain the aspirant. Even in the long term, containment implies that the aspirant must invest more for longer to develop nuclear weapons. But the critics are also correct: if the aspirant keeps investing he will eventually discover nuclear weapons even under containment. Thus, in the long-term containment delays but does not stop the aspirant's progress.

I'll now show that the long-term effects are more perverse than first thought. It turns out

[17] Most recently, those opposed to the JCPOA with Iran used this phrase (Izewicz, 2017), but it is commonly used to condemn sanction efforts.

that any time a great power chooses to contain a nuclear aspirant, they alter their strategic incentives in ways that guarantee the long-run probability that an aspirant will discover nuclear weapons is larger. The reason is that great powers can only credibly threaten war when the aspirant's program has a high risk of producing nuclear weapons in the immediate future. Containment does not alter the risk of proliferation that the great power is willing to accept. Rather, containment only increases the time it takes for the aspirant to reach that threshold. It turns out that containment's effects on short-term risk imply that the great power must wait longer before she can credibly threaten containment. As a result, the overall risk that the great power is willing to accept is higher.

In a perfect world, the great power would not use containment but instead would deter the aspirant through the credible threat of war. However, once the aspirant has invested in his nuclear program, the great power's short-term incentives drive her to select containment and kick the can down the road. Anticipating containment, the aspirant invests creating opportunities to develop nuclear weapons that he otherwise would not have had.

3.1 Dynamic nuclear proliferation

To study these dynamics, I transform the two-period baseline model ($t \in \{1, 2\}$) into an infinite horizon model $t \in \{1, 2, 3...\infty\}$. Most features of the model are the same. I still assume both players can enter a terminal sub-game war. I assume that there is an enforceable status quo (q_0) and that nuclear weapons shifts A's military power from p_0 to $p_0 + \Delta$. Furthermore, I still assume the same period payoffs from war, containment and peaceful offers and that players discount the future by a constant δ.[18]

The main difference between this extension and the baseline model is that A's nuclear investment and B's choice to enact containment now have cumulative effects. The more A invests in nuclear research, the closer he gets to discovering the bomb. Informally, I conceptualize τ_t as A's cumulative nuclear research effort from the first period up to the

[18] For simplicity, I set $\pi = 1$ in this analysis. Thus, this extension only considers the case where A and B's preferences diverge the most. But the results are substantively the same if I allow allow π to vary.

present period t. I will build a model where A acquires nuclear weapons only when A's cumulative research effort surpasses a point τ_N. That is, A gets nuclear weapons once $\tau_t > \tau_N$.

More precisely, I define τ_t conditional on whether or not A chose to invest in nuclear research in every period up until the present period t:

$$\tau_t = \sum_{j=1}^{t} r_j. \tag{15}$$

Subscript t reflects the current period, and j is some period in the history of the game $j \in \{1 : t\}$. r_j is an indicator function equal to one if A invested in her nuclear program in period j and zero otherwise. Every period that A invests in his military, τ_t increases by 1.

τ_N is the amount of nuclear research A must complete to acquire nuclear weapons that includes a random component and a strategic component. To get at the random component, I assume that at the beginning of the game Nature draws τ_n from a uniform distribution supported on $[0, \theta]$.[19] I assume both players know τ_n is drawn from $U[0, \theta]$. However, neither player observes the realization of τ_n. This assumption reflects the uncertainty of the nuclear research process. Both aspirants and great powers do not know when the aspirant will make a research breakthrough. But both know that if the aspirant keeps investing in nuclear research she will get closer but neither knows exactly when it will come.

If B never enacts containment, then $\tau_n = \tau_N$. However, B's containment policy increases the amount of research that A must complete to discover nuclear weapons. That is, suppose B enacts containment in period t_c and τ_c is A's cumulative investment in period t_c given the number of times A has invested in nuclear research up until that point, then:

$$\tau_N | \tau_c = (\tau_n - \tau_c + 1)k + \tau_c - 1 \tag{16}$$

[19] I use a simple density function to illustrate the intuition. But proposition 3.1 below holds for any continuous distribution function supported on the non-negative real numbers given that I can solve for an equilibrium where containment appears on the path.

In this equation $k > 1$ is the effectiveness of containment. It extends the number of periods it takes for A to discover nuclear weapons. Notice that k only effects τ_N for $\tau_n - \tau_c + 1$ investments. A's investment in prior periods is not effected. This assumption implies that containment only affects A's nuclear research in periods after containment takes effect and does not effect A's research in prior periods. This has an intuitive substantive interpretation. Containment makes it more difficult to conduct research and develop new technologies. However, aspirants keep all the research progress they made before they were burdened by sanctions or export controls.

This difference leads to the following change in the sequence of moves. In any period t:

- Nature gives A nuclear weapons if $\tau_{t-1} > \tau_N$ and not otherwise.
- A chooses to either: invest in his nuclear program setting $r_t = 1$, or not invest and setting $r_t = 0$.
- B chooses to: bargain peacefully, bargain under containment, or fight a war.
  - If B chooses war, enter sub-game war. Bargaining stops.
  - If B chooses containment set $t_c = t$ and $\tau_c = \tau_t$. B pays a cost c in that period. Bargaining continues.
  - If B chooses a peaceful offer, bargaining continues.
- A chooses war or accepts B's offer.
- Period t pay-offs are realized and the game repeats.

As in the baseline model, and consistent with existing research on nuclear containment, the effect of A's nuclear investment is still delayed. B's competition choices influence A's investment choice before A can profit from nuclear weapons research.

3.2 Analysis

My core argument is that when B chooses containment, she raises the long-run risk of proliferation. This claim relies on counter-factual reasoning. If D was unable to select containment, the risk of proliferation would be lower that it is in the game when B chooses containment holding all the parameters of the game constant. In Appendix B, I establish that there is a unique SPNE in which we observe containment on the equilibrium path. In Lemma B.5, I solve the model in the counter-factual world were containment is not allowed.

In the text, I provide a more intuitive understanding of the argument in two steps. First, I define long-term and short-term risk of proliferation. Second, I write out my core claim as a proposition, then provide an intuition for why it holds using pictures.

I define the short-term risk of proliferation as players' expectations that A will discover nuclear weapons in the next period and reference it $1 - \lambda_t | t_c$. Suppose in period $t-1$ B has not chosen containment, then in $t-1$ players believe that A will discover nuclear weapons at the beginning of period t with probability:

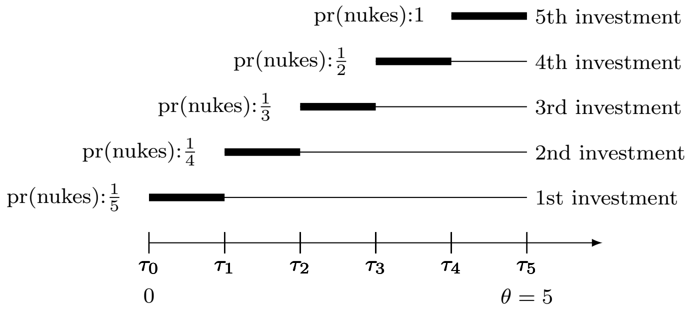
$$1 - \lambda_t | t_c \geq t = \frac{r_{t-1}}{\theta - \tau_{t-1}}. \tag{17}$$

The numerator equals 1 if A invested in nuclear research in period $t-1$. If A did not conduct research there is no short-term risk of proliferation ($1 - \lambda_t = 0$). The denominator is A and B's beliefs about all the plausible values that τ_N could take given that τ_n was drawn from a uniform distribution $[0, \theta]$ and both players have observed that A did not discover nuclear weapons in the first $t-1$ periods.

I illustrate how players' beliefs shift with A's investments through Figure 4(a). The Figure assumes that $\theta = 5$, and B never selects containment. The x-axis marks how the counter τ_t moves with A's nuclear investments. Above the x-axis are 5 lines. The total length of each line (including the thick and thin parts) represent players' beliefs about τ_n at different stages of the game. The thick part represents the draws of τ_n that will lead A to discover nuclear weapons in the next period. Each period that A invests and does not discover nuclear weapons the line gets shorter because players observe the fact that A did not get nuclear weapons and rule out draws of τ_n. Following each investment, the proportion of feasible types that discover nuclear weapons is larger than it was in the past, because the players rule out low values of τ_n each period. It follows that players increase their confidence that A will discover nuclear weapons with every investment because there are fewer draws of τ_n that are possible.

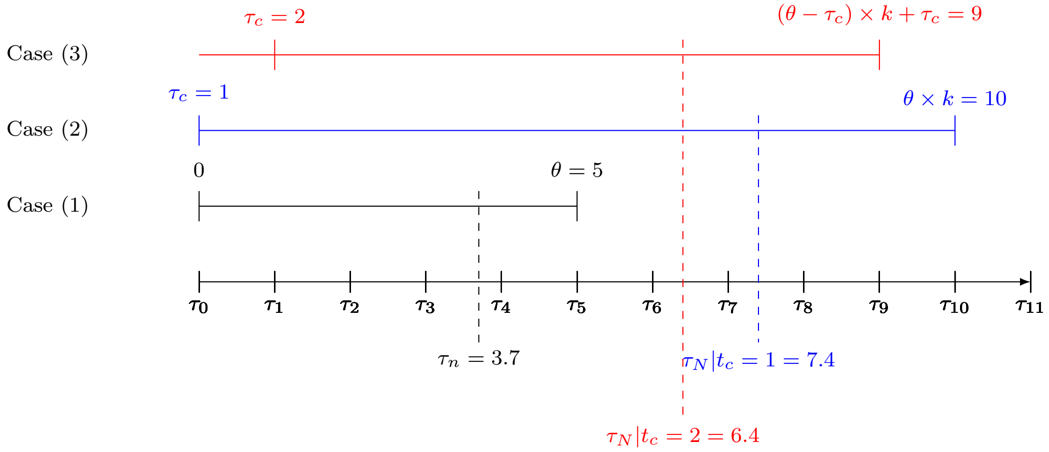
Figure 4: Understanding the cumulative investment function

(a) How beliefs update



The thin + thick line lengths represent players' beliefs about the total possible values of τ_n at each stage of the game given that A has not yet discovered nuclear weapons. The thick line represent the types who will discover nuclear weapons following A's τ_tth research attempt. The probability A's attempt is successful is the thick line over the total line length.

(b) Effects of containment



Each case plots the effects of containment at different points in the game holding $\theta = 5$, $k = 2$. In case (a) B never chooses containment. In case (b) B chooses containment in the period of A's first research attempt. It effects all of A's research attempts. In case (c) B chooses containment in the period of A's second research attempt. It effects all but A's first research attempt. The dashed line shows the consequences of containment on proliferation given a specific draw $\tau_n = 3.7$. The different line lengths represent how containment affects the players' total expectations about proliferation.

Containment influences players' beliefs about the short-term risk of proliferation by increasing the time it takes for A to discover nuclear weapons. If B has chosen containment in period $t-1$ or an earlier period, then both players believe that A will discover nuclear weapons at the beginning of period t with probability:

$$1-\lambda_t|t_c<t=\frac{r_t}{k\left[\theta-\tau_c+1\right]+\tau_c-\tau_{t-1}} \tag{18}$$

The core difference is that the denominator factors in the effects of containment on τ_N. I illustrate these effects through Figure 4(b), which builds on 4(a). In this example, containment doubles the number of investments A must make to achieve nuclear weapons ($k=2$). Case (1) colored black assumes that B never chooses containment. Case (2) colored blue assumes that A enacts containment before the first investment comes into effect. Case (3) colored red, assumes that B enacts containment following A's second investment.

The Figure emphasizes two points. First, the longer B waits to enact containment, the less containment effects A's research program. To illustrate this, I marked a sample draw of $\tau_n=3.7$. When B does not choose containment (case (1)) A discovers nuclear weapons after 4 research attempts. Containment extends the number of research attempts that A must make to surpass τ_N. If B enacts containment straight away (case (2)) then containment extends τ_N the most and it takes 8 research attempts for A to discover nuclear weapons. If B waits to enact containment (case (3)), then containment extends τ_N by less and A discovers nuclear weapons after fewer investments.

Second, containment increases players' beliefs about the number of investments it takes for A to acquire nuclear weapons. Each case marks the maximum possible value that τ_N could take given the respective containment choices. That is, it marks τ_N under the assumption that $\tau_n=\theta$. Clearly, the more effective containment is, the longer it will take for A to acquire nuclear weapons. But both players know this, and this extends their beliefs about how long it will take.

While short-term risk is important, my major claim is that containment raises the *long-run* risk of proliferation. By long-run risk, I mean B's expectation that A will acquire nuclear weapons at some point in the game given the equilibrium strategies that both states play $s_A^*(), s_B^*()$. Figure 4(b) is also helpful for understanding what I mean by long-term risk. Suppose A's strategy is to research for the first four periods of the game, then then stops and never research again ($r_t : \{1, 1, 1, 1, 0, 0, 0....0\}$). We can analyze the long-run risk of proliferation given B's different containment choices.

In all three scenarios, B accepts some risk that A acquires nuclear weapons during the game. In particular, if $\tau_N < \tau_4$ then A discovers nuclear weapons given that A will invest for 4 periods. We can therefore think about the long-run risk of proliferation as the realizations of τ_N where A would acquire nuclear weapons (i.e. $\tau_4 - 0$) divided by the total length of the line $k\left[\theta - \tau_c + 1\right] + \tau_c$.[20] Holding the number of A's nuclear research investments constant, the long-run risk of proliferation is lowest if B chooses containment straight away.

The different effects of short and long-term risk are consistent with how US policy-makers weigh their options to thwart aspirants. Policy-makers' calls for preventive war are most persuasive when they believe proliferation is imminent. When policy-makers believe that an aspirant will take decades to discover nuclear weapons, it is difficult to make the case for war. For example, American hawks in the late 1940s could not find support for preventive war against the Soviet Union because it was widely believed that the Soviets were a decade away from discovering nuclear weapon. It was not that the Americans did not fear a nuclear Soviet Union. Rather, they thought it was a problem for a later date.

These different effects are also conform with how policy-makers understand the effects of using containment to kick the can down the road. The direct effect of containment (k) delays A's successful nuclear research. Once B enacts containment it takes A longer to discover nuclear weapons. But B knows that containment decreases the chance that A will acquire nuclear in the short-term and factors that into her decision to threaten war. It follows that B's credible threat of preventive war is also delayed because A's chance of a successful nuclear investment is also delayed.

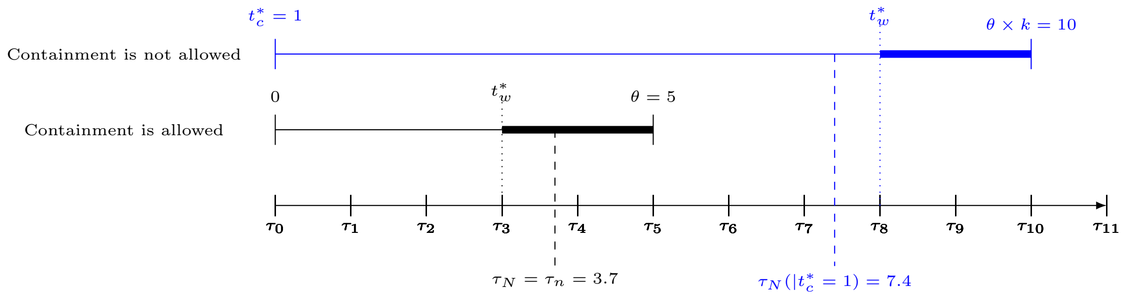
[20] These equations are simpler because I assume that τ_n is drawn uniformly. But changing this distribution has no effect on my argument, it just complicates the conditions.

What the informal literature has not yet grappled with, is how these competing effects balance out. Does containment delay A's nuclear program by more or less than it delays B's credible threat of war? How does this influence the long-run risk of proliferation?

> **Proposition 3.1** *For any set of parameters where we observe containment on the path in a SPNE, the long-run risk of nuclear proliferation is larger than it would have been given A and B's equilibrium strategies for the same set of parameters if B was not allowed to choose containment.*

See Appendix B.2. I use Figure 5, which builds on the same values as Figure 4, to explain why containment always raises the long-run risk of proliferation. In Figure 5, the blue line (thick+thin parts) depicts equilibrium behavior in the game where containment is possible. The black line depicts equilibrium behavior in the counter-factual case where B cannot enact containment.

Figure 5: Why containment increases the risk of proliferation



Each of the two lines represents a unique equilibrium in the game given $R = .001$, $p_0 = .3$, $\Delta = .3$, $w = .2$, $c = .05$, $\theta = 5$, $k = 2$. The bottom line is the counter-factual game where containment is not allowed. The top line is the main game where containment is allowed and occurs on the equilibrium path following A's first nuclear research attempt. t_w^* marks the period where B's threat of major war is credible. The thick lines represent all the draws of τ_n that do not discover nuclear weapons. The thin line represents all the types that will discover nuclear weapons. The dashed lines emphasize a specific draw $\tau_n = 3.7$.

In the example, I assume that Nature draws $\tau_n = 3.7$ from a uniform distribution $[0, 5]$. In the counter-factual game (containment is impossible) if A invests for four periods and B does not fight a preventive war then A acquires nuclear weapons. In the model where containment is allowed, B enacts containment following A's first-period investment. This stretches the number of investments that A must make to discover nuclear weapons by k such that $\tau_N = k \times \tau_n = 7.4$. If A invests for eight periods and B does not fight a preventive war then A acquires nuclear weapons. I mark the period t^*_w that A's investment surpasses B's risk threshold $1-\lambda^*_w$. At t^*_w, B can credibly threaten war if A invests in nuclear research. If $\tau_N < t^*_w$ then A discovers nuclear weapons given equilibrium strategies.

The thick parts of each line represent the distance between when B can credibly threaten war and deter A from future investment (t^*_w); and the number of periods that remain before A will acquire nuclear weapons with certainty (θ in the case with no containment, θk in the case with containment). The first thing to notice is that the thick parts of each lines are the same length. This is not a coincidence. We already showed that B can only threaten war when the short-term risk of proliferation surpasses B's risk threshold $(1 - \lambda^*_w)$. This risk is based on the number of remaining proliferation attempts, not the number of attempts that have passed, or whether B chose containment along the way. All that matters is that the risk of proliferation in the next period was sufficiently high.

The second thing to notice is that containment always extends the time it takes for B to credibly threaten preventive war. In the plot, the thin part of the blue line is longer than the thin part of the black line. This is also not a coincidence. When B enacts containment, each period carries a smaller short-term risk of proliferation. As a result, it takes longer for A's investments to surpass the critical risk threshold defined in equation 18.

The long-run risk of proliferation is simply the thin part of the line over the total line length. When containment is impossible the long-run risk is: $\frac{t^*_w}{\theta}$. In the counter-factual case it is $\frac{t^*_w}{\theta k}$.[21] Clearly, containment extends both the total line length (the number of investments until $1 - \lambda_t = 1$), and the time it takes for B to enact preventive war. However, it disproportionately extends the time until preventive war. This follows because B sets t^*_w based on a fixed risk preventive war $1-\lambda^*_w$ that B bases on the number of periods remaining.

[21] The value θk follows because B set containment in the first period. If B set containment later, this value would be adjusted based on equation 18.

If containment increases the long-run risk of proliferation then why would B ever choose it? In a perfect world, B would understand these long-term risks and avoid containment to reduce the chance of proliferation. However, B faces short-term incentives to choose containment precisely because containment delays how long it will take A to discover nuclear weapons. When containment is cheap and effective, B prefers to capitalize on this short-term benefit, even if it raises the long-run risk of proliferation.

3.3 Illustration: Containment undermines war in US-North Korean relations (1994-2003)

The perverse effects of containment follow because containment undermines the credible threat of preventive war. In the dynamic model, B delays war by choosing containment even as A increases the chance that his program will succeed. If A thought his nuclear investment would trigger war, he would stop. However, A anticipates containment not war. This creates opportunities for A to keep investing. Containment buys A more time to develop his nuclear program leading to an overall greater probability of proliferation.

US policy towards North Korea under the Clinton Administration provides some evidence that containment undermines the credible threat of war in a dynamic setting; raising the risk of proliferation. In April 1993, the International Atomic Energy Agency[22] (IAEA) reported that North Korea had secretly advanced its nuclear weapons program by attempting to reprocess spent fuel rods to create a large volume of highly enriched Uranium. This processing brought North Korea to the brink of discovering a workable nuclear device.[23]

This news alarmed American policy-makers leading to calls for military intervention against North Korea. President Clinton warned that if North Korea developed, or used, atomic weapons that he "would quickly and overwhelmingly retaliate... It would mean the

[22] The IAEA is the International Organization responsible for monitoring nuclear research and reporting weapons activities.

[23] This is not the last step of the process. Even with a device, North Korea would need to develop delivery systems and miniaturize their device to use it as a weapon.

end of their country as they know it." He then stated that "North Korea is just one of many renegade nations that would like to have nuclear weapons and be unaccountable for them, and we can't let it happen.[24]" Foreign-policy focused lawmakers from both major parties supported calls for intervention because they preferred to use force, rather than live with a nuclear North Korea. As a result, "the U.S. efforts to stop North Korea from developing nuclear weapons came close to a war that could have killed as many as a million people on the Korean peninsula."

Despite many calls for intervention between 1993 and 1994, intervention did not come. Instead, the US and North Korea signed the Agreed Framework (1994), which delayed but did not stop North Korea's nuclear program. Under the Agreed Framework North Korea agreed to freeze graphite-moderated nuclear reactors, and remain a party to the NPT.[25] In exchange, the US would provide North Korea with light water reactors and economic inducements.

North Korea's obligations under the Agreed Framework hindered the North Korean nuclear weapons program, reducing North Korea's chance of imminent proliferation.[26] But it was only a short-term fix. North Korea kept its nuclear technologies in tact, and could re-start its program unilaterally.

The US foreign policy elite were concerned that the Agreed Framework would not stop North Korea's nuclear program for long. Many feared that North Korea would continue its program in secret. As Senator John McCain put it, "North Korea's threat to reprocess its 8000 spent fuel rods... represents, by my reckoning, the tenth time Pyongyang reneged on a commitment to the United States.[27]" Even the Clinton Administration admitted to "serious concerns about possible continuing nuclear weapons-related work in the DPRK....[including the] development, testing, deployment, and export by the DPRK of ballistic missiles of

[24] Williams (993)

[25] The Agreement also required North Korea to eventually dismantle its heavy water re-actors—but only after the US had built them light water reactors. This would take years.

[26] To be clear, the agreement also lifted economic sanctions and provided economic inducements. But these inducements were carefully crafted so as not to enhance the North Korean nuclear program.

[27] McCain (1995)

increasing range, including those potentially capable of reaching the territory of the United States.[28]" These positions demonstrate that many feared that the Agreed Framework would only hinder but not stop North Korea's progress towards the bomb.

These concerns turned out to be correct. In 1998, the North Koreans tested rockets,[29] and accepted information from Pakistan to overcome technological challenges North Korea faced in understanding the uranium enrichment process. In 2000, the Clinton Administration learned that North Korea had built a secret uranium enrichment facility inside Mount Chonma. Then, according to a CIA report to Congress, North Korea attempted in late 2001 to acquire "centrifuge related materials in large quantities to support a uranium enrichment program.[30]" In light of these developments, the US imposed increasingly tough sanctions on North Korea and took extreme steps to slow down its nuclear program (Niksch, 2003). In the end, the sanctions were not enough and North Korea acquired nuclear weapons.

Given that these risks were well understood, it is surprising that Clinton did not follow through with his threats of force in 1993, and instead opted for the Agreed Framework. It is even more surprising that Clinton did not resort to war in 1998 once he discovered that North Korea had developed secret enrichment facilities and was slowly working towards the bomb. Instead, Clinton opted for tougher and tougher containment measures to further delay North Korea's realization of a nuclear weapon.

Evidence suggests that a major factor in Clinton's calculus was the immediate cost of war versus the effectiveness of delaying North Korea's program. In March 1994, Clinton reviewed war plans for a full-scale 140,000 solider-strong invasion of North Korea. Then commander in Korea, Gen. Gary Luck, told Clinton that the US could successfully conquer North Korea within six months but it would cost "One million killed and one trillion spent." The president replied, "No one told me that before." Following this conversation, Clinton instructed the State Department to make further concessions to the North Koreans in the

[28] Perry (1999)

[29] which verified their ballistic missile capabilities

[30] Niksch (2003)

hopes of stalling their nuclear program.[31]

Ashton Carter (who served as Assistant Secretary of Defense for International Security Policy from 1993-1996 and senior advisor on the Clinton administration's North Korea policy review) recounted how the Clinton Administration thought about trade-offs and priorities in 1994. In Carter's view, "It is such a disaster for our security in many ways to allow North Korea to go nuclear that we needed to run then... substantial risks to avoid the greater danger of a nuclear North Korea... [However,] We reckoned there would be many, many tens of thousands of deaths: American, South Korean, North Korean, combatant, non-combatant... God forbid that kind of war ever starts on the Korean Peninsula. The loss of life is horrific." As a result, of these trade-offs, Clinton took an opportunity to "*freeze* North Korea's plutonium program[32]" and stall their overall nuclear development, rather than end the program through war.

This evidence conforms to the predictions of my model. The Clinton Administration weighed three unpalatable choices. It could do nothing and live with an imminent nuclear North Korea, fight at enormous cost, or try to stall the North Korean program at low cost. As Carter makes clear, the US did not want to live with a nuclear-armed North Korea and they were willing to fight to prevent it if North Korea's nuclear program was on the verge of success. However, the Administration preferred to stall North Korea's program and delay they cost of war, rather than accept the enormous cost of war straight away. While stopping North Korea was the goal, it was better to kick the can down the road then pay the costs of intervention. In this way, the effectiveness of containment facilitated the US to delay more extreme measures because it reduced the short-term risk of proliferation.

There is some evidence that North Korea thought about the credibility of US preventive war, took steps to undermine the credibility of that threat, and altered their rate of proliferation in response.[33] This has led some policymakers who interacted with North Korean

[31] For a good discussion of the exchange see Mosettig (2014).

[32] Quotes from Carter (2003)

[33] Unfortunately, owing to the secrecy of the Regime, I could not find direct evidence of North Korea's decision-making.

negotiators to claim that North Korea has relied on “the ability of sanctions to avert war.[34]”

Some of this evidence lies in the timing of North Korean conventional and nuclear programs in the context of the end of the Cold War. Through the early 1980s, North Korea advanced their nuclear program under the Soviet umbrella. They broke ground on the Yongbyon Reactor in 1980 and completed a reprocessing plant on that facility in 1984. When the Cold War came to a close North Korea could no longer rely on the Soviets to deter military intervention. As a result, they halted their nuclear programs and turned their attention to developing their conventional forces(Sun Lee, 2006; Choi, 1985). In the mid-to-late 1980s, the North Koreans underwent a period of conventional military modernization. They did not modernize all their forces. Rather they focused on forces they could use to inflict enormous damage on Seoul in the event that they were attacked. Analysts reason that North Korea chose this force posture to raise the cost of a military intervention against them (Anderson, 2017; Michishita, 2006; Sun Lee, 2006; Choi, 1985). It was only once this modernization effort was complete did they start their enrichment programs in 1993.

North Korea’s bizarre negotiating strategy during 1993-1994 suggests that they did not fear war at that point and this allowed them to negotiate a settlement where they did not give up their progress towards the bomb. At the start of the 1993-1994 negotiations American negotiators demanded that the North Koreans immediately return to the NPT, dismantle their re-actors and allow for comprehensive inspections.[35] Rather than concede, the North Koreans demanded that the US build them light water re-actors, reduce their military presence in the Korean Peninsular, and pledge a policy of balance towards South and North Korea (Quinones, 1993). They refused to dismantle any facilities until these demands were met. During in-person negotiations, the North Koreans responded to US questions by quoting passages from Gone With The Wind, rather than engage questions presented by the US delegates (Mosettig, 2014). In the end, the Agreed Framework that they and the United States signed looks much closer to North Korea’s initial position than what the

[34] Harrell and Zarate (2018)

[35] The American negotiators offered economic inducements to compensate the North Koreans.

United States had wanted. Such an outcome seems unlikely if North Koreans thought that US threats of war were serious. Typically, a state that fears military intervention does not bargain so hard.[36]

It is also clear that North Koreans anticipated sanctions and other punitive measures and were undeterred by them. After all, North Korea lived for decades under increasingly tighter US-led global sanctions and continued to develop nuclear weapons. If sanctions could have deterred, then North Korea would have terminated its program once the sanctions were put into place.

The North Korean Regime's observable behavior suggests that it took steps to make war unattractive, and moderated its progress towards the bomb when US threats seemed the largest. The Kim Regime also negotiated fiercely following the United States' public threats of war, rather than make concessions.

4 Conclusion

Cases of deterrence are difficult to observe. This led some analysts to argue that containment has positive effects that we have overlooked. I have shown that hidden effects of containment cuts two ways: containment undermines the credible threat of war causing deterrence to fail. Through formal analysis I have shown that these negative effects arise in the worst possible conditions: against potential aspirants who are industrialized and whose preferences diverge from the United States. Further, in dynamic settings they raise the long-run risk of proliferation.

These differences lead to dramatically different policy implications. Policy-makers that focus on the secret success of containment might invest in coalitions of anti-nuclear sanctioning states to make containment as cheap and effective as possible. This would increase the

[36] North Korean behavior was similar in 1999 when US Secretary of Defense William Perry visited North Korea. Perry first met with a North Korean General who told him "This meeting was not my idea... I was directed to meet with you. I don't think we should even be talking about giving up nuclear weapons." It seems highly unusual that a delegation that seriously feared war would so blatantly create conflict with their rival.

conditions under which the US could credibly threaten containment. My theory shows that making containment attractive may undermine the credible threat of war. Policy-makers should exercise caution before they make containment the most credible tool in their kit. Rather, they should lower the cost of containment when they can also lower the cost of preventive war. Strangely, it may be in their interest to make containment more difficult (and therefore less credible) against their worst adversaries.

Future research might integrate inducements and institutional protocols into a model that includes containment to understand if these positive measures for ending proliferation are enhanced or undermined by sanctions.

References

Anderson, N. D. (2017, nov). Explaining North Korea's Nuclear Ambitions: Power and Position on the Korean Peninsula. *Australian Journal of International Affairs 71*(6), 621–641.

Bailey, M. A., A. Strezhnev, and E. Voeten (2017, feb). Estimating Dynamic State Preferences from United Nations Voting Data. *Journal of Conflict Resolution 61*(2), 430–456.

Baldwin, D. A. (2000, jan). The Sanctions Debate and the Logic of Choice. *International Security 24*(3), 80–107.

Banks, A. S. and K. A. Wilson (2014). Cross-National Time-Series Data Archive.

Bas, M. A. and A. J. Coe (2016). A Dynamic Theory of Nuclear Proliferation and Preventive War. *International Organization 70*(4), 655.

Bas, M. A. and A. J. Coe (2018, sep). Give Peace a (Second) Chance: A Theory of Nonproliferation Deals. *International Studies Quarterly 62*(3), 606–617.

Carnegie, A. and A. Carson (2018, may). The Spotlight's Harsh Glare: Rethinking Publicity and International Order. *International Organization 72*(03), 627–657.

Carter, A. (2003, mar). Interview: Ashton Carter. *Frontline*.

Choi, Y. (1985, mar). The North Korean Military Buildup and Its Impact on North Korean Military Strategy in the 1980s. *Asian Survey 25*(3), 341–355.

Coe, A. J. (2018, oct). Containing Rogues: A Theory of Asymmetric Arming. *The Journal of Politics 80*(4), 1197–1210.

Coe, A. J. and J. Vaynman (2015, oct). Collusion and the Nuclear Nonproliferation Regime. *The Journal of Politics 77*(4), 983–997.

Debs, A. and N. P. Monteiro (2014, jan). Known Unknowns: Power Shifts, Uncertainty, and War. *International Organization 68*(01), 1–31.

Diamond, P. A. (1982, apr). Wage Determination and Efficiency in Search Equilibrium. *The Review of Economic Studies 49*(2), 217.

Drezner, D. W. (2011, mar). Sanctions Sometimes Smart: Targeted Sanctions in Theory and Practice. *International Studies Review 13*(1), 96–108.

Feaver, P. D. and E. M. S. Niou (1996, jun). Managing Nuclear Proliferation: Condemn, Strike, or Assist? *International Studies Quarterly 40*(2), 209.

Gartzke, E. and M. Kroenig (2014, apr). Nuclear Posture, Nonproliferation Policy, and the Spread of Nuclear Weapons. *Journal of Conflict Resolution 58*(3), 395–401.

Harrell, P. and J. Zarate (2018). How to Successfully Sanction North Korea. *Foreign Affairs*.

Hersman, R. K. C. and R. Peters (2006, nov). Nuclear U-Turns. *The Nonproliferation Review 13*(3), 539–553.

Izewicz, P. (2017, apr). Iran's Ballistic Missile Programme. *Non-proliferation papers 57.*

Jackson, M. O. (2009, dec). Strategic Militarization, Deterrence and Wars. *Quarterly Journal of Political Science 4*(4), 279–313.

Jo, D.-J. and E. Gartzke (2007, feb). Determinants of Nuclear Weapons Proliferation. *Journal of Conflict Resolution 51*(1), 167–194.

Kahl, C. H. (2012, mar). Not Time to Attack Iran. *Foreign Affairs*.

Krainin, C. (2017, jan). Preventive War as a Result of Long-Term Shifts in Power. *Political Science Research and Methods 5*(01), 103–121.

Kreps, S. E. and M. Fuhrmann (2011, apr). Attacking the Atom: Does Bombing Nuclear Facilities Affect Proliferation? *Journal of Strategic Studies 34*(2), 161–187.

Kroenig, M. (2012, jan). Time to Attack Iran. *Foreign Affairs*.

McCain, J. (1995). Address on the North Korean Agreed Framework.

McCormack, D. and H. Pascoe (2015, dec). Sanctions and Preventive War. *Journal of Conflict Resolution*, 0022002715620471–.

Mehta, R. N. and R. E. Whitlark (2017, sep). The Benefits and Burdens of Nuclear Latency. *International Studies Quarterly 61*(3), 517–528.

Michishita, N. (2006, dec). Coercing to reconcile: North Korea's response to US hegemony'. *Journal of Strategic Studies 29*(6), 1015–1040.

Miller, N. L. (2014, sep). The Secret Success of Nonproliferation Sanctions. *International Organization 68*(04), 913–944.

Monteiro, N. P. and A. Debs (2014, oct). The Strategic Logic of Nuclear Proliferation. *International Security 39*(2), 7–51.

Mosettig, M. (2014, oct). 20 years later, commemorating a war averted.

Narang, N. and R. N. Mehta (2017, sep). The Unforeseen Consequences of Extended Deterrence. *Journal of Conflict Resolution*, 002200271772902.

Niksch, L. (2003). North Korea's Nuclear Weapons Program. Technical report, Congressional Research Service, Washington D.C.

Paul, T. V., P. M. Morgan, and J. J. Wirtz (2009). *Complex Deterrence: Strategy in the Global Age.* Chicago: University of Chicago Press.

Perry, W. (1999). Review of United States Policy Toward North Korea: Findings and Recommendations. Technical report, Office of the North Korea Policy Coordinator, United States Department of State, Washington D.C.

Powell, R. (1999). *In the Shadow of Power: States and Strategies in International Politics.* Princeton, N.J: Princeton University Press.

Powell, R. (2006, jan). War as a Commitment Problem. *International Organization 60*(01).

Quinones, C. K. (1993). Resolution of the Nuclear Issue: Elements to be Considered.

Reiter, D. (2005, jul). PREVENTIVE ATTACKS AGAINST NUCLEAR PROGRAMS AND THE SUCCESS AT OSIRAQ. *The Nonproliferation Review 12*(2), 355–371.

Rogerson, R., R. Shimer, and R. Wright (2005, nov). Search-Theoretic Models of the Labor Market: A Survey. *Journal of Economic Literature 43*(4), 959–988.

Smith, B. C. and W. Spaniel (2018, jan). Introducing v-CLEAR: a latent variable approach to measuring nuclear proficiency. *Conflict Management and Peace Science*, 073889421774161.

Spaniel, W. and P. Bils (2017). Policy Bargaining and International Conflict. *Journal of Theoretical Politics*.

Spaniel, W. and B. C. Smith (2015, mar). Sanctions, Uncertainty, and Leader Tenure. *International Studies Quarterly 59*(4), n/a–n/a.

Sun Lee, D. (2006, may). US Preventive War against North Korea. *Asian Security 2*(1), 1–23.

Volpe, T. A. (2017, jul). Atomic Leverage: Compellence with Nuclear Latency. *Security Studies 26*(3), 517–544.

Williams, D. (993, jul). U.S. Warns N. Korea on Nuclear Weapons.

A Solution to simple model reported in the paper.

A.1 Lemma 2.2: A's second period minimum demand

Period 2 is the terminal period. Thus, A has no value from investing in nuclear research. Suppose she did, she would pay a cost R in the second period, but would not be able to benefit from a third-period lottery. By the same logic, B cannot profit from containment. Thus, I focus on B's incentives to make an offer. By definition, A accepts an offer q_2^*. Clearly, B won't offer more because B values the pie. Suppose B offers less than q_2^*, A rejects it in favor of war. But B's utility from war is less than his utility from an offer q_2^* that is accepted. It follows that B's best offer is q_2^*. There are no more off-path deviations to consider.

A.2 Total expected utilities given different first period strategies

I'll now write out each player's first period expected utility under the assumption that B will offer A q_0 in the first period and that both players anticipate second period pay-offs written in Lemma 2.2, and the associated probabilities with each strategy.

Suppose A does not invest in the first period, then each player's total expected utility in the first period is:

$$U_1^A(s^A(NI), s^B(q_1 = q_0)) : (1 - \pi(1 - p_0) - w)(1 + \delta) \tag{19}$$

$$U_1^B(s^A(NI), s^B(q_1 = q_0)) : (1 - p_0\pi + w)(1 + \delta) \tag{20}$$

Notice A's utility is identical to her war pay-off. It follows that A will accept this offer. Suppose A invests in nuclear research in the first period and B makes peaceful offers, then each player's first period expected utility is:

$$U_1^A(s^A(I), s^B(q_1 = q_0, q_2^*, c_1 = 0)) : (1 - \pi(1 - p_0) - w)(1 + \delta) + \delta\Delta\pi(1 - \lambda) - R \tag{21}$$

$$U_1^B(s^A(I), s^B(q_1 = q_0, q_2^*, c_1 = 0)) : (1 - p_0\pi + w)(1 + \delta) - \delta\Delta\pi(1 - \lambda) \tag{22}$$

The additional $\Delta\pi(1 - \lambda)$ in A's utility (and taken from B's) follows from A's $1 - \lambda$ chance of getting nuclear weapons in the second period. Of course, this comes at a cost R.

A's utility from accepting this offer is larger than her war pay-off because at the point where A chooses to accept the offer or not, she has already invested in nuclear research. It follows that A will accept an offer q_0 assuming that she has already invested. Since B cannot make a smaller offer, B will make this offer holding constant her containment and war choices in the first period.

Suppose A invests in the first round and B contains, then expected total utilities are:

$$U_1^A(s^A(I, NI), s^B(c, q_2^*)) : (1 - \pi(1 - p_1) - w)(1 + \delta) + \delta\Delta\pi(1 - \lambda_c) - R \tag{23}$$

$$U_1^B(s^A(I, NI), s^B(c, q_2^*)) : (1 - p_1\pi + w)(1 + \delta) - \delta\Delta\pi(1 - \lambda_c) - c \tag{24}$$

If A invests in the first round and B chooses war:

$$U_1^A(s^A(I, w), s^B(w, w)) : (1 - \pi(1 - p_1) - w)(1 + \delta) - R \tag{25}$$

$$U_1^B(s^A(I, w), s^B(w, w)) : (1 - p_1\pi - w)(1 + \delta) \tag{26}$$

There are no more outcomes to consider. This covers all the strategy pairs.

Lemma A.1 *War cannot appear on the equilibrium path in any SPNE*

We've shown that war does not appear on the path in the second period. So I focus on each player's first period incentives. If A does not invest in the first period, then B's best reply is to make two peaceful offers $q_1 = q_2 = q_0$. B's value from these offers is always larger than B's utility from war. If A invests, B can prefer war to a peaceful offer. Yet A never prefers to invest in nuclear research and face war because A must pay an additional R but gets no benefit from it. It follows that if A's investment induces war.

A.3 Lemma 2.3: A does not invest.

The best A can do from investment is the condition where B plays $s^B(q_1^*, q_2^*)$. The inequality in Lemma 2.3 follows from comparing A's Utility from: $U_1^A(s^A(NI, NI), s^B(q_1^*, q_2^*)) > U_1^A(s^A(I, NI), s^B(q_1^*, q_2^*))$.

A.4 Lemma 2.4: B does not compete.

Inequality 9 follows from comparing: $U_1^B(s^A(I, NI), s^B(q_1^*, q_2^*)) > U_1^B(s^A(I, w), s^B(w, w))$. Inequality 10 follows from comparing $U_1^B(s^A(I, NI), s^B(q_1^*, q_2^*)) > U_1^B(s^A(I, NI), s^B(c, q_2^*))$. Since B only has three options, it follows that when these inequalities are jointly satisfied peaceful offers are better than any form of competition. However, if either fails, then at least one form of competition is better.

A.5 Lemma 2.5: Sanctions deter

I focus on strategic choices in period 1 given total expected utilities defined above. Following Lemma 2.4, B's best reply to A's investment is containment if inequality 10 is not satisfied and:

$$U_1^B(s^A(I, NI), s^B(c, q_2^*)) > U_1^B(s^A(I, w), s^B(w, w)) \equiv 2w(1+\delta) > \Delta\pi\delta(1-\lambda_c) + c \quad (27)$$

When these conditions are jointly satisfied, B will respond to A's investment with containment. Anticipating this, A chooses to invest or not. A does not invest if $U_1^A(s^A(I, NI), s^B(c, q_2^*)) < U_1^A(s^A(NI, NI), s^B(q_1^*, q_2^*))$ which is satisfied when $r > \Delta\pi\delta(1-\lambda_c)$ (as reported in Lemma 2.5). By my definition of deterrence, it must be that A would have invested if B made a peaceful offer. This solves for the $\Delta\pi\delta(1-\lambda) > r$ also reported in Lemma 2.5.

All that's left to do is demonstrate that these inequalities can be jointly satisfied. Notice that c only appears in inequalities 10 and 27 (which are easily solved for c) and r only appears in inequality 11 (and can be solved for positive values). Since c and r vary independently, I can find values of these parameters such that both inequalities hold. This completes the proof.

A.6 Lemma 2.6: war deters

I focus on strategic choices in period 1 given total expected utilities defined above. Following Lemma 2.4, B's best reply to A's investment is war if inequality 10 is satisfied and inequality 27 is not.

When these conditions are jointly satisfied, B will respond to A's investment with war. Anticipating this, A chooses to invest or not. We've already shown that A does not invest if A will face war. Thus, A does not invest. By my definition of deterrence, it must be that A would have invested if B made a peaceful offer. This solves for the $\Delta\pi\delta(1-\lambda) > r$.

By the same logic written for section A.5, these inequalities can be jointly satisfied (here c must be sufficiently high). This completes the proof.

A.7 Proposition 2.7: Containment undermines deterrence

The proof focuses on the two-competition region where B prefers both forms of competition to a peaceful offer in the first period. Thus, it must be that inequality 9 and 10 are jointly not satisfied. Further, it must be that in the first period, B prefers the risk of proliferation under containment than fighting a war in the first round: $U_1^B(s^A(I, w), s^B(w, w)) < U_1^B(s^A(I, NI), s^B(c, q_2^*))$ which solves for $2w(1+\delta) > \Delta\pi\delta(1-\lambda_c) + c$ as reported in Proposition 2.7. When these conditions are jointly satisfied B responds to A's nuclear investment to containment.

A's first round decision depends on A's expected value for investment given that A will face containment. In section A.5 I defined the condition where A will invest anticipating containment and have re-reported this result in Proposition 2.7.

By the same logic written for section A.5, these inequalities can be jointly satisfied (here I must be sufficiently low). This completes the proof.

This satisfies all the elements for deterrence to fail. A would have invested if B made a peaceful offer. B's threat of competition was credible but A invested anyway.

Turning to corollary 2.8. I start by making the informal description a little more precise. Define the full set of parameters X as those that can support equilibrium behavior described in Proposition 2.7, and $x \in X$ as a specific set of parameters. Then $s^A(x), s^B(x)$ describes the equilibrium strategies in the game that produce Proposition 2.7. Define $\bar{S}^A(x) \times \bar{S}^B(x)$ as the set of strategy pairs that form an equilibrium in the counter-factual game given parameters x. Where $\bar{s}^A(x) \times \bar{s}^B(x)$ is a specific strategy pair that forms an equilibrium. To re-state the corollary, for any set of parameters $x \in X$ any pair of strategies that form a PBE in the counter-factual world: $\bar{s}^A(x) \times \bar{s}^B(x)$ include the following choices: In both periods, A does not invest in his nuclear program. If A did invest in the first period, B would reply with war.[37]

The proof is identical to the proof where war deters A reported in Section A.6.

[37] The reason I must consider a variety of strategy pairs is that B could make any offer $q_1^* \leq q_0$ in equilibrium and A would just reject it. This opens up the possibility for multiple equilibrium. However, these equilibrium produce identical expected utilities for both players.

B Solution to the model with shifting risk of proliferation

First, I show that the game must converge to three stable sub-games in any SPNE. I use these sub-games to rule out possible sub-games and focus the analysis. I also define players' expected utilities once they enter these sub-games. Second, I prove some facts about the result, which allow me to focus on a specific sub-set of strategies on the equilibrium path. Third, I search for equilibrium conditions.

B.1 Stable Subgames

I say players enter a stable sub-game when they play the exact same pure strategy in that period and all subsequent periods leading to constant expected utilities across periods. More precisely, suppose a period t where both players' equilibrium strategy is $s_t^A(x), s_t^B(y)$ yielding a one round payoff $U_t^A(X), U_t^B(Y)$ and total expected utility $\sum_{i=t}^{\infty} \delta^{i-t} U_i^A(X)$, then a sub-game is stable if for any $t' > t$ $s_{t'}^A(x) = s_t^A(x)$ $s_{t'}^B(y) = s_t^B(y)$, $EU_{t'}^A(X) = EU_t^A(X)$, $EU_t^B(Y) = EU_{t'}^B(Y)$.

The first stable sub-game is war. If either player chooses war in any period, they are both forced into an absorbing sub-game for all future rounds.

Lemma B.1 *Suppose war occurs in period t, then in every period $t' > t$ A does not invest in her nuclear research, B does not contain. Players' total expected utilities in period t are:*

$$EU_t^B(war) : \frac{1 - p_t - w}{1 - \delta} \tag{28}$$

$$EU_t^A(war, NI) : \frac{p_t - w}{1 - \delta} \tag{29}$$

These expected values for war only depend on t in so far as A has acquired nuclear weapons. If $t < t_N$ than $p_t = p_0$, otherwise $p_t = p_0 + \Delta$. The proof is obvious. A has no benefit from future nuclear investment. B has no benefit from containment. It follows that they do not play these costly strategies in any period.

Remark War cannot appear on the equilibrium path in a SPNE.

The proof is identical to the equivalent Lemma in the baseline game. When B can threaten war, A prefers not to invest. Thus, war cannot appear. Although war does not appear on the path, it defines two points. First, it defines S's minimum demand from an offer. Second, it defines the point where S is deterred from nuclear research.

Define, $q_N^* = p_0 + \Delta - w$ as the offer that leaves A indifferent with fighting given that A has discovered nuclear weapons. The second stable state arises as soon as A discovers nuclear weapons.

Lemma B.2 *Suppose A discovers nuclear weapons at the onset of round t, then the subgame in round t is stable. In period $t' \geq t$, A does not invest in nuclear research, B offers A $q_N^* = p_0 + \Delta - w$, A accepts. Leading to expected utilities in any period:*

$$EU^B(t > t_N) : \frac{1 - p_0 - \Delta - w}{1 - \delta} \tag{30}$$

$$EU_t^A(t > t_N) : \frac{p_0 + \Delta - w}{1 - \delta} \tag{31}$$

Once A acquires nuclear weapons, A never invests in nuclear research because it costs her R but cannot provide any additional benefit. Similarly, B cannot profit from containment. Since power is stationary then A's minimum demand in each period is equivalent to her war payoff.

Define, q_{0N}^* as the offer that leaves A indifferent with fighting a war under the assumption that A has not discovered nuclear weapons, that A did not invest in her nuclear program, and that A will not invest in nuclear research in any subsequent period. Given this strategy profile, the offer that leaves A indifferent with fighting a war is: $\sum_{i=0}^{\infty} \delta^i q_{0N}^* = EU_t^A(War)$. This implies, $q_{0N}^* = p_0 - w$.

The third stable state arises when A's best reply is not to invest in her nuclear program in an arbitrary period before A acquires nuclear weapons.

Lemma B.3 *Suppose a Sub-game Perfect Nash Equilibrium where on the path A chooses not to invest in her nuclear program at t and A has not yet discovered nuclear weapons, then the subgame starting in period t is stable. In period t and all following periods, A does not invest in her nuclear program, B offers A* $q^*_{0N} = p_0 - w = q_0$*, A accepts. This leads to expected utilities:*

$$EU^B(t < t_N|NI) : \frac{1 - p_0 - w}{1 - \delta} \tag{32}$$

$$EU^A_t(t < t_N|NI) : \frac{p_0 - w}{1 - \delta} \tag{33}$$

The reason that this is stable is that if A's best response is not to invest in some arbitrary period t, then the strategic setting is identical in period $t + 1$. That is, $\tau_t = \tau_{t+1}$. It follows that A does not invest at $t+1$, because A did not invest at t and the strategic incentives are the same. By induction, it must be that A never chooses to invest in subsequent rounds.

Lemma B.3 implies that in any equilibrium where A invests in his nuclear program at t, A must have invested in his nuclear program in all prior periods. Further, if A does not invest in the first period, then A never invests. It immediately follows that:

Remark (a) If A invests in her nuclear program, she does so in the first period and then for a finite number of sequential periods. (b) If on the path, A does not invest in her nuclear program at t, she never invests in her nuclear program again.

Remark B.1 follows directly from the sub-game in Lemma B.3. Once A stops investing on the path, she never invests again even if A can acquire nuclear weapons. As a result, if A did not invest in the first period, she never invests.

This result effectively turns A's investment choice into a stopping rule. That is, A sets a period t_R where A chooses to stop nuclear research. On the equilibrium path, A can start investing in period 1, and continue until either A acquires nuclear weapons (leading to subgame described in Lemma B.2), or a fixed point where A stops investing (leading to sub-game described in Lemma B.3). Alternatively, A can never invests in nuclear research and Lemma B.3 describes equilibrium behavior in every period.

Let t_R be an integer that reports the number of periods that A will invest in nuclear research if he is not stopped from doing so (through war), and he has not discovered nuclear weapons.[38] B's strategy can also be described as a decision to enact war and containment at a specific period conditional on A's stopping rule. Let t_c, t_w be integers that identify the period in which B enacts war /containment if A has invested in that round.

A's equilibrium strategy s^A is defined by:

- An investment rule $t^*_R|s^B(t^*_c, t^*_w))$ that defines the number of rounds A will invest before he stops investing.
- An accept/reject for status quo/war rule conditional on B's offer.

A set's these rules to maximize his expected utility conditional on $s^B()$.

B's equilibrium strategy s^B is defined by:

- A contain rule $t^*_c|t^*_R$ that defines the number of nuclear investments that B will tolerate before she contains A.
- A preventive war rule $t^*_w|t^*_R, t^*_c$ that defines the number of nuclear investments that B will tolerate before she fights a preventive war.

B sets these rules to maximize her expected utility conditional on $s^A()$

Proposition B.4 *There exists an equilibrium where we observe containment on the path. When the following inequalities can be solved for a constant t, $1 \leq t < t^*_w$.*

[38] Since A invests in consecutive rounds starting at 1, I need not distinguish between the number of rounds in which A invests and the specific round of the game.

$$\frac{\delta\Delta}{(\theta-t+1)k(1-\delta)^2}\left[k-1-k\delta^{\theta-t-\frac{\delta\Delta-2w}{2w(1-\delta)}}+\delta^{(\theta-t)k-\frac{\delta\Delta-2w}{2w(1-\delta)}}\right]>c \tag{34}$$

$$\frac{\delta\Delta}{k(\theta-t+1)(1-\delta)^2(1-\delta+\frac{\delta}{\theta-t+1})}\left[(1-\delta)(k-1)-\delta^{(\theta-t-1)k+1-\frac{\delta\Delta-2w}{2w(1-\delta)}}+\delta^{(\theta-t)k-\frac{\delta\Delta-2w}{2w(1-\delta)}}\right]>c \tag{35}$$

and

$$\delta\Delta > 2w \tag{36}$$

then I can always find a cost $R^* > 0$ sufficiently small where the following strategies are unique best replies in a Sub-game Perfect Nash Equilibrium. If A has not yet discovered nuclear weapons,

- *A sets an investment rule $t_R^* = t_w^* - 1$. A invests in every period $t \in \{1 : t_w^* - 1\}$ then does not invest in any period $t \geq t_w^*$.*

- *B sets a war rule $t_w^* = (\theta - t_c^*)k + t_c^* - \frac{\delta\Delta-2w}{2w(1-\delta)}$. In any period $t < t_w^*$, B will not choose war even if A invests. However, B will choose war if A invests in her nuclear program in any period $t \geq t_w^*$*

- *B sets a containment rule t_c^* that satisfies $1 \leq t_c^* < t_w^* - 1$. In particular, t_c^* is the smallest t that satisfies inequalities 34 and 35. In any period $t < t_c^*$, B will not choose containment even if A invests. However, B will choose containment if A invests in her nuclear program in any period $t \geq t_c^*$.*

- *A accepts all offers $q_t \geq q_{N0}^* = q_0$ on the path and rejects all smaller offers in favor of war off the path.*

- *B offers $q_t = q_{N0}^*$ in every period unless A discovers nuclear weapons.*

If A has discovered nuclear weapons ($\tau_N | t_c^* < t_w^*$) then in that period and all subsequent periods B offers A his minimum demand ($p + \Delta - w$), which A accepts. A does not invest in his nuclear program.

The reason containment can occur on the path is that B always selects containment at least 2 periods earlier than major war $t_c^* < t_w^* - 1$. However, A only stops her nuclear research in the period right before B chooses preventive war: $t_R^* = t_w^* - 1$. It follows that if $\tau_N > t_c^*$ that we observe containment on the equilibrium path. Otherwise, we observe nuclear proliferation before containment takes place.

I'll show two things. First, conditions 34 and 35 define the thresholds that determine when B's best strategy is to enact containment straight away if B observes A invest in his nuclear program. As soon as these conditions are jointly satisfied, B prefers to contain rather than either wait one period to contain, or never contain. It follows that in the first period (the lowest possible t) that satisfies these conditions, B selects containment.

Second, for any parameters of the game that can satisfy conditions 34 and 35, which produces a t_c^* such that $1 \leq t_c^* < t_w^*$, there must be some values of $R > 0$ sufficiently small that A prefers to invest in every period $t \in \{1 : t_w^* - 1\}$. It follows that A's best reply to B's threat of containment is to keep investing in his military.

First, I'll show B's strategy is incentive compatible holding A's strategy constant. On the path, $t_w^* > t_c^*$. As a result, I first identify when B can credibly threaten war following A's investment: t_w^* for any history of the game that includes sequential investments t_R and assumes that A has not yet discovered nuclear weapons. B can credibly threaten war when B prefers to fight in the current period rather than wait one round to fight: $\frac{1-p+w}{1-\delta} + \delta\lambda\frac{1-p-w}{1-\delta} + \delta(1-\lambda_t)\frac{1-p+w-\Delta}{1-\delta} > \frac{1-p-w}{1-\delta}$. This solves for

$$1 - \lambda_w^* \geq \frac{2w(1-\delta)}{\delta(\Delta - 2w)}. \tag{37}$$

This is B's critical risk threshold for war. When this inequality is not satisfied, B cannot credibly threaten war if A invests. But as soon as it is satisfied, B can credibly threaten war. Define $\tau_w^* = t_w^*$ as the number of investments A must make such that it is the first round that this risk threshold is satisfied: $1 - \lambda_{\tau_w^*} > 1 - \lambda_w^* > 1 - \lambda_{\tau_w^*-1}$. I can locate the number if investments it takes to reach t_w^* by subbing in equation 37.

$$t_w^*|(1 \leq t_c < t_w^*) = (\theta - t_c)k + t_c - \frac{\delta\Delta - 2w}{2w(1-\delta)}, \tag{38}$$

and

$$t_w^*|(t_c = \emptyset) = \theta - \frac{\delta\Delta - 2w}{2w(1-\delta)}. \tag{39}$$

The final thing to check is that B prefers to fight a war, rather than let A get nuclear weapons with certainty. B's risk threshold must satisfy $0 < 1 - \lambda_w^* < 1$. Following equation 37, B prefers war to letting A get nuclear weapons with certainty if $\Delta > 2w$ and $\delta\Delta > 2w$. These are both satisfied when and only when equilibrium condition 36 is satisfied as desired.

To be clear, equations 38 and 39 provide different values for $t_w^*|t_c$. They describe the point at which B can credibly threaten war, given a history of the game t_c, t_R. Working backwards, B's decision to set a containment rule must factor in her expected utility from setting that rule conditional on when B can credibly threaten war, and what it means for A's investment choice, and the probability that A will acquire nuclear weapons.

For B to set containment in some period $t < t_w^*$, it must be that in that period B prefers to enact containment straight away, rather than wait, or never contain A. In any period $t < t_w^*|t_c$, B's utility from never setting containment is:

Starting with the case where B never selects containment on the path, I can write B's utility in period $t < t_w$ as:

$$\begin{aligned} EU_t^B|t_c = \emptyset, t < t_w = 1 - p + w + \frac{\delta(1-\lambda_t)(1-p+w-\Delta)}{1-\delta} + \delta(1-\lambda_t)\left(1-p+w+\frac{\delta(1-p+w-\Delta)}{1-\delta}\right) \\ + \delta(1-\lambda_t)\left((1-p+w)(1+\delta) + \frac{\delta^2(1-p+w-\Delta)}{1-\delta}\right)... \\ + \delta(1-\lambda_t)\left((1-p+w)(1+\delta+\delta^2... + \delta^{t_w^*(|t_c=\emptyset)-t-1}) + \frac{\delta^{t_w^*(|t_c=\emptyset)-t}(1-p+w-\Delta)}{1-\delta}\right) \\ + \left(\delta[1-(t_w^*(|t_c=\emptyset)-t)(1-\lambda_t)]\frac{1-p+w}{1-\delta}\right). \end{aligned} \tag{40}$$

A key element of these equations is that for an arbitrary period t, B's expectation that A will discover nuclear weapons following A's investment in period $t+1$ is the same as B's period t expectation that A will discover nuclear weapons in period $t+2$, $t+3$, and so on (it is λ_t). B's expectation only shifts once A does not discover nuclear at $t+1$. As a result, I can write B's expected utility in any period t using a constant probability: λ_t.

I now describe each term in the equation. The first term (red), is B's first period payoff from not fighting $(1-p+w)$. Since B already knows that A has not acquired nuclear weapons, B's period t payoff is not affected by λ_t. The second term (blue) is weighted by B's expectation that A will acquire nuclear weapons at the beginning of period $t+1$ $(1-\lambda_t)$. This event would trigger a stable-subgame defined in Lemma B.2. So it includes B's payoff (discounted by δ) from the stream of offers B would need to make in that contingency: $\frac{1-p+w-\Delta}{1-\delta}$.

The third term (gray) is weighted by B's expectation that A will not acquire nuclear weapons following the first investment but will acquire them following the second investment. As described above, this is also probability $1-\lambda$. B's value in this contingency is what B gets in period $t+1$ given that A has not yet discovered nuclear weapons $(1-p+w)$ plus what B gets from entering the stable sub game defined in Lemma B.2 when D discovers nuclear weapons in $t+2$: $\frac{\delta(1-p+w-\Delta)}{1-\delta}$.

I then (in black) identify a sequence of expected pay-offs that follow for periods $t+2, t+3, ...$ using the same logic.

The final term (green) is weighted by B's expectation that A will not discover nuclear weapons in the first $t_w - 1$ periods (with probability $[1-(t_w-t)(1-\lambda)]$). If true, then the game enters stable sub-game defined in Lemma B.3 and B benefit from that sub-game is $\frac{1-p+w}{1-\delta}$.

Solving the inequality we get:

$$EU_t^B | t_c = \emptyset, t < t_w = \frac{1-p+w}{1-\delta} - \frac{\Delta(1-\lambda_t)\delta(1-\delta^{t_w^*(|t_c=\emptyset)-t})}{(1-\delta)^2} \tag{41}$$

I now define B's expected total utility in period t under the assumptions that B will select containment in some period $t_c \geq t$, $t_c < t_w^*|t_c$. That is, B could set containment in the current period, or some future period before t_w^*.

$$
\begin{aligned}
EU_t^B|t \leq t_c < t_w = 1 - p + w + \frac{\delta(1-\lambda_t)(1-p+w-\Delta)}{1-\delta} + ... \\
+\delta(1-\lambda_t)\left((1-p+w)(1+\delta+...+\delta^{t_c-t-2}) + \frac{\delta^{t_c-t-1}(1-p+w-\Delta)}{1-\delta}\right) - \delta^{t_c-t}\left[1-(1-\lambda_t)(t_c-t)\right]c \\
+\frac{\delta(1-\lambda_t)}{k}\left((1-p+w)(1+\delta...+\delta^{t_c-t-1}) + \frac{\delta^{t_c-t}(1-p+w-\Delta)}{1-\delta}\right)... \\
+\frac{\delta(1-\lambda_t)}{k}\left((1-p+w)(1+\delta...+\delta^{t_w^*(|t\leq t_c<t_w)-t-1}) + \frac{\delta^{t_w^*(|t\leq t_c<t_w)-t}(1-p+w-\Delta)}{1-\delta}\right) \\
+\left(\delta[1-(t_c-t)(1-\lambda_t)-(t_w^*(|t \leq t_c < t_w)-t_c)(1-\lambda_t)/k]\frac{1-p+w}{1-\delta}\right)
\end{aligned}
\tag{42}
$$

The logic of this equation is similar to the set-up for inequality 41. There are som differences that start in the period $t = t_c - 1$ (colored in red). In that period, B anticipates that she pays a cost c at t_c (colored blue), and accepts a lower probability that A's nuclear investment will succeed (colored green and reflected by reduced probability $\frac{(1-\lambda)}{k}$). The final term (colored orange) reflects B's expectation that A will not discover nuclear weapons in the $t_c - t$ period before containment, and the $t_w^*(|t \leq t_c < t_w) - t_c$ periods between containment and when B can credibly threaten war. Solving the inequality we get:

$$
EU_t^B|t \leq t_c < t_w = \frac{1-p+w}{1-\delta} - \delta^{t_c-t}\left[1-(1-\lambda_t)(t_c-t)\right]c - \frac{\Delta(1-\lambda_t)\delta}{1-\delta}\left[\frac{1-\delta^{t_c-t}}{1-\delta} + \frac{\delta^{t_c-t} - \delta^{t_w^*(|t\leq t_c<t_w)-t}}{k(1-\delta)}\right]
\tag{43}
$$

There is a special case of this equation: B's value from enacting containment straight away $t_c = t$.

$$
EU_t^B|t = t_c < t_w = \frac{1-p+w}{1-\delta} - c - \frac{\Delta(1-\lambda_t)\delta}{1-\delta}\left[\frac{1-\delta^{t_w^*(|t=t_c)-t}}{k(1-\delta)}\right]
\tag{44}
$$

My goal is to find a $t < t_w^*|t_c = \emptyset$ ($t < \theta - \frac{\delta\Delta - 2w}{2w(1-\delta)}$) such that B's expected utility from setting $t_c = t$, is larger than B's utility from either delaying containment or never selecting containment: $EU_t^B(|t = t_c < t_w) > max(EU_t^B(|t < t_c < t_w))$ and also $EU_t^B(|t = t_c < t_w) > EU_t^B(|t_c = \emptyset)$.

To see this can hold, consider B's utility equation 43 as a function of t_c for a fixed t. The middle term captures the cost B pays for enacting containment c. This value is *decreasing* in t_c because delaying containment delays the amount of time until B incurs this cost ($\delta^{t_c - t}$) and also increases the possibility that B will never need to pay it because A will discover nuclear weapons $[1 - (1-\lambda_t)(t_c - t)]$. The final term captures B's expectation that A will discover nuclear weapons sooner rather than later, as a result of delayed containment and the costly consequences of that contingency. This term is *increasing* in t_c because delaying containment longer implies proliferation is likely to come sooner. It follows that for a fixed t, equation 43 as a function of $t_c < t_w^*|t_c$ has at most one critical value for $t < t_c < t_w^*$. Thus, it is either strictly increasing, decreasing, or a convex or concave function.

It follows that if I can find a value of t that allows me to jointly satisfy setting $t = t_c$ such that $EU_t^B(|t_c = t) > EU_t^B(|t_c = t+1)$ and $EU_t^B(|t = t_c) > EU_t^B(|t_c = \emptyset)$, then B's best reply must be containment at some point on the path.

Starting with B's preference for $t_c = t$ over $t_c = \emptyset$ for a fixed t:

$$\frac{\delta\Delta(1-\lambda_t)}{k(1-\delta)^2}\left[k - 1 - k\delta^{t_w^*(|t_c=\emptyset)-t} + \delta^{t_w^*(|t_c=t)-t}\right] > c \tag{45}$$

Subbing in values for t_w^* and λ gives the equilibrium condition 34.

B's preference for $t_c = t$ over $t_c = t+1$ for a fixed t:

$$\frac{1-p+w}{1-\delta} - c - \frac{\Delta(1-\lambda_t)\delta}{1-\delta}\left[\frac{1-\delta^{t_w^*(|t=t_c)-t}}{k(1-\delta)}\right] > \frac{1-p+w}{1-\delta} - \delta\lambda_t c - \frac{\Delta(1-\lambda_t)\delta}{1-\delta}\left[\frac{k - k\delta + \delta - \delta^{t_w^*(|t=t_c+1)-t}}{k(1-\delta)}\right] \tag{46}$$

$$\frac{\delta\Delta(1-\lambda_t)}{k(1-\delta)^2(1-\delta\lambda_t)}\left[(1-\delta)(k-1) - \delta^{t_w^*(|t_c=t+1)-t} + \delta^{t_w^*(|t_c=t)-t}\right] > c \tag{47}$$

Subbing in values for t_w^* and λ_t gives the equilibrium condition 35.

The LHS of inequality 34 and 35 are positive for $k > 1$, $\theta > 2$. Thus, there must be some c sufficiently low where these inequalities are jointly satisfied. When this is true, B's strategy reported in proposition B.4 is a best reply to A's strategy.

I now analyze A's incentives to set $t_R^* = t_w^* - 1$ given B plays $s^B()$. I won't solve for the exact conditions that A invests in the face of containment. Rather, I'll show that for any fixed parameters of the game that imply B's best strategy is to set t_c^* that satisfies $1 \leq t_c^* < t_w^* < \theta$, and given $s^A()$, that I can find a values of R sufficiently small that $s^A()$ is A's best strategy.

In Lemma B.3 I showed, that if A chose not to invest, the game entered a stable subgame where B offered $p_t - w$ each period and A accepted. If A has not yet discovered nuclear weapons, this produces an expected utility for A $\frac{p-w}{1-\delta}$ in every period of that sub-game.

I also showed that if A's investment will trigger preventive war, A will not invest. As a result, A's investment rule caps the total number of A's investments at: $t_R = t_w^* - 1$. Exactly where A sets t_R^* depends on A's expected benefits from investment. If the cost of investment is high relative to the potential gains, A may set $t_R^* = 0$. That is, A never invests in her nuclear program. I now search for conditions that drive A to set set $t_R^* = t_w^* - 1$ given B's equilibrium reply t_w^*, t_c^*.

Notice by definition of $1 - \lambda_t$ each of A's investments produces a larger expectation of success than the investment before (with the exception of the period that containment is enacted). As a result, suppose any period $t < t_c^* - 1$ in which A's best strategy is to invest, it must be that A's best strategy is to invest at $t + 1$. Similarly, containment reduces A's expected gains from that investment and all subsequent investments. However, once B has enacted containment, A's incentives to invest increase in every subsequent period assuming that B will not enact preventive war. It follows that for any period t that satisfies $t_c^* \geq t < t_w^* - 1$, if A's best strategy is to invest in her nuclear program, then A's best strategy is to invest in her nuclear program in period $t + 1$.

It follows that if A's utility from investing at $t = 1$ and $t = t_c^*$ dominates not investing, then A's must prefer to invest all periods $t \in \{1 : t_w^* - 1\}$.

In the first period, A's from investment is:

$$
\begin{aligned}
EU_1^A(|t_c^* = t, t_R = t_w^* - 1) = \frac{p-w}{1-\delta} - R - \frac{\delta R(1-\delta^{t_w^*})}{1-\delta}\left[1-(1-\lambda_1)(t_c-1) - \frac{1-\lambda_1}{k}(t_w^* - t_c^*)\right] \\
+ \frac{(\Delta - R(1-\delta))(1-\lambda_1)}{(1-\delta)^3}\left[\delta(1-\delta^{t_c-1}) + \frac{1}{k}\delta^{t_c}(1-\delta^{t_w^*-t_c^*})\right] \quad (48)
\end{aligned}
$$

A's expected total utility from investing in the first period (given what A expects will happen if she does) is larger than not investing at all if:

$$
\begin{aligned}
0 < \frac{(\Delta - R(1-\delta))(1-\lambda_1)}{(1-\delta)^3}\left[\delta(1-\delta^{t_c-1}) + \frac{1}{k}\delta^{t_c}(1-\delta^{t_w^*-t_c^*})\right] \\
- R - \frac{\delta R(1-\delta^{t_w^*})}{1-\delta}\left[1-(1-\lambda_1)(t_c-1) - \frac{1-\lambda_1}{k}(t_w^* - t_c^*)\right] \quad (49)
\end{aligned}
$$

For any t_c that satisfies $1 \leq t_c < t_w^*$, and assuming that $0 < (1-\lambda_1)(t_c-1) - \frac{1-\lambda_1}{k}(t_w^* - t_c^*) < 1$[39], the RHS is strictly increasing in Δ and strictly decreasing in R. Further, the RHS is positive when $R = 0$, $\Delta > 0$. It follows that for any $\Delta > 0$, I can find an $R > 0$ such that the inequality is satisfied.

In period $t = t_c^*$, A's utility from investment knowing that investment will trigger containment is:

$$
\begin{aligned}
EU_{t_c^*}^A(|t_c^*, t_R = t_w^* - 1) = \frac{p-w}{1-\delta} - R + \frac{(1-\lambda_t)\delta(1-\delta^{t_w^*-t_c})}{k(1-\delta)^3}[\Delta - R(1-\delta)] \\
- \frac{\delta R(1-\delta^{t_w^*-t_c})}{1-\delta}\left[1 - \frac{1}{k}(1-\lambda_t)(t_w^* - t_c)\right] \quad (50)
\end{aligned}
$$

A's expected total utility from investing at $t = t_c^*$ (given what A expects will happen if she does) is larger than not investing in that period if:

[39] which must be true since this $t_w^* < \theta$.

$$0 < \frac{(1-\lambda_t)\delta(1-\delta^{t_w^*-t_c})}{k(1-\delta)^3}\left[\Delta - R(1-\delta)\right] - \frac{\delta R(1-\delta^{t_w^*-t_c})}{1-\delta}\left[1 - \frac{1}{k}(1-\lambda_t)(t_w^* - t_c)\right] - R \tag{51}$$

For any t_c that satisfies $1 \leq t_c < t_w^*$, the RHS is strictly increasing in Δ and strictly decreasing in R. Further, the RHS is positive when $R = 0$, $\Delta > 0$. It follows that for any $\Delta > 0$, I can find an $R > 0$ such that the inequality is satisfied.

Since R uniquely governs A's incentives, I must be able to find a sufficiently small value for R such that $s^A()$ is A's best reply to $s^B()$. When this is true, and the conditions reported in proposition B.4 form an SPNE.

B.2 Proposition 3.1: Containment increases the long-run risk of proliferation.

I prove proposition 3.1 in three steps. First, I define A and B's strategies in the counter-factual world where containment is impossible. Second, I define exactly what I mean by long-run risk of proliferation in both cases. Third, I compare the long-run risks of proliferation in the counter-factual world to the world where containment appears on the path and show that the risk is always greater in a world with containment.

Let's start with what would have happened if B was not allowed to use containment:

Lemma B.5 *Suppose a counter-factual game where containment is impossible. In that game, for any set of parameters $x \in X$ that produce equilibrium behavior described in proposition B.4, the following strategies describe A and B's equilibrium choices.*

- *A sets $t^{\bar{*}}_R = t^{\bar{*}}_w - 1$.*

- *B sets $t^{\bar{*}}_w = \theta - \frac{\delta\Delta - 2w}{2w(1-\delta)}$.*

All offers in all subgames are identical to those described in proposition B.4.

Starting with B's incentives, I've already shown that B sets t_w^* based on B's short-term risk threshold $1 - \lambda_w^* = \frac{2w(1-\delta)}{\delta(\Delta-2w)}$. I also showed that for any set of parameters where we observe containment on the equilibrium path (see proposition B.4), we reach a period t_w^* in which B prefers to respond to A's nuclear investment with war rather than allow A to acquire nuclear weapons with certainty. It follows that in the counter-factual game, given the same set of parameters, B must set $t_w^* < \theta$.

Turning to A's incentives, I've shown that A never invests when B can credibly threaten preventive war in a period that A invests. Thus, A must set $t_R^* < t_w^* - 1$. Further, A is only willing to play the strategy defined in proposition B.4, if A prefers to invest in the face of containment (which reduces A's gain from investment). It follows that A must also prefer investment knowing that B will not respond with competition under any condition $x \in X$.

Finally, none of B's offering rules or A's accept/reject rules depended on t_c^*. It follows that these are not effected when containment is not allowed.

I now turn to the definitions of long-run risk of nuclear proliferation. I'll start by summarizing some critical values that we have already solved for and then use them to define what I mean by long-run risk. I have already solved for the period in which B can credibly promise to fight a war given A's history of on-path investments in both versions of the game:

$$\bar{t_w^*}|(t_c = \emptyset) = \theta - \frac{\delta\Delta - 2w}{2w(1-\delta)} \tag{52}$$

$$t_w^*|(1 \leq t_c < t_w^*) = (\theta - t_c)k + t_c - \frac{\delta\Delta - 2w}{2w(1-\delta)}, \tag{53}$$

Although B's risk threshold λ_w^* did not depend on the history of investments and containment, the time it took to reach this threshold did. The differences in $t_w^*, \bar{t_w^*}$ are a function of containment. I also showed that the period in which A will stop investing $\bar{t_R^*} = \bar{t_w^*} - 1$, $t_R^* = t_w^* - 1$.

Furthermore, we know the probability that A will acquire nuclear weapons in any period that he invest. In the counter-factual game, there is an equal probability that A will acquire nuclear weapons in each period: $\frac{1}{\theta}$.

We can use these values to compute the cumulative probability that A discovers nuclear weapons as:

$$\frac{\sum_{i=1}^{t_R^{\bar{*}}} \frac{1}{\theta}}{\sum_{i=1}^{\theta} \frac{1}{\theta}} = \frac{\theta - \frac{\delta\Delta - 2w}{2w(1-\delta)}}{\theta} \tag{54}$$

The denominator is the cumulative sums given that A invests in every period of the game. This reflects the total possible risk. The numerator is the cumulative risk given the total number of investments that A makes on the path. The long run risk is the proportion of realized risk against total risk.[40]

We can do the same thing in the game where containment is allowed. We know that A will discover nuclear weapons with probability $\frac{1}{\theta}$ in each period $t \in \{1 : t_c^* - 1\}$. A will discover nuclear weapons with probability $\frac{1}{(\theta - t_c^*)k}$ in each period $t \in \{t_c^* : t_R^*\}$.

We can use these values to compute the cumulative probability that A discovers nuclear weapons as:

$$\frac{\sum_{i=1}^{t_c^*-1} \frac{1}{\theta} + \sum_{j=t_c^*}^{t_R^*} \frac{1}{(\theta - t_c^*)k}}{\sum_{i=1}^{t_c^*-1} \frac{1}{\theta} + \sum_{j=t_c^*}^{\theta} \frac{1}{(\theta - t_c^*)k}} = \frac{(\theta - t_c^*)k + t_c^* - \frac{\delta\Delta - 2w}{2w(1-\delta)}}{(\theta - t_c^*)k + t_c^*} \tag{55}$$

When I say that containment increases the long-run risk of proliferation, I mean that the probability defined in 54 is larger than 55 for any value of t_c^* given the implications it has for $t_w^*, t_w^{\bar{*}}$. Comparing these two inequalities:

$$\frac{(\theta - t_c)k + t_c - \frac{\delta\Delta - 2w}{2w(1-\delta)}}{(\theta - t_c)k + t_c} > \frac{\theta - \frac{\delta\Delta - 2w}{2w(1-\delta)}}{\theta} \tag{56}$$

$$\theta(k-1) > t_c(k-1) \tag{57}$$

We've already shown that $t_c^* < t_w^* < \theta$ on the path. Thus, the inequality is always satisfied. To be clear, this inequality is satisfied for any value of $t_c^* < t_w^*$. It follows that

[40] I use sums because I consider uniform distributions and fixed periods. But suppose τ_N was drawn from some $F()$ supported on positive real numbers, I get an equivalent result if I integrate over that function.

containment must always raise the long-run risk of proliferation given that it implies a delayed $\bar{t_w^*}$. This is enough to establish proposition 3.1.