How Fast and How Expensive? Uncertainty and Incentives in Nuclear Negotiations

William Spaniel*

February 5, 2019

Abstract

There is a growing consensus among researchers and policymakers alike that negotiated settlements can convince would-be proliferators to forgo nuclear weapons. However, uncertainty over the cost of weapons and development times interferes with the bargaining process. To sort out the incentives, I apply Bayesian mechanism design to a class of nuclear negotiation games. This methodology uncovers many properties that must hold across equilibria of all game forms. First, the probability of proliferation, the size of a settlement, equilibrium payoffs, and the value of bargaining must be monotonic in the proliferator's cost and development time. Second, uncertainty over development time represents a greater challenge to negotiations between military rivals, as no game form may exist that guarantees nonproliferation. Finally, game forms with nonproliferation equilibria are more likely to exist between military rivals than allies and when expected proliferation speeds are fast.

*Assistant Professor, Department of Political Science, University of Pittsburgh. (williamspaniel@gmail.com, http://williamspaniel.com).

1 Introduction

On July 14, 2015, the United States and Iran agreed to the landmark Joint Comprehensive Plan of Action. Colloquially known as the Iran Deal, the JCPOA traded economic concessions and improved diplomatic ties from the world community in exchange for a pause to Iran's nuclear program. Although whether Iran will comply in the long term remains uncertain, some policymakers view the JCPOA as a framework for future negotiations with would-be proliferators.

Contemporaneously, international relations scholars have seen a growing literature on bargaining over nuclear weapons, both in qualitative and formal treatments.[1] But the feasibility of settlements does not guarantee their implementation. Researchers have pinpointed resolve and self-reliance dispositions (Hymans 2006; Bleek and Lorber 2013), costs and technological capability (Singh and Way 2004; Jo and Gartzke 2007; Brown and Kaplow 2014; Fuhrmann and Tkach 2015), and development times (Kroenig 2009; Bas and Coe 2012) as determinants of proliferation behaviors. Unfortunately for nonproliferation advocates, opponents of potential nuclear states cannot always observe these factors. Such uncertainty casts a fog on the bargaining process, possibly leading to insufficient offers.

To better understand the incentives, one might develop a game theoretical model to analyze the connection between incomplete information and proliferation. However, existing models of nuclear proliferation diverge on a number of fronts, including the timing of moves and observability of past actions. Constructing a single model that features uncertainty may provide some insights, but the question will remain whether those results generalize. In short, standard modeling approaches cannot speak to what must be true across many different forms of negotiations.

Fortunately, game theory provides a powerful solution: mechanism design. Rather than solve only one model, mechanism design demonstrates what is true for a class of games. Prior work has demonstrated the utility of mechanism design in international relations for crisis bargaining (Banks 1990; Fey and Ramsey 2011), and so I apply the approach in the context of nuclear proliferation.

Many interesting results emerge. First, using the revelation principle (Myerson

[1] Reiss 1988; Paul 2000; Levite 2003; Rublee 2009; Benson and Wen 2011; Spaniel 2015; Bas and Coe 2016; Debs and Monteiro 2017; Volpe 2017.

1979), I generate a series of monotonicity properties. In all equilibria, types with lower costs and faster development times proliferate more frequently, receive better settlements when they do not, and have higher overall utilities than types with higher costs and slower development times. The fact that it is possible to write down game forms with equilibria with these properties should not be surprising; the results reflect how types with cheaper and quicker paths to the bomb should have greater coercive bargaining leverage. What is surprising is that they are *required* properties of equilibria; it is impossible to construct equilibria without those properties. Given that game theory has a nasty tendency to prove the unexpected, it is remarkable to observe such consistency across a large class of games.[2]

Second, I characterize the conditions under which equilibria exist that guarantee a nonproliferation resolution. In general, uncertainty over development speed is a more difficult problem to solve. This is because uncertainty over proliferation cost has no direct impact over a nonproliferator's payoff. In contrast, uncertainty over development speed implies that the nonproliferator does not know how soon it will suffer the consequences of bargaining failure. Thus, the nonproliferator faces greater temptation to walk away, hoping that it will only suffer consequences in the long term. These additional demands from the nonproliferator create significant constraints on settlements, making nonproliferation outcomes more difficult to ensure. I also show that this central difference is not sensitive to more complicated modeling specifications, including unobservable violations and counter-proliferation sanctions.

Third, nonproliferation settlements may be easier to achieve between security rivals than between friendly countries. At first, deals between allies may seem more likely due to their warm relations. However, for demands to be compatible, the benefits of nuclear weapons for the proliferator cannot be too great to the loss for the nonproliferator. For security rivals, the gains cancel out the losses, allowing the remaining costs of nuclear weapons to be the foundation of a bargaining range. In contrast, negotiations between friends may fail simply because nuclear weapons can accrue benefits external to the bilateral relationship.

Finally, faster average development times make nonproliferation equilibria *more*

[2]In the appendix, I also prove two propositions demonstrating that types of proliferators with high costs and long development times for nuclear weapons benefit the most from guaranteed nonproliferation outcomes. This is somewhat surprising given that one may suspect that types with cheaper and faster paths to a bomb can better leverage their threat to proliferate.

likely to exist. Intuitively, one might believe that slower times facilitate nonproliferation because potential builders will find them less valuable. However, that logic incentivizes nonproliferators to pull support from agreements that would satisfy types with fast development times. In turn, it may be impossible to construct nonproliferation equilibria. This counterintuitively means that countries like North Korea—whose observable factors all indicate a slow proliferation speed—may be more likely to develop nuclear weapons.

From a technical standpoint, this paper is closest to Fey and Ramsay 2011, which exploits mechanism design to find general results in crisis bargaining games.[3] War shares a common inefficiency puzzle with nuclear proliferation. However, beyond the clear substantive differences in application, my findings depart from crisis bargaining in a few ways. First, the sum benefits of proliferation may exceed the externalities, as the United States commonly negotiates with allies regarding proliferation. Second, it is not obvious how uncertainty over development times maps onto Fey and Ramsay's results; I show it is closer to uncertainty over power than the costs of fighting. Third, uncertainty over development speed is more complicated than those scenarios, as it affects both the expected outcome and the expected externalities suffered. The model indicates that it is more difficult to obtain a guaranteed settlement as a consequence. Fourth, the extensions I investigate are special nuclear proliferation. The analysis therefore has no parallel in Fey and Ramsay.

Given the subject matter, this paper is of obvious interest to scholars of nuclear proliferation. Due to the connections with Fey and Ramsay, it also speaks to the broader literature on bargaining and security. More subtly, the results have implications for the international institutions literature. Mechanism design by its nature studies what institutions can and cannot do given the constraints of those who might participate in them. Moreover, comparing the regular case with the compliance problem shows that assured nonproliferation is easier to obtain when the nonproliferation regime provides better information about potential proliferators' activities. This suggests that the core

[3]Fey and Ramsay 2011 itself extends Banks 1990, making Banks' work close to mine as well. Mechanism design approaches to the existence of efficient equilibria in bargaining games date back to Myerson and Satterthwaite 1983. Their setup involves an efficient trade between two parties. In my model, countries "trade" nondevelopment of nuclear weapons, with efficiency based on the nonproduction of the externality. However, such standard economic trades do not have interdependent values for bargaining failure, which Fey and Ramsay cite as a key difference when bringing mechanism design to international relations.

principles of the Non-Proliferation Treaty and the International Atomic Energy Agency have general theoretical support. The results also mesh well with core institution theory on information provision (Koremenos et al. 2001; Fang 2008).

2 Nuclear Negotiations and Barriers to Agreement

For traditional security theorists—who see security as "the only necessary and sufficient cause of nuclear proliferation" (Thayer 1995, 486)—nuclear negotiations may seem puzzling. However, there are two reasons why states may prefer to reach an agreement. On the potential proliferator's side, nuclear weapons are expensive. Schwartz (1998, 4) prices American development of nuclear weapons at $409 billion and deployment expenses at $3.241 trillion from 1940 to 1996, in constant 1996 dollars. This may seem striking, as many internal projections of nascent nuclear programs arrive at lower figures. However, atomic reactions are only the first step to having a nuclear deterrent; remaining infrastructure and upkeep expenses account for the difference. Potential proliferators would therefore benefit from leveraging the *threat* to develop a nuclear weapon. Indeed, deals could generate the security and policy concessions the states desire without the cost.

Both friends and rivals of potential proliferators also have a vested interest in halting nuclear advances. Providing inducements comes at a price—after all, the concessions must override a potential proliferator's incentive to build—but such agreements stop the production of negative externalities. These externalities can be large. Even if a new nuclear state takes great care over its command and control, safety has limits (Sagan 1993), and the cost of an accidental nuclear war would be staggering. An additional nuclear weapons state also causes a setback for the nonproliferation regime and creates another potential source for rogue nuclear weapons. Meanwhile, counteracting a rival's nuclear arsenal can drain a government's coffers; for instance, the United States spent $937.2 billion in constant 1996 dollars on defending against the bomb (Schwartz 1998, 4). Alternatively, rivals may have to give concessions post-proliferation under the threat of nuclear coercion. Even allies may fear their newly-empowered friends will deviate from a common trajectory.

That said, room to negotiate does not guarantee an agreement. Nonproliferators have incentive to offer just enough to induce compliance, as concessions come at a price to nonproliferators. But there are challenges to knowing the best offer. The cost of nuclear weapons is one source of uncertainty. Military expenditures are notoriously difficult to track (Kaplow and Gartzke 2016). Nuclear weapons add another layer of complexity, as a large percentage of expenditures are on scientific research and development. Meanwhile, whether a country has access to rogue nuclear materials and designs impacts the cost of a program. But potential proliferators have private information about that availability. Indeed, the nonproliferation regime had a hard time tracking the A.Q. Khan network even as Khan appeared to have a footprint in Libya, North Korea, and Iran (ElBradei 2011, 164-179).

A second source of uncertainty is resolve. In bargaining models, resolve refers to how much an actor cares about the issues at stake versus the material costs of a dispute. It is an internal attribute of leaders, making it difficult for other countries to track. For example, few doubt that Japan could build a nuclear weapon in short order at a low monetary cost. However, in the 1960s, American policymakers became uncertain of how serious Japan was in maintaining its nonproliferation norm. Despite the Japanese constitution renouncing war as a foreign policy tool, the director of Japan's future defense agency—and future Prime Minister—Yasuhiro Nakasone believed they nuclear weapons might still be legal. He commissioned a white paper on the subject (Campbell and Sunohara 2004, 222), perhaps signaling that the security issues Japan faced were worth finding a constitutional loophole. If Japanese leadership was serious, the United States would have to offer more; if this was a bluff, smaller offers would be satisfactory.

A third source of uncertainty is the length of time until a potential proliferator secures a nuclear weapon. Proliferation predictions are often inexact, with guesses sometimes missing by a decade (Richelson 2007; Montgomery and Mount 2014). Take, for example, U.S. intelligence estimates of the Soviet program. During World War II, the Soviet Union was "denied territory"—Washington had no agents on the ground (Gordin 2009, 82). Meanwhile, American efforts to stymie German espionage on the Manhattan Project meant that Russian spies went overlooked (Gaddis 1997, 93). As a result, U.S. intelligence was vague and underestimated Soviet progress, with the best-guess set at 1953 (Holloway 1994, 220). Yet Stalin's first test occurred in 1949. Alternatively, progress toward a nuclear weapon requires states to pass many scientific hurdles. Because successes occur stochastically and only the potential proliferator observes them, other states may not know where their bargaining partner is in the proliferation process

(Bas and Coe 2016; Bas and Coe 2017).

These informational asymmetries matter. To negotiate nonproliferation solutions, potential proliferators need to receive concessions commensurate with the value they expect to receive by proliferating. Higher resolve means a greater value for the end goal compared to the cost. Lower costs make the proliferation route more desirable. And faster progress leads to a quicker return on the investment. In turn, more resolved types with low costs and faster turnaround times need greater concessions to find deals acceptable. Given that inducements are expensive to concede, nonproliferators may wish to gamble on stingier offers so as not to overpay.

The alternative solution is to learn about the proliferator's resolve, cost, and speed to a nuclear weapon. However, eliminating this uncertainty may prove difficult. Potential proliferators have incentive to misrepresent the desirability of building a bomb. After all, if countries who could proliferate easily receive more concessions through bargaining, a country facing greater hurdles would want to bluff strength. It is therefore not surprising that President Zulfikar Ali Bhutto once remarked that Pakistanis would "eat grass" if it were necessary to build a bomb (Khan 2012).

3 Mechanism Design and the Revelation Principle

To sort out the information problem, it would be useful to have a tool that could investigate equilibria across a class of nuclear negotiation games. Such a tool would remove any concern that an equilibrium result is an artifact of a specific functional form. Fortunately, mechanism design provides such a method. Mechanism design is not often used in international relations, so this section serves as a brief primer.[4]

In short, I exploit a finding known as *the revelation principle*. The revelation principle says that every Bayesian Nash equilibrium has an incentive-compatible direct mechanism that produces an identical outcome. A direct mechanism is a mapping of player types to outcomes. An incentive compatible mechanism is therefore a mechanism for which no type has incentive to lie about its type. In turn, the revelation principle says that one can always write down an incentive compatible mapping from types to outcomes that mimics the equilibrium.

[4]For a fuller treatment with international relations applications and examples, see Fey and Ramsay 2011 (152-157).

To understand why such mechanisms always exist, it helps to take a step back and think broadly about equilibrium's meaning. Game theory maps inputs such as strategies available, payoff functions, and private information to outcomes. Equilibrium formalizes how that mapping behaves and generates the strategies that the players select over the course of a game. Direct mechanisms skip the process of players choosing strategies and instead immediately assigns outcomes based on the type a player reports to the mechanism designer.

The revelation principle works because the mechanism designer functions as a middleman. When working through an interaction with a middleman, players do not adopt strategies themselves. Rather, they tell the middleman their type, and the middleman plays the game on their behalf. This generates outcomes, which the middleman can then assign to those players without those player ever having actually played the game.

Why don't players have incentive to lie about their types? This is the beauty of the revelation principle. The direct mechanism mimics what the equilibrium play would have been. Thus, lying to the middleman is redundant—if there is any incentive to misrepresent, the middleman can bluff on a particular type's behalf. Reporting anything else causes the middleman to bluff in a different way, potentially causing a type to take greater risks than it would want to.

The revelation principle's contrapositive offers an additional application: if *no* incentive compatible direct mechanism with certain qualities exists, then no Bayesian Nash equilibrium with those qualities exists either. This is how a researcher can draw general conclusions about incentive structures without having to analyze every game form imaginable. For example, suppose that it is impossible to create an incentive structure such that the probability of proliferation is 0 and there is no incentive to lie to the designer. Then the contrapositive says that no equilibrium of *any* game features a zero probability of proliferation. This trick allows me to make strong claims about what is possible in nuclear negotiations, and I exploit it in the following section.

4 Incentives in Nuclear Negotiations

Consider a class of games in which one state can develop nuclear weapons and another can provide inducements to end the process. I call the potential developer the Proliferator and the inducement provider the Nonproliferator. The game can end in either development or a settlement. Denote a settlement as a transfer of x from the Nonproliferator to the Proliferator. Let the Nonproliferator's utility for a transfer be $u_N(x) = -x$ and the Proliferator's utility be $u_P(x) = x$. This transfer captures the yearly economic aid, military support, policy concessions, or pure bribery that are commonplace in nuclear negotiations.

Proliferation payoffs are more complicated. Let the Proliferator's utility for that outcome be $u_P(b, \delta, c) = \delta b - c$. The parameter $b > 0$ represents the benefits of nuclear weapons. This includes the security gains, policy concessions, and additional prestige a country may acquire by proliferating. The parameter $\delta \in [0, 1]$ captures the speed of the proliferation process, with larger values indicating more rapid development. Using a discount factor in this manner is common in games of shifting power (Fearon 1995).[5] The straightforward interpretation of δ is the time delay to proliferation. However, for risk-neutral actors, δ equivalently represents the *expected* delay value.[6] Lastly, $c > 0$ reflects the cost of nuclear weapons, weighted by the Proliferator's resolve.[7]

Meanwhile, let the Nonproliferator's utility for nuclearization be $u_N(e, \delta) = -\delta e$, where $e > 0$ represents the externality the Nonproliferator suffers from the Proliferator acquiring nuclear weapons. Note that the development speed determines the Nonproliferator's payoff, as it does not suffer the externalities until the Proliferator actually obtains a weapon.

I do not make any assumptions about whether the Proliferator gains more than the Nonproliferator loses through the introduction of nuclear weapons. In alliance relationships, one might imagine that the security benefits the Proliferator gains causes no pain to the Nonproliferator. Here, e captures the additional risks of accidental nuclear annihilation, environmental costs, and damage to the nonproliferation regime. In such cases, the relationship between e and b could break either way. On the other hand, for purely coercive rivalries, one might imagine that $e > b$. This is because the

[5]The interpretation of δ is often a time value of preferences. But because rising powers do not exercise their newfound strength until the next period, δ implicitly defines how long it takes to reach that period.

[6]This type of stochasticity can arise when breaking scientific barriers occurs probabilistically (Bas and Coe 2012). With uncertainty over δ, the key feature is that the Proliferator has better knowledge of the likelihood of the underlying likelihood of success. My key theoretical results hold if the Proliferator has risk averse preferences, but modeling those preferences complicates notation.

[7]This standardization is common in the formal security literature. Higher levels of resolve mean that the Proliferator internalizes the monetary costs at lower rates. Thus, c is decreasing in resolve.

Nonproliferator necessarily loses whatever security benefits go to the Proliferator, and the Nonproliferator also suffers the other costs described above. Whether $e > b$ has unexpected implications for the existence of settlements. I return to this point later.

As the utility functions indicated, the analysis below separates two sources of uncertainty: the Proliferator's cost to build c and its development speed δ. This information is private to the Proliferator, and its payoff depends on its cost and speed type. In contrast, the Nonproliferator's payoff only depends on the proliferation speed, with faster development bringing the negative payoff more quickly. The Nonproliferator only has a prior belief about the Proliferator's type. To formalize this, in the first case, let the Nonproliferator's prior belief on c be bounded on the interval $[\underline{c}, \overline{c}]$, with cumulative distribution function $F(c)$. Meanwhile, in the second case, let the Nonproliferator's prior belief on δ be bounded on the interval $[\underline{\delta}, \overline{\delta}]$, with cumulative distribution function $T(\delta)$. I assume that both distribution functions are differentiable everywhere and let $f(c)$ and $t(\delta)$ represent the associated density functions.

With those preliminaries set, let G represent the game form. A game form consists of a set of actions for each player i, denoted A_i, and an outcome function $g(a_P, a_N)$, where $(a_P, a_N) \in A = A_P \times A_N$. In words, the function g takes actions from each player as its inputs and maps those to an outcome. I define a **nuclear negotiation game** as a game that sees either a settlement or proliferation as its outcome. This class of games is infinitely large. Simple structures like the ultimatum game are in the set. However, complicated interactions featuring many rounds of proposals or cheap talk messages are also included.

Now consider the game's outcome space, denoted $\Omega = \{p, x\}$. Despite the limitless game forms, the outcome space consists of just two elements: a proliferation outcome (p) and a transfer (x). The outcome function therefore only has two components: a probability of the proliferation outcome $\pi \in [0, 1]$ and the value of the settlement.

Following past mechanism design work on conflict (Banks 1990; Fey and Ramsay 2011), I impose the *voluntary agreements* requirement on the game form. Voluntary agreements mandates that each player may unilaterally choose a strategy that provides it with at least its payoff for failed negotiations. In other words, the game cannot force an actor to receive a settlement that generates a worse payoff than the proliferation outcome. This axiom takes seriously the problem of anarchy and unenforceable contracts in international relations. Because states are not bound to agreements, they must find the terms of settlement preferable to the alternative to maintain a pact.

A careful reader will notice that this setup does not explicitly allow for a preventive war outcome. This is intentional. In cases where preventive war is credible, the existing literature shows that creating game forms with nonproliferation equilibria is easy. For example, when power shifts are completely observable, states internalize the threat of preventive war and take measures to avoid it (Chadefaux 2011; Debs and Monteiro 2014; Spaniel 2015). Indeed, I develop such a nuclear negotiations game with preventive war in the appendix to illustrate this. However, in such games, the credible preventive war threat renders informational dynamics irrelevant. I therefore focus here on cases where preventive war is not credible.[8]

Such cases are common empirically. For example, consider Washington's negotiations with Tokyo and Seoul over their potential nuclear programs. There was no reasonable expectation that the United States would go to war with Japan or South Korea if bargaining failed. Although it is obvious that preventive action does not work against allies, the same logic sometimes applies to bitter rivals. American officials believed that preventive war with the Soviet Union was impractical or would be too expensive (Silverstone 2007, 51-75), rendering an intervention impossible even if intelligence reports had better estimated Moscow's progress toward a bomb. The United States and Soviet Union had a similar disinterest in terminating China's program (Burr and Richelson 2000).

With the setup complete, I turn to the results. I break down the findings into two categories: when the Nonproliferator knows δ but not c and when the Nonproliferator knows c but not δ.

4.1 Uncertainty over Costs

I begin with a series of monotonicity results. Holding fixed an equilibrium of a nuclear negotiation game, these findings state what must be true when comparing a lower cost type to a higher cost type.

[8]More rigorously, the central results here apply to a broader class of games featuring preventive war given a simple assumption. If the utility of preventive war is worse than the utility for allowing a power shift to transpire, then the Nonproliferator would never choose that route anyway, and the Proliferator would see any threats to do so as incredible. The states can then only bargain to resolve the dispute, as I have modeled. I address one caveat to this later.

Proposition 1. *Suppose the proliferation cost c is private information but the development speed δ is common knowledge. Let s^* be any equilibrium of a game form G. The following must be features of that equilibrium:*

1. *The probability of proliferation weakly decreases in c.*
2. *For all types with a positive probability of settlement, the value of the settlement weakly decreases in c.*
3. *The proliferator's equilibrium utility weakly decreases in c.*

The proofs for these qualitative features of all equilibria do not provide much intuition, so I save them for the appendix. Instead, I explain the underlying logic of each result below.

Let c and c' be the costs of proliferation for two types, where $c' > c$. The first monotonicity result states that, within an equilibrium of any game form, the probability of proliferation for type c' must be no greater than the probability of proliferation for type c.[9] This is a product of incentive compatibility. The c' cost type must not want to mimic the c type, and vice versa. Recall that all types receive the same utility for a transfer x. Thus, the only disincentive for misrepresentation is the proliferation outcome because the utility for that depends on c. The weakly higher probability of proliferation for lower cost types disincentivizes higher types from lying, as doing so increases the portion of the time a higher type pays its larger cost for proliferation. The relationship is weak because equilibria may exist in which settlement occurs with certainty.

This first result provides a helpful narrative of the United States' negotiations with Pakistan. Throughout Pakistan's nuclear progressions, the United States struggled to understand its bargaining partner's resolve. Within a short time span, one analyst described the program as a "train that [was] going down the track very fast" and that he was "not sure anything [would] turn it off"[10]; at the other end of the spectrum, another

[9] The peculiar phrasing clarifies that this is *not* a comparative static that holds across all models. Within the same model, it is possible to increase overall proliferation costs and increase the probability of proliferation. However, within the equilibrium of those models, lower cost types still proliferate at least as often as higher cost types.

[10] NSA EBB No. 333, Document 42, `https://nsarchive2.gwu.edu/nukevault/ebb333/doc42.pdf`.

believed that the U.S. was likely "to achieve a quick resolution of the nuclear reprocessing issue."[11] Proposition 1 claims that in *any* nuclear negotiation game, a "railroad track" type must be (weakly) more likely to proliferate. True to form, Pakistan pushed forward. Indeed, a Pakistani Foreign Secretary flaunted their progress specifically as a credible message of the country's commitment to proliferation (Ahmed 1999, 190).

Moving on, the second monotonicity result states that, within the same equilibrium of the same game form, a type with cost c' cannot receive a larger settlement than a type with cost c if settlements occur for those types.[12] This is a product of the first result, voluntary agreements, and incentive compatibility. Imagine that a low cost type could receive a better settlement by pretending to be a high cost type. Voluntary agreements require that the low cost type's settlement be at least as good as its proliferation payoff; because the high cost type allegedly receives a larger settlement than the low cost type, the low cost type's settlement conditional on lying must be greater than if it tells the truth. Moreover, as a consequence of the first monotonicity result, reporting a higher cost also results in receiving that settlement at least as often as its proliferation outcome. But this violates incentive compatibility: lying would generate a greater payoff some of the time while coming at no risk. So a lower cost type's settlement must be at least as large as a higher cost type's.

The final monotonicity result states that a low cost type's overall utility must be at least as large as a higher cost type's. If the outcome is a settlement with certainty, this is a straightforward consequence of the second monotonicity result. If proliferation occurs with positive probability, suppose that a higher cost type received a greater payoff. Recall that proliferation is cheaper for the low cost type and the probability of proliferation is at least as large for the low type. Thus, the only way the high cost type can do better is if its settlement value is large enough that it overcompensates for those problems. But if that were the case, the low cost type could exploit the benefit by mimicking that type. This violates incentive compatibility. Therefore, a low cost type must receive at least as great a payoff as a high cost type.

Collectively, the monotonicity results explain what must be true when comparing two cost types. They do not speak to the possible outcomes overall. In particular,

[11]NSA EBB No. 333, Document 2, `https://nsarchive2.gwu.edu/nukevault/ebb333/doc02.pdf`.

[12]The case where settlements do not occur for those types is uninteresting as the value of a settlement for them is vacuous.

nonproliferation advocates may wonder when it is possible to guarantee an agreement. I therefore define a *nonproliferation equilibrium* as an equilibrium in which the probability of proliferation is 0. The following proposition generates insight in that regard:

Proposition 2. *Suppose the proliferation cost c is private information but the development speed δ is common knowledge. A game form with a nonproliferation equilibrium exists if and only if $b \leq \frac{\underline{c}}{\delta} + e$. In such equilibria, all types receive the same settlement.*

Why can't there be differential settlements? The appendix gives the proof, but a clear intuition elucidates the claim. Suppose there were. Every type would want to receive the highest settlement possible and therefore report the type associated with that highest settlement. The only disincentive for false reporting is the possibility of proliferation with a greater probability, which could convince higher cost types to not mimic lower cost types. However, guaranteed nonproliferation prohibits that disincentive. In turn, the only way to keep each type honest is to assign them the same payoff regardless of what they report.[13]

Voluntary agreements further narrows what that settlement looks like. Recall that voluntary agreements requires that each type receive at least its proliferation payoff from a settlement. The largest of these proliferation payoffs comes from the lowest cost type, which earns $\delta b - \underline{c}$ by proliferating. Therefore the constraint on x for the lowest cost type is

$$x \geq \delta b - \underline{c} \tag{1}$$

But because incentive compatibility requires the settlement be identical across the board, x must be greater than $\delta b - \underline{c}$ for all types.

The Nonproliferator creates a second voluntary agreement constraint. Specifically, the Nonproliferator must receive at least its payoff for assured proliferation. Recalling that the Nonproliferator receives $-x$ for a settlement, voluntary agreements then mandates

$$-x \geq -\delta e \tag{2}$$

[13]The identical settlements requirement is why, as I prove in the appendix, that types with the highest cost benefit the most from the deal. Subtracting each type's reservation value from the transfer must result in a larger difference for a type with a lower payoff. The same logic applies to the slowest type with the alternative source of uncertainty.

Solving Line 2 for x and stringing it together with Line 1, a settlement x simultaneously satisfies both voluntary agreements conditions if:

$$\delta b - \underline{c} \leq x \leq \delta e \tag{3}$$

Such an x exists if:

$$\delta b - \underline{c} \leq \delta e$$

$$b \leq \frac{\underline{c}}{\delta} + e \tag{4}$$

This is the condition given in Proposition 2.

However, the above logic only showed that Line 4 must hold to obtain guaranteed nonproliferation. What remains to be seen is whether game forms exist that guarantee nonproliferation given those parameters. Fortunately, showing existence is trivial. Fix an x that satisfies Line 3. Consider a game in which both players simultaneously choose to consent to x or not. If both players consent, then x is implemented. Otherwise, the Proliferator builds. Both players consenting is an equilibrium—vetoing does not yield a higher payoff by construction of x.[14]

At the same time, Line 4 does not say whether a specific game form has a nonproliferation equilibrium. For example, in an ultimatum game in which the Nonproliferator makes a take-it-or-leave-it offer, proliferation results in equilibrium for certain prior beliefs. Instead, Proposition 2 makes a more general claim about whether negotiations could possibly result in nonproliferation.

Despite that limitation, Line 4 gives a clear necessary condition for a nonproliferation equilibrium to exist. The right hand side is increasing in $\underline{c}$ and e but decreasing in δ. Thus, Line 4's condition is more likely to hold when the cost and externalities of proliferation are high but the speed of development is slow. The changes to $\underline{c}$ and δ make proliferation less attractive to the Proliferator and thus more inclined to find a settlement preferable. Meanwhile, larger values of e provide a similar incentive to the Nonproliferator.

By placing further structural assumptions on the payoffs, the following claim refines

[14] This game form also satisfies voluntary agreements because either may exercise its veto to obtain the proliferation payoff.

when game forms with nonproliferation equilibria exist:

Corollary 1. *Suppose the proliferation cost c is private information, the development speed δ is common knowledge, and $e > b$. Then there exists a game form with a nonproliferation equilibrium.*

Put differently, a sufficient condition for the existence of game forms with nonproliferation equilibria is that the benefit derived from nuclear weapons results in an equivalent loss for the Nonproliferator. One way this could happen is if the parties are security rivals. The intuition here is straightforward. In these cases, proliferation results in inefficiencies for both parties. Meanwhile, the Nonproliferator's security loss offsets the Proliferator's security benefit. In turn, transferring the value of the security benefit from the Nonproliferator to the Proliferator guarantees that both parties enjoy the settlement. On the other hand, when the security benefits for proliferation only marginally hurt the Nonproliferator, the most the Nonproliferator is willing to concede may be less than the minimum amount necessary to induce the lowest cost type to accept. Every equilibrium contains some positive probability of proliferation in these situations.

Although that logic is intuitive, it contrasts with the focus in the inducements literature. Existing studies tend to focus on how ideologically-friendly countries can form security agreements to convince would-be proliferators to end their programs. Work on Japan (Paul 2000; Levite 2003) and South Korea (Campbell and Sunohara 2004; Lanoszka 2013) are emblematic in this regard. These agreements have inherent appeal, as friends would seem to have an easier time negotiating with one another to assure continued positive relations. Research on negotiations between rivals—or potential rivals—is comparatively sparse (Drezner 1999; Spaniel 2015; Volpe 2017).

Negotiations may in fact be easier overall for friends than enemies due to non-informational mechanisms. For example, other potential causes of bargaining failure like commitment problems may be less present with allies. Allies may also have better information. Nevertheless, agreements between allies first require a bargaining range to exist. Security relationships guarantee that existence; the same guarantee does not hold for friends. And for every Japan and South Korea, there is a United Kingdom or France where the benefits the United States could provide were insufficient to induce nonproliferation. The key conclusion here is that researchers should not take for granted the apparent ease of negotiating with friends.

4.2 Uncertainty over Speed

I now continue to the case where the Nonproliferator knows the Proliferator's cost but only has a prior belief about the speed δ. As before, I begin with monotonicity results.

Proposition 3. *Suppose the development speed δ is private information but the proliferation cost c is common knowledge. Let s^* be any equilibrium of a game form G. The following must be features of that equilibrium:*

1. *The probability of proliferation weakly increases in δ.*
2. *For all types with a positive probability of settlement, the value of the settlement weakly increases in δ.*
3. *The proliferator's equilibrium utility weakly increases in δ.*

Put differently, holding fixed an equilibrium of a game form, types that can rapidly proliferate are at least as likely to proliferate as slower types, receive at least as great of a settlement if a settlement is reached, and receive at least as great of an overall payoff. A careful examination of these claims reveal that they are analogous to the monotonicity results with uncertainty over costs. The intuition is identical (noting that in this case *higher* values of δ represent superior proliferation threats), so I leave all proof work for the appendix.

The first point comports well with the literature on sensitive nuclear assistance. In any bargaining game, types that privately know they are relatively quick to develop a weapon are more likely to do so. One cause of this private information is the secret transfer of sensitive nuclear technology. If an analyst observes *ex post* that a country received such transfers, Proposition 3 predicts such countries proliferate more often. Kroenig (2009) shows that this is the case empirically.

It also provides insight on why the Soviet Union proliferated in 1949. American intelligence analysts faced uncertainty over Stalin's speed to a nuclear weapon (Richelson 2007; Gordin 2009; Bas and Coe 2012, 667). Unbeknownst to Washington, the Soviet Union was a "fast" type due to Moscow's penetration of the American nuclear program

(Montgomery and Mount 2014, appendix). Proposition 3 states that in all equilibria of nuclear negotiation games, such a type is more likely to proliferate than slower types. Indeed, the Soviet Union ended America's nuclear monopoly in 1949.

Despite Proposition 3's similarity to Proposition 1, concluding that uncertainty over costs and uncertainty over development speed function identically based on this would be hasty. The following proposition shows substantial differences when it comes to the existence of assured nonproliferation settlements:

Proposition 4. *Suppose the development speed δ is private information but the proliferation cost c is common knowledge. A game form with a nonproliferation equilibrium exists if and only if $\overline{\delta}b \leq c + e\left(\overline{\delta} - \int_{\underline{\delta}}^{\overline{\delta}} T(\delta)d\delta\right)$. In such equilibria, all types receive the same settlement.*

Recall that the condition necessary for a nonproliferation equilibrium with uncertainty over costs only depended on b, $\underline{c}$, δ, and e. The condition here includes all of those (swapping $\underline{c}$ with $\overline{\delta}$) plus a measure of the shape of δ's distribution function. Why does this matter here but not before? Again, the proof is illustrative. To preview the answer, uncertainty over costs creates *independent* values for proliferation; the Proliferator's cost type has no baring on the Nonproliferator's payoff. In contrast, uncertainty over development speed creates *interdependent* values for proliferation[15]; how fast the Proliferator obtains a nuclear weapon affects the Nonproliferator's payoff because it determines how soon the Nonproliferator suffers the externality.

As before, assured nonproliferation implies that payoffs are only a function of the settlement offer. Because all types internalize a settlement identically, incentive compatibility requires the settlement amount x to be equivalent for all types. Voluntary agreements additionally mandates that

$$x \geq \overline{\delta}b - c \tag{5}$$

otherwise the highest type would prefer proliferating.

Voluntary agreements further requires the settlement generate a better payoff for the Nonproliferator than if it forced all types to develop nuclear weapons. Unlike

[15] One may wonder how results change if the Nonproliferator is uncertain about both δ and c. Because interdependent values make agreements more difficult and interdependence exists with multiple forms of uncertainty, the results would more closely match uncertainty over just δ than uncertainty over just c.

with uncertainty over costs, the Nonproliferator's expectation over δ forms part of the constraint. In particular, the Nonproliferator's payoff is e weighed by the probability distribution over δ. Comparing that to the Nonproliferator's payoff for an accepted offer generates the following voluntary agreements constraint:

$$-x \geq -e \int_{\underline{\delta}}^{\overline{\delta}} t(\delta)\delta d\delta \tag{6}$$

Solving for x, integrating by parts, and stringing the resulting inequality with Line 5 yields:

$$\overline{\delta}b - c \leq x \leq e \left(\overline{\delta} - \int_{\underline{\delta}}^{\overline{\delta}} T(\delta) d\delta \right) \tag{7}$$

Such an x exists if:

$$\overline{\delta}b \leq c + e \left(\overline{\delta} - \int_{\underline{\delta}}^{\overline{\delta}} T(\delta) d\delta \right) \tag{8}$$

This is the condition given in Proposition 4. Despite the complicated notation, it has a straightforward interpretation. Recall that $\overline{\delta}$ represents the fastest possible time to proliferation. The left side therefore calculates the Proliferator's best possible value for building a weapon. The right hand side includes the inefficiencies of bargaining failure, with $\overline{\delta} - \int_{\underline{\delta}}^{\overline{\delta}} T(\delta)d\delta$ reflecting anticipated proliferation speeds given the Nonproliferator's prior belief.

Like before, Line 8 is only a necessary condition. However, existence is easy to show. Simply consider the same mutual consent game outlined with uncertainty over costs. Now change the settlement to x to any that satisfies Line 7. The same caveat about how Line 8 is not a sufficient condition also applies.

With a little more work, Proposition 4 yields a counterintuitive prediction. Consider two distributions, $T_1(\delta)$ and $T_2(\delta)$. Let $T_1(\delta)$ first order stochastically dominate $T_2(\delta)$, meaning that $\int_{\underline{\delta}}^{\overline{\delta}} T_2(\delta)d\delta > \int_{\underline{\delta}}^{\overline{\delta}} T_1(\delta)d\delta$. This yields the following proposition:

Proposition 5. *Suppose a nonproliferation equilibrium exists for distribution $T_2(\delta)$. Then a nonproliferation equilibrium also exists for distribution $T_1(\delta)$. However, existence of a nonproliferation equilibrium for $T_1(\delta)$ does not imply existence of a nonproliferation equilibrium for $T_2(\delta)$.*

Roughly, Proposition 5 states that the existence of nonproliferation equilibria decrease in the first order stochasticity of the proliferation speed's distribution. One can observe this by noting that the right hand side of Line 8 is decreasing in the integral.

First order stochastic dominance may seem like a mere technical condition, but it has a critical implication for the empirical study of nuclear proliferation. Holding fixed the support, distribution $T_1(\delta)$ places greater probability on faster proliferation times than distribution $T_2(\delta)$. Proposition 5 counterintuitively says that the greater weight on *faster* proliferation times facilitates nonproliferation equilibria. Bluntly, if a country believes its opponent is likely to obtain a nuclear weapon quickly, reaching an agreement is *easier* in the sense that nonproliferation equilibria are more likely to exist. This is despite how faster expected proliferation times make nuclear weapons better investments on average.[16]

To illustrate the logic and empirical consequences, consider a case like North Korea. Between its low gross domestic product and its lack of industrialization, all observable indicators suggested that North Korea would have a long time table to nuclear weapons. Intuitively, one would expect such circumstances to make proliferation less likely; indeed, many empirical studies argue that more technologically sophisticated countries are more likely to build a bomb (Singh and Way 2004; Jo and Gartzke 2007; Fuhrmann 2012; Brown and Kaplow 2014). But Proposition 5 states that skepticism about North Korea's proliferation capabilities could prohibit a nonproliferation equilibrium.

A recounting of the constraints help make sense of the claim. For a nonproliferation equilibrium to exist, two things must be true. First, incentive compatibility requires the Proliferator receive an amount at least as large as the fastest type's proliferation value. Second, voluntary agreements requires the Nonproliferator to prefer giving that amount to taking the expected externality across all the possible proliferation speeds. When proliferation times are fast, making the deal looks more attractive; the Nonproliferator does not lose much by paying the necessary amount when the distribution is heavily slanted toward fast times anyway. When proliferation times are slow, rolling the dice looks more attractive. As a result, slower proliferation times prevent the existence of

[16]This mechanism holds within the situations where preventive war is incredible. Outside of that case, there is a caveat. Preventive war looks more attractive when proliferation speeds are faster, as the opponent suffers the consequences sooner. Thus, putting more weight on types for which an opponent would want to fight facilitates nonproliferation equilibria under these alternate scope conditions through a preventive war mechanism.

nonproliferation equilibria.

This logic manifested itself in negotiations with North Korea in the 1990s. In 1994, the Clinton administration signed the "Agreed Framework," which traded a freeze in nuclear weapons for normalization of relations and energy transfers. But Democrats lost control of Congress the following January, and Republicans believed they could obtain North Korea's compliance at a lower price. They therefore blocked funds (Martin 2002; Chinoy 2008, 8), forcing Clinton to take a more aggressive approach. However, North Korea's actions indicate that it believed it could do better by building nuclear weapons. Bargaining thus broke down.

To be clear, Proposition 5 only makes a claim about the existence of nonproliferation equilibria. It does not state what must be true across comparative statics for all nuclear negotiation games. Indeed, decreasing an inefficiency can lead to more failed negotiations in many bargaining models. Consequently, the mechanism design results do not invalidate the theoretical claims that the aforementioned empirical scholars have linked to statistical patterns. Proposition 5 instead flags a limitation of those theoretical claims. Observable factors that suggest a slow time to proliferation do not assure fewer instances of proliferation.

Moving on, adding structure to the externality once again grants additional empirical leverage:

Corollary 2. *Suppose the development speed δ is private information, the proliferation cost c is common knowledge, and $e > b$. This is not sufficient for the existence a game form with a nonproliferation equilibrium.*

In other words, the externality outweighing the benefit guarantees the existence of nonproliferation equilibria with uncertainty over costs; it does not guarantee the existence of nonproliferation equilibria with uncertainty over development speed. As previewed earlier, the key difference is that uncertainty over development speed results in interdependent values. If the Nonproliferator does not know the Proliferator's development cost, it can still rest assured that the Proliferator is willing to accept δb—that amount is sufficient even for a type with zero cost. With uncertainty over development speed, it does not know whether the Proliferator is willing to accept $\delta b - c$ precisely because the Nonproliferator does not know the value of δ. And as the discussion of the distribution function elucidated, the Nonproliferator may prefer allowing all types to proliferate to reaching a settlement acceptable to the fastest type.

Instead, the existence of a nonproliferation equilibrium hinges on the development cost and the externality. Let $\epsilon = e - b$ represent the non-security related exeternalities of proliferation, including environmental damage and risk of accidental launch. Substituting $e = b + \epsilon$ into Line 8 and rearranging changes the cutpoint to

$$b\left(\int_{\underline{\delta}}^{\overline{\delta}} T(\delta)d\delta\right) \leq c + \epsilon\left(\overline{\delta} - \int_{\underline{\delta}}^{\overline{\delta}} T(\delta)d\delta\right). \tag{9}$$

Large enough costs and externalities cause the minimum amount necessary to appease the fastest type to overlap with the expected loss the Nonproliferator suffers if all types proliferate.

The left side of Line 9 indicates that researchers and policymakers would benefit from more work on the value of nuclear weapons. To reiterate, the integral measures how quickly the Nonproliferator expects the Proliferator to obtain a weapon. However, this expectation becomes irrelevant as the benefit descends to 0; it does not matter whether proliferation will be fast or slow if nuclear weapons do not provide any benefits. If nuclear weapons are not useful for coercion and instead serve as deterrents on specialized issues (Sechser and Fuhrmann 2017), then game forms with nonproliferation equilibria should be plentiful. But if nuclear weapons fulfill a traditional military role or states worry about brinkmanship (Schelling 1960, 187-204; Beardsley and Asal 2009), then nonproliferation equilibria may prove elusive.

Combined, Propositions 2 and 4 draw a clear connection to the crisis bargaining literature in the special case of security relationships. Fey and Ramsay (2011) show that uncertainty over the costs of war guarantee the existence of always peaceful equilibria; in contrast, uncertainty over power and sufficiently low costs of war mean that no such equilibria can exist. A similar logic prevails here. The unexpected finding is that the probability of victory in war works somewhat analogously to time to proliferation.

However, the result is not identical to Fey and Ramsay. With uncertainty over the probability of victory in war, the quantity of the inefficiency is common knowledge. Here, the proliferator's cost is common knowledge but the Nonproliferator does not know its externality; this is because a longer time to proliferation delays the externality for a longer period of time. The uncertainty manifests itself in how ϵ is multiplied by $\overline{\delta} - \int_{\underline{\delta}}^{\overline{\delta}} T(\delta)d\delta$ in Line 9. Because $\overline{\delta} > \int_{\underline{\delta}}^{\overline{\delta}} T(\delta)d\delta$, the full effect of the externality does not shift the cutpoint. Moreover, probability distributions that place a greater share on lower speeds (that is, probability distributions that are first order stochastically dominated) exacerbate the effect. This creates a double whammy when combined with the similar problem with b from earlier. In short, implicit uncertainty over the externality can make the nonexistence of efficient equilibria substantially more likely than in the crisis bargaining case.

On a related note, a clear problem with reductions of externalities further differentiates Proposition 4 from Proposition 2. Fix a game form and a nonproliferation equilibrium in the $e = b + \epsilon$ case. With uncertainty over development costs, shrinking ϵ down to 0 still maintains the Line 4. That is, the existence of that settlement is not sensitive to the non-security externalities. In contrast, with uncertainty of development speed, if c is sufficiently small, reducing ϵ eventually invalidates Line 8. Put differently, the existence of game forms with nonproliferation equilibria is sensitive to externality reductions.

This is a real problem. The international community works to reduce the possibility of accidental launch (Caldwell 1987) and stops trafficking through the Proliferation Security Initiative.[17] Spaniel (2018) suggests that these institutions do not have a clear perverse effect on bargaining because uncertainty in nuclear proliferation only creates independent values. Ignoring uncertainty over development speed leads to the wrong inference, as these efforts can unambiguously contribute to proliferation in some models.

As a final point about Proposition 4, note that many nuclear negotiations feature third parties in the discussions. For example, although most narratives of the Joint Comprehensive Plan of Action focus on the bilateral relationship between the United States and Iran, the official agreement also includes China, Russia, the United Kingdom, France, Germany, and the European Union. These parties can provide extra subsidies to further incentivize an agreement. Such subsidies conditional on nonproliferation act as an additional (opportunity) cost to bargaining breakdown. Including these subsidies in Lines 5, 6, and 7 and then solving for the subsidy reveals that the minimum subsidy necessary for nonproliferation equilibria to exist is $\overline{\delta} - c - e \int_{\underline{\delta}}^{\overline{\delta}} T(\delta) d\delta$.

[17] Some policymakers advocating greater efforts to prevent outright theft (Hecker 2006).

5 Extensions

I now examine two extensions, motivated by substantive proliferation concerns.

5.1 Voluntary Compliance

First, formal models of nuclear proliferation often emphasize the clandestine nature of such weapons programs. That is, opposing countries cannot easily distinguish between when a state adhering to nonproliferation norms and a state secretly building a nuclear weapon (Debs and Monteiro 2014; Jelnov, Tauman, and Zeckhauser 2017). This raises a challenge in reaching a negotiated settlement. Thus far, the model has assumed that when a Proliferator agrees to a deal, it commits to the terms. In practice, though, a Proliferator could potentially enjoy the benefits of an agreement today, build weapons in private, and then enjoy the benefit later on. If a Nonproliferator cannot observe the violation, then it cannot retract the concessions to disincentivize this behavior.

To incorporate incentives to cheat on an agreement into the mechanism design framework, I now add another axiom: *voluntary compliance*. Voluntary compliance takes voluntary agreements one step further and mandates that if the Proliferator builds a weapon, it receives the settlement quantity x in the interim. That is, its proliferation payoff is now $(1-\delta)x + \delta b - c$. The $\delta b - c$ remain identical to before, whereas x is the offer that the Proliferator pretends to accept and $1-\delta$ measures the length of time between when it receives that offer and when it proliferates.[18]

Fortunately, compliance problems do not eliminate the existence of nonproliferation equilibria. However, they do place greater constraints on the parameters. Consider uncertainty over the proliferation cost and the minimum offer necessary to satisfy the lowest cost type's voluntary compliance constraint. It earns x by accepting and now $(1-\delta)x + \delta b - \underline{c}$ by building. Thus, for the lowest cost type to accept, it must be that

$$x \geq (1-\delta)x + \delta b - \underline{c}$$

$$x \geq b - \frac{\underline{c}}{\delta} \tag{10}$$

[18] The value x does not persist afterward because the Nonproliferator can withdraw the concession once it observes the violation.

Meanwhile, the Nonproliferator's constraint remains identical to Line 2. This is because voluntary compliance does not change the most the Nonproliferator is willing to give up to appease the Proliferator. That calculation is made using the Nonproliferator's payoff for agreeing with no types whatsoever, and so compliance is irrelevant. Thus, stringing together Line 2 with Line 10 and showing the conditions under which a mutually acceptable settlement exists generates:

$$b \leq \frac{c}{\delta} + \delta\epsilon \tag{11}$$

Comparing Line 11 to Line 4, the parameters for nonproliferation equilibria are harder to fulfill with voluntary compliance.

The same property is true with uncertainty over proliferation speed. Working through the formulae analogous to Lines 10 and 11 shows that the conditions under which a mutually acceptable agreement exists is:

$$\overline{\delta}b \geq c + \overline{\delta}e\left(\overline{\delta} - \int_{\underline{\delta}}^{\overline{\delta}} T(\delta)d\delta\right) \tag{12}$$

Like before, comparing Line 12 to Line 8 shows that the parameters for nonproliferation equilibria are harder to fulfill with voluntary compliance.

Both of these results have a straightforward intuition: the inability to observe the Proliferator's decision means that the Nonproliferator must pay a premium to buy compliance. This makes the Nonproliferator more inclined to abandon negotiations, thereby making nonproliferation equilibria more difficult.

Though intuitive, these general results broadly validate the nonproliferation regime's policies. One of the International Atomic Energy Agency's main tasks is to monitor the status of states' nuclear programs to ensure they comply with nonproliferation norms. The recent push for ratification of the Nonproliferation Treaty's Additional Protocol furthers their reach. Insofar as these policies provide information to states providing concessions, the mechanism design results indicate that they promote the existence of nonproliferation equilibria.[19]

[19]This section compared two extreme cases, one where a Nonproliferator could immediately observe noncompliance and another where the Nonproliferator did not observe noncompliance until the Proliferator acquired a weapon. If the IAEA is only partially effective and can only broadcast noncompliance more quickly than in its absence, nonproliferation equilibria are still easier to construct as claimed.

5.2 Sanctions

Although this mechanism design framework analyzes what is possible during negotiations, a little more structure can answer questions about how sanctions affect the process. Suppose that if bargaining fails, some subgame follows in which the Nonproliferator can sanction the Proliferator. To keep the framework general, I do not impose functional forms on that process. Rather, suppose that the sanctioning decision ends with the Nonproliferator choosing a sanctions level $s \geq 0$. The Proliferator's payoff is now $\phi(\delta, s)b - c$, where $\phi(\delta, s)$ maps a speed type and a sanctions level to a time to proliferation.[20]

I make a few sensible assumptions about the discount function. First, $\phi(\delta, 0) = \delta$, meaning that no sanctions return us to the original model. Second, $\frac{\partial \phi}{\partial s} < 0$, meaning that higher sanctions imply lower development speeds regardless of the Proliferator's baseline capability. Finally, $\frac{\partial \phi}{\partial \delta} > 0$, meaning that higher baseline speeds yield faster times to proliferation for any level of sanctions.

Meanwhile, the Nonproliferator's payoff is $-\phi(\delta, s)e - k(s)$, where $k(s)$ maps the sanctions level to a cost of implementation. Assume that $k(s) = 0$ and $k'(s) > 0$, so that no sanctions are free and more sanctions are increasingly expensive.

With uncertainty over c, the existing results carry over cleanly. Let s^* be the equilibrium quantity of sanctions the Nonproliferator would impose. Then obtaining acceptance from all types of the Proliferator requires appeasing the lowest cost type, or:

$$x \geq \phi(\delta, s^*)b - \underline{c} \tag{13}$$

This is identical to the main model's constraint but now clarifies where origins of the delay.

The Nonproliferator's constraint is a little different. Now the payment must satisfy:

$$-x \geq -\phi(\delta, s^*)e - k(s^*) \tag{14}$$

The only real change here is that Line 14 is weakly greater than Line 1. This is because $s = 0$ recovers Line 1 exactly. Thus, if the Nonproliferator employs any

[20] One could conceptualize the sanctions as lowering the chances that the proliferating leader will survive until development (Marinov 2005) or slow the shift (McCormack and Pascoe 2017).

sanctions, it must be because it performs better by doing so. Connecting Lines 13 and 14 shrinks the condition for mutually acceptable settlements but otherwise does not alter the main model's fundamental inferences.

Now consider uncertainty over proliferation speed. This presents a slight wrinkle. The value s^* could depend on δ. Let $\hat{s}$ represent the largest credible sanction that the Nonproliferator could possibly impose. Obtaining all types of the Proliferator to accept requires that appeasing the fastest type under that threat, or:

$$x \geq \phi(\overline{\delta}, \hat{s})b - c \tag{15}$$

Note that $\hat{s}$ might not correspond with the lowest type. This is because a nonproliferation equilibrium requires that the sanctions subgame never be reached. In turn, in some equilibrium, the Proliferator could adopt the off-the-path belief that results in the highest level of sanctioning even if it is not from the fastest type.

Meanwhile, let $\tilde{s}$ be the optimal sanctions level the Nonproliferator would impose if all types proliferated. Then working through the expectation, the Nonproliferator requires that:

$$-x \geq -e \int_{\underline{\delta}}^{\overline{\delta}} t(\delta)\phi(\delta, \tilde{s})d\delta - k(\tilde{s}) \tag{16}$$

Like the previous source of uncertainty, Line 16 is weakly greater than Line 6. Thus, the condition for mutually acceptable agreements is smaller but does not change the main model's fundamental inferences.

6 Conclusion

This paper introduces mechanism design to the proliferation literature to explore how uncertainty affects the bargaining process. Two principle findings emerged. First, types with faster and cheaper routes to nuclear weapons are generally advantaged in bargaining; they fare no worse than types with slower and costlier pathways and receive at least as good of nonproliferation settlements. This is because these types have superior proliferation payoffs, and indeed they develop nuclear weapons at least as often as other types. Second, nonproliferation equilibria are harder to generate when opponents do not know the development speed than when they do not know the cost of weapons. This is because uncertainty over speed affects the proliferation payoffs for both sides. Proliferators therefore have incentive to pretend to be the fastest types, while nonproliferators find generous offers less tempting in comparison to allowing all types to proliferate.

A hasty reading of the methodology may lead one to conclude that this paper is the final word on nuclear negotiations. That is not the takeaway. Although mechanism design gives general results for a class of games, three important limitations remain. First, the general principles only apply to nuclear negotiation games as I have defined them. But alternative assumptions are also worth analyzing. To wit, many researchers find a link between domestic politics and nuclear proliferation (e.g., Sagan 1996; Fuhrmann and Horowitz 2014). My setup only includes two unitary actor states. Although many domestic concerns only tweak existing parameters, others issues may not have such simple patches. In those cases, my general findings may or may not apply.

Second, the results presented here are not comparative statics in the traditional sense. Propositions 1 and 3 give monotonicity results among types *given* a type distribution and game form. Meanwhile, Propositions 2 and 4 give conditions for the existence of nonproliferation equilibria. What is missing is how changing a game's input alters qualitative outcomes of interest holding fixed the type distribution and game form. For example, when does increasing the cost proliferation decrease the probability of proliferation? This type of question is left for future manuscripts, which can develop specific game forms and use traditional equilibrium analysis.

Third, future research could explore the results of a more dynamic bargaining environment. Although the nuclear negotiations game definition allows for multiple periods of bargaining, it only permits outcomes with agreements or proliferation. It therefore helps explain that dichotomy but lacks the power to explore middling investment in a weapon that stops short of proliferation. Additional work on this would help reveal the dynamics of nuclear reversals.

7 Appendix

7.1 Proof of Proposition 1

I now prove the monotonicity results for c, in three steps.

The probability of proliferation is weakly decreasing in the cost of proliferation. Let $u(t|c)$ denote the utility for a type t for reporting c, $\pi(c)$ map a reported cost type to a probability of proliferation and $x(c)$ map a reported cost type to a settlement amount x. Incentive compatibility proves the result. Consider two types, c and $c' > c$. Satisfying incentive compatibility requires neither type wish to report as the other type. For the c type, this mandates:

$$u(c|c) \geq u(c|c')$$
$$\pi(c)(\delta b - c) + (1 - \pi(c))x(c) \geq \pi(c')(\delta b - c) + (1 - \pi(c'))x(c') \tag{17}$$

Thus, the c type must not want to report as c' and obtain the c' type's probability of proliferation and settlement amount.

Applying this same principle to the c' type gives:

$$u(c'|c') \geq u(c'|c)$$
$$\pi(c')(\delta b - c') + (1 - \pi(c'))x(c') \geq \pi(c)(\delta b - c') + (1 - \pi(c))x(c) \tag{18}$$

Adding Line 17 to Line 18 and simplifying produces:

$$\pi(c) \geq \pi(c')$$

This is the claimed result.

For all types with a positive probability of settlement, the value of the settlement is weakly decreasing in the cost of proliferation. The previous monotonicity result gave $\pi(c) \geq \pi(c')$. Thus, consider two (exhaustive) possible equilibrium cases: $\pi(c) = \pi(c')$ and $\pi(c) > \pi(c')$. First, suppose that $\pi(c) = \pi(c')$. Then Line 17 immediately reduces to $x(c) \geq x(c')$, which is the intended claim.

Second, suppose that $\pi(c) > \pi(c')$. For proof by contradiction, suppose $x(c') > x(c)$. From Line 17, incentive compatibility for the c type is violated if:

$$\pi(c')(\delta b - c) + (1 - \pi(c'))x(c') > \pi(c)(\delta b - c) + (1 - \pi(c))x(c) \tag{19}$$

Because $x(c') > x(c)$ as an assumption for the proof by contradiction, I can rewrite

$x(c')$ as $x(c) + \epsilon$, where $\epsilon > 0$. Substituting this into Line 19 and rearranging terms yields:

$$\pi(c')(\delta b - c) + (1 - \pi(c'))(x(c) + \epsilon) > \pi(c)(\delta b - c) + (1 - \pi(c))x(c)$$

$$\epsilon(1 - \pi(c')) > -(\pi(c) - \pi(c'))(x(c) - (\delta b - c))$$

Because $\pi(c') < \pi(c) \leq 1$, $1 - \pi(c') > 0$. Combined with the fact that $\epsilon > 0$, the left hand side is strictly positive. Meanwhile, $\pi(c) - \pi(c') > 0$ and (by voluntary agreements) $x(c) - (\delta b - c) \geq 0$. Thus, the right hand side is weakly negative. This demonstrates that incentive compatibility is violated, completing the proof by contradiction.

The proliferator's equilibrium utility is weakly decreasing in the cost of proliferation. I consider two cases. First, suppose both types proliferate with positive probability. For proof by contradiction, suppose $u(c') > u(c)$. Then the assumption for the proof by contradiction gives:

$$\pi(c')(\delta b - c') + (1 - \pi(c'))x(c') > \pi(c)(\delta b - c) + (1 - \pi(c))x(c) \tag{20}$$

Note that $\pi(c)(\delta b - c) + (1 - \pi(c))x(c)$ appears both here and in Line 17's incentive compatibility constraint. Stringing those inequalities together and removing that middle value yields:

$$\pi(c')(\delta b - c') + (1 - \pi(c'))x(c') > \pi(c')(\delta b - c) + (1 - \pi(c'))x(c')$$

$$c > c'$$

This violates the assumption about types, completing the proof by contradiction for this case.

Second, suppose at most one type proliferates with positive probability. I opt for a direct proof this time. The first monotonicity result implies that the c' type settles with probability 1 in this case. Substituting $\pi(c') = 0$ into Line 17 generates:

$$\pi(c)(\delta b - c) + (1 - \pi(c))x(c) \geq x(c') \tag{21}$$

The left hand side is the c type's equilibrium utility. The right hand side is the c'

type's equilibrium utility. The inequality is the desired claim.

7.2 Proof of Proposition 2

The main text proved the proposition assuming that all settlement offers must be identical in nonproliferation equilibria. I now prove that this must be true. Let $x(c)$ map a reported type c to a settlement division x and $\pi(c)$ map a reported type c to a probability of proliferation π. Recall that incentive compatibility mandates

$$\pi(c)(\delta b - c) + (1 - \pi(c))x(c) \geq \pi(c')(\delta b - c) + (1 - \pi(c'))x(c') \tag{22}$$

for all c and c'. To interpret, the left hand side of the inequality is the utility type c receives for reporting its true type, a convex combination of its proliferation payoff and the settlement the mechanism gives to type c. The right hand side is the payoff type c receives for reporting c' instead; this potentially manipulates the probability of proliferation and the settlement received but does not change type c's true cost of proliferation. Incentive compatibility requires the payoff for the truth be greater than the payoff for lying.

Proposition 2 searches for equilibria in which the probability of proliferation is 0. Substituting $\pi(c) = \pi(c') = 0$ reduces Line 22 to $x(c) \geq x(c')$.[21] But because Line 22 must be true for all c and c'—including the interchange of the two—it must also be that $x(c') \geq x(c)$. The only way to accomplish this is if $x(c) = x(c')$—that is, all types receive the exact same settlement.

7.3 Proof of Proposition 3

The probability of proliferation is weakly increasing in the development speed. Let $u(t|\delta)$ denote the utility for type t reporting δ, $\pi(\delta)$ map a reported speed type to a probability of proliferation, and $x(\delta)$ map a reported speed type to a settlement amount x. Incentive compatibility proves the first result. Consider two types, δ and $\delta' > \delta$. Satisfying incentive compatibility requires neither type wish to report as the other type. For the δ' type, this mandates:

[21] Because payoffs for agreements are the same in the baseline model and with voluntary compliance, this same step implies that the settlement must be identical across all types with voluntary compliance in nonproliferation equilibria.

$$u(\delta'|\delta') \geq u(\delta'|\delta)$$

$$\pi(\delta')(\delta' b - c) + (1 - \pi(\delta'))x(\delta') \geq \pi(\delta)(\delta' b - c) + (1 - \pi(\delta))x(\delta) \tag{23}$$

Applying this same principle to the δ type gives:

$$u(\delta|\delta) \geq u(\delta|\delta')$$

$$\pi(\delta)(\delta b - c) + (1 - \pi(\delta))x(\delta) \geq \pi(\delta')(\delta b - c) + (1 - \pi(\delta'))x(\delta') \tag{24}$$

Adding Line 23 to Line 24 and simplifying produces:

$$\pi(\delta') \geq \pi(\delta)$$

This is the claimed result.

For all types with a positive probability of settlement, the value of the settlement is weakly increasing in the development speed. The previous monotonicity result gave $\pi(\delta') \geq \pi(\delta)$. Thus, consider two (exhaustive) possible equilibrium cases: $\pi(\delta') = \pi(\delta)$ and $\pi(\delta') > \pi(\delta)$. First, suppose that $\pi(\delta') = \pi(\delta)$. Then Line 23 immediately reduces to $x(\delta') \geq x(\delta)$, which is the intended claim.

Second, suppose that $\pi(\delta') > \pi(\delta)$. For proof by contradiction, suppose $x(\delta) > x(\delta')$. From Line 23, incentive compatibility for the δ' type is violated if:

$$\pi(\delta)(\delta' b - c) + (1 - \pi(\delta))x(\delta) > \pi(\delta')(\delta' b - c) + (1 - \pi(\delta'))x(\delta') \tag{25}$$

Because $x(\delta) > x(\delta')$ as an assumption of the proof by contradiction, I can rewrite $x(\delta)$ as $x(\delta') + \epsilon$, where $\epsilon > 0$. Substituting this into Line 25 and rearranging terms yields:

$$\pi(\delta)(\delta' b - c) + (1 - \pi(\delta))(x(\delta') + \epsilon) > \pi(\delta')(\delta' b - c) + (1 - \pi(\delta'))x(\delta')$$

$$\epsilon(1 - \pi(\delta)) > -(\pi(\delta') - \pi(\delta))(x(\delta') - (\delta' b - c))$$

Because $\pi(\delta) < \pi(\delta') \leq 1$, $1 - \pi(\delta) > 0$. Combined with the fact that $\epsilon > 0$, the left hand side is strictly positive. Meanwhile, $\pi(\delta') - \pi(\delta) > 0$ and (by voluntary agreements)

$x(\delta') - (\delta' b - c) \geq 0$. Thus, the right hand side is weakly negative. This demonstrates that incentive compatibility is violated, completing the proof by contradiction.

The proliferator's equilibrium utility is weakly increasing in the development speed. I consider two cases. First, suppose both types proliferate with positive probability. For proof by contradiction, suppose $u(\delta) > u(\delta')$. Then the assumption for the proof by contradiction gives:

$$\pi(\delta)(\delta b - c) + (1 - \pi(\delta))x(\delta) > \pi(\delta')(\delta' b - c) + (1 - \pi(\delta')x(\delta') \tag{26}$$

Note that $\pi(\delta')(\delta' b - c) + (1 - \pi(\delta')x(\delta')$ appears both here and in Line 23's incentive compatibility constraint. Stringing those inequalities together and removing that middle value yields:

$$\pi(\delta)(\delta b - c) + (1 - \pi(\delta))x(\delta) > \pi(\delta)(\delta' b - c) + (1 - \pi(\delta))x(\delta)$$

$$\delta > \delta'$$

This violates the assumption about types, completing the proof by contradiction for this case.

Second, suppose at most one type proliferates with positive probability. I opt for a direct proof this time. The first monotonicity result implies that the δ type settles with probability 1 in this case. Substituting $\pi(\delta) = 0$ into Line 23 generates:

$$\pi(\delta')(\delta' b - c) + (1 - \pi(\delta')x(\delta') \geq x(\delta) \tag{27}$$

The left hand side is the δ' type's equilibrium utility. The right hand side is the δ type's equilibrium utility. The inequality is the desired claim.

7.4 Proof of Proposition 4

The main text proved the proposition assuming that all settlement offers must be identical in nonproliferation equilibria. I now prove that this must be true. Let $x(\delta)$ map a reported type δ to a settlement division x and $\pi(\delta)$ map a reported type δ to a probability of proliferation π. Recall that incentive compatibility mandates

$$\pi(\delta)(\delta b - c) + (1 - \pi(\delta))x(\delta) \geq \pi(\delta')(\delta b - c) + (1 - \pi(\delta'))x(\delta') \tag{28}$$

for all δ and δ'. To interpret, the left hand side of the inequality is the utility type δ receives for reporting its true type, a convex combination of its proliferation payoff and the settlement the mechanism gives to type δ. The right hand side is the payoff type δ receives for reporting δ' instead; this potentially manipulates the probability of proliferation and the settlement received but does not change type δ's true cost of proliferation. Incentive compatibility requires the payoff for the truth be greater than the payoff for lying.

Proposition 4 searches for equilibria in which the probability of proliferation is 0. Substituting $\pi(\delta) = \pi(\delta') = 0$ reduces Line 28 to $x(\delta) \geq x(\delta')$. But because Line 28 must be true for all δ and δ'—including the interchange of the two—it must also be that $x(\delta') \geq x(\delta)$. The only way to accomplish this is if $x(\delta) = x(\delta')$—that is, all types receive the exact same settlement.

7.5 Additional Results: Welfare Gains from Bargaining

I now provide additional results on who benefits more from nonproliferation agreements:

Proposition 6. *Suppose the proliferation cost c is private information but the development speed δ is common knowledge. Let s^* be any nonproliferation equilibrium of a game form G. Then the Proliferator's welfare gains through negotiation strictly increase in the cost of proliferation.*

Define the welfare gains of negotiation as the equilibrium utility a Proliferator receives minus its proliferation payoff. Proposition 6 states that lower cost types do not benefit from the ability to negotiate as much as higher cost types. To see this, recall that guaranteed nonproliferation requires a single settlement for all types. Call that amount x. Then the difference is $x - (\delta b - c)$. This is clearly increasing in c. In words, if all types receive the same settlement, then the type with the highest cost enjoys the settlement more than anyone else.

Although Proposition 6 is mathematically apparent, it tempers interpretations of negotiation outcomes. Traditional bargaining theory indicates that types with better outside options perform better within negotiated settlements. In regard to the actual settlement terms, such types do not fare any worse. But they also cannot leverage a proliferation threat to do better than others when nonproliferation is guaranteed. In turn, higher cost types outperform their circumstances more than lower cost types.

An analogous result applies for uncertainty over development speeds:

Proposition 7. *Suppose the development speed δ is private information but the proliferation cost c is common knowledge. Let s^* be any nonproliferation equilibrium of a game form G. Then the Proliferator's welfare gains through negotiation strictly increase in the development speed.*

The proof and commentary are similar to above.

7.6 Ease of Nonproliferation Equilibria When Preventive War Is Credible

The main paper claimed that obtaining game forms with nonproliferation equilibria is trivial when preventive war is credible. I demonstrate that here.

Consider the following simple extensive form. The Nonproliferator begins by making an offer x. the Proliferator accepts or builds in response. Accepting implements the proposed transfer. Building forces the Nonproliferator to choose between allowing proliferation to occur and fighting a preventive war. Proliferation results in the payoffs $\delta b - c$ for the Proliferator and $-\delta e$ for N.[22] Preventive war results in war payoffs $-w_P - c$ for the Proliferator and $-w_N$ for N, where $w_i > 0$. Note that this matches the definition of a nuclear negotiations game, except for the addition of a preventive war outcome.

Solving for the complete information game's equilibrium illustrates how credible preventive war eliminates the informational incentives described in the main text. By backward induction, the Nonproliferator prevents if $-w_N > -\delta e$, or $w_N < \delta e$.

Suppose this condition holds. Then the Proliferator's build decision becomes straightforward. It earns x by accepting and $-w_P - c$ for build-induced preventive war. Because x is weakly positive, the Proliferator must accept all offers.

Moving to the initial step, the Nonproliferator also has a clear optimal course of action. No matter what it offers in this first stage, the Proliferator accepts. The

[22]The Proliferator's preventive war payoff includes a wasted investment cost in addition to whatever else it suffers from conflict. This is consistent with standard specifications, but the claims I make here are not sensitive to its inclusion.

Nonproliferator's payoff is strictly decreasing in the transfer, so it offers the minimum transfer: $x = 0$.

As it relates to informational dynamics, note that no step in the process here relied on the Nonproliferator knowing the Proliferator's specific cost of proliferation. No matter how c is distributed and no matter how e relates to b, a deal exists. It is, however, trivial.

A similar story holds for uncertainty over δ, with the caveat from the main text. Suppose $w_N < \underline{\delta}e$. Then the Nonproliferator prefers preventing against all types. It can therefore still credibly threaten preventive war. Internalizing that, the Proliferator accepts all offers, and the Nonproliferator strikes a (trivial) deal at the beginning.

In short, credible preventive war allows the Nonproliferator to ignore its information problem. It does not have to work through any risk-return tradeoff because it can instead leverage the threat of war to force the Proliferator into compliance. Game forms with nonproliferation equilibria—like this one—are therefore easy to develop. The case where preventive war is not credible inverts these incentives and generates the results from the main paper.

8 Works Cited

Ahmed, Samina. 1999. "Pakistan's Nuclear Weapons Program: Turning Points and Nuclear Choices." *International Security* 23 (4): 178-204.

Banks, Jeffrey S. 1990. "Equilibrium Behavior in Crisis Bargaining Games." *American Journal of Political Science* 34 (3): 599-614.

Bas, Muhammet A. and Andrew J. Coe. 2012. "Arms Diffusion and War." *Journal of Conflict Resolution* 56 (4): 651-674.

Bas, Muhammet A. and Andrew J. Coe. 2016. "A Dynamic Theory of Nuclear Proliferation and Preventive War." *International Organization* 70 (4): 655-685.

Bas, Muhammet A. and Andrew J. Coe. 2017. "Give Peace a (Second) Chance: A Theory of Nonproliferation Deals." *International Studies Quarterly* Forthcoming.

Beardsley, Kyle and Victor Asal. 2009. "Winning with the Bomb." *Journal of Conflict Resolution* 53 (2): 235-55.

Benson, Brett V. and Quan Wen. 2011. "A Bargaining Model of Nuclear Weapons:

Development and Disarmament." In *Causes and Consequences of Nuclear Proliferation* Eds. Robert Rauchhaus, Matthew Kroenig, and Erik Gartzke. New York: Routledge Press.

Bleek, Philipp C. and Eric B. Lorber. 2013. "Security Guarantees and Allied Nuclear Proliferation." *Journal of Conflict Resolution* 58 (3): 429-454.

Brown, Robert L. and Jeffrey M. Kaplow. 2014. "Talking Peace, Making Weapons: IAEA Technical Cooperation and Nuclear Proliferation." *Journal of Conflict Resolution* 58 (3): 402-428.

Burr, William and Jeffrey T. Richelson. 2000. "Whether to 'Strangle the Baby in the Cradle': The United States and the Chinese Nuclear Program, 1960-64." *International Security* 25 (3): 54-99.

Caldwell, Dan. 1987. "Permissive Action Links: A Description and Proposal." *Survival* 29 (3): 224-238.

Campbell, Kurt M. and Tsuyoshi Sunohara. 2004. "Japan: Thinking the Unthinkable." In *The Nuclear Tipping Point: Why States Reconsider Their Nuclear Choices* Eds. Kurt M. Campbell, Robert J. Einhorn, and Mitchell B. Reiss. Washington D.C.: Brookings Institution Press.

Chadefaux, Thomas. 2011. "Bargaining over Power: When Do Shifts in Power Lead to War?" *International Theory* 3 (2): 228-253.

Chinoy, Mike. 2008. *Meltdown: The Inside Story of the North Korean Nuclear Crisis* New York: St. Martin's Press.

Debs, Alexandre and Nuno Monteiro. 2014. "Known Unknowns: Power Shifts, Uncertainty, and War." *International Organization* 68 (1): 1-31.

Debs, Alexandre and Nuno Monteiro. 2017. *Nuclear Politics: The Strategic Causes of Proliferation.* Cambridge: Cambridge University Press.

Drezner, Daniel W. 1999. *The Sanctions Paradox: Economic Statecraft and International Relations.* Cambridge: Cambridge University Press.

ElBradei, Mohammed. 2011. *The Age of Deception: Nuclear Diplomacy in Treacherous Times.* New York: Metropolitan Books.

Fang, Songying. 2008. "The Informational Role of International Institutions and Domestic Politics." *American Journal of Political Science* 52(2): 304-21.

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49 (3): 379-414.

Fey, Mark and Kristopher W. Ramsay. 2011. “Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict.” *American Journal of Political Science* 55 (1): 149-169.

Fuhrmann, Matthew. 2012. *Atomic Assistance: How “Atoms for Peace” Programs Cause Nuclear Insecurity.* Ithaca: Cornell University Press.

Fuhrmann, Matthew and Michael C. Horowitz. 2015. “Leaders, Rebel Experience, and Nuclear Proliferation.” *Journal of Politics* 77 (1): 72-87.

Gaddis, John Lewis. 1997. *We Now Know: Rethinking Cold War History.* Oxford: Oxford University Press.

Gordin, Michael D. 2009. *Red Cloud at Dawn: Truman, Stalin, and the End of the Nuclear Monopoly.* New York: Farrar, Straus and Giroux.

Hecker, Siegfried S. 2006. “Toward a Comprehensive Safeguards System: Keeping Fissile Materials out of Terrorists’ Hands.” *The ANNALS of the American Academy of Political and Social Science* 607 (1): 121-132.

Holloway, David. 1994. *Stalin and the Bomb: The Soviet Union and Atomic Energy, 1939-1956.* New Haven: Yale University Press.

Hymans, Jacques E.C. 2006. *The Psychology of Nuclear Proliferation: Identity, Emotions and Foreign Policy.* New York: Cambridge University Press.

Jelnov, Artyom, Yair Tauman, and Richard Zeckhauser. 2017. “Attacking the Unknown Weapons of a Potential Bomb Builder: The Impact of Intelligence on the Strategic Interaction.” *Games and Economic Behavior* 104: 177-189.

Jo, Dong-Joon and Erik Gartzke. 2007. “Determinants of Nuclear Weapons Proliferation.” *Journal of Conflict Resolution* 51 (1): 167-194.

Kaplow, Jeffrey M. and Erik Gartzke. 2016. “The Determinants of Uncertainty in International Relations.” Manuscript, College of William and Mary.

Koremenos, Barbara, Charles Lipson, and Duncan Snidal. 2001. “The Rational Design of International Institutions.” *International Organization* 55(4): 761-99.

Khan, Feroz. 2012. *Eating Grass: The Making of the Pakistani Bomb.* Stanford: Stanford University Press.

Kroenig, Matthew. 2009. “Importing the Bomb: Sensitive Nuclear Assistance and Proliferation.” *Journal of Conflict Resolution* 53 (2): 161-180.

Lanoszka, Alexander. 2013. “Protection States Trust? Major Power Patronage, Nuclear Behavior, and Alliance Dynamics.” Manuscript, Princeton University.

Levite, Ariel E. 2003. "Never Say Never Again: Nuclear Reversal Revisited." *International Security* 27 (3): 59-88.

Marniov, Nikolay. 2005. "Do Economic Sanctions Destabilize Country Leaders?" *American Journal of Political Science* 49 (3): 564-576.

Martin, Curtis H. 2002. "Rewarding North Korea: Theoretical Perspectives on the 1994 Agreed Framework." *Journal of Peace Research* 39 (1): 51-68.

McCormack, Daniel and Henry Pascoe. 2017. "Sanctions and Preventive War." *Journal of Conflict Resolution* 61 (8): 1711-1739.

Montgomery, Alexander H. and Adam Mount. 2014. "Misestimation: Explaining US Failures to Predict Nuclear Weapons Programs." *Intelligence and National Security* 29 (3): 357-386.

Myerson, Roger B. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47 (1): 61-74.

Myerson, Roger B. and Mark A. Satterthwaite. 1983. "Efficient Mechanisms for Bilateral Trading." *Journal of Economic Theory* 29 (2): 265-281.

Paul, T.V. 2000. *Power versus Prudence: Why Nations Forgo Nuclear Weapons.* Montreal: McGill-Queen's University Press.

Reiss, Mitchell B. 1988. *Without the Bomb: The Politics of Nuclear Nonproliferation.* New York: Columbia University Press.

Richelson, Jeffrey T. 2007. *Spying on the Bomb: American Nuclear Intelligence from Nazi Germany to Iran and North Korea.* New York: Norton

Rublee, Maria Rost. 2009. *Nonproliferation Norms: Why States Choose Nuclear Restraint.* Athens: University of Georgia Press.

Sagan, Scott D. 1993. *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons.* Princeton: Princeton University Press.

Sagan, Scott D. 1996. "Why Do States Build Nuclear Weapons?: Three Models in Search of a Bomb." *International Security* 21 (3): 54-86.

Schelling, Thomas C. 1960. *The Strategy of Conflict.* Cambridge: Harvard University Press.

Schwartz, Stephen I. 1998. *Atomic Audit: The Cost and Consequences of U.S. Nuclear Weapons since 1940.* Washington D.C.: The Brookings Institute.

Sechser, Todd S. and Matthew Fuhrmann. 2017. *Nuclear Weapons and Coercive Diplomacy.* Cambridge: Cambridge University Press.

Silverstone, Scott A. 2007. *Preventive War and American Democracy.* New York: Routledge.

Singh, Sonali and Christopher R. Way. 2004. “The Correlates of Nuclear Proliferation: A Quantitative Test.” *Journal of Conflict Resolution* 48 (6): 859-885.

Spaniel, William. 2015. “Arms Negotiations, War Exhaustion, and the Credibility of Preventive War.” *International Interactions* 41 (5): 832-856.

Spaniel, William. 2018. “Only Here to Help? Bargaining and the Perverse Incentives of International Institutions.” *International Studies Quarterly* 62 (1): 14-22.

Thayer, Bradley A. 1995. “The Causes of Nuclear Proliferation and the Utility of the Nuclear Non-Proliferation Regime.” *Security Studies* 4 (3): 463-519.

Volpe, Tristan A. 2017. “Atomic Leverage: Compellence with Nuclear Latency.” *Security Studies* 26 (3): 517-544.