Give Peace a (Second) Chance: A Theory of Nonproliferation Deals

Muhammet A. Bas and Andrew J. Coe*

Abstract

We formally analyze the possibilities for a deal to avoid proliferation or conflict between a state that might develop nuclear weapons and another that might use force or diplomacy to stop it. We find that the prospects for a deal are best either when the proliferant's program is rudimentary, or when it has made enough progress that the observing state believes its fruition is close. Deals at the latter point are underpinned by a newly credible threat to attack or abandon the proliferant, and yet are more generous than earlier deals. The viability of an early deal can be undermined by the anticipation of a better, later one, and a proliferant might continue a program as much to secure a better agreement as to actually obtain nuclear weapons. We test and find support for the model's predictions against the historical record of dealmaking over states' nuclear programs.

Word Count: 12,983

*We are grateful to Robert Powell, Dustin Tingley, and Jane Vaynman for comments on earlier versions of this paper, and to Målfrid Braut-Hegghammer, Taylor Dalton, Ryan Musto, and Jane Vaynman for help with particular empirical cases. Andrew Coe thanks the Stanton Foundation for its support of this research through its Nuclear Security Fellowship, and the Council on Foreign Relations for hosting the author during his fellowship.

Introduction

Since the nuclear era began, the United States has negotiated with other states to try to end their possible pursuit of nuclear weapons. Sometimes a nonproliferation deal is made, and other times not. Some deals last, while others are only temporary. Sometimes negotiations are refused for years, only to be renewed and a deal agreed just as the risk of conflict seems highest. Why are deals made in some instances but not others, and why are the negotiations sometimes so protracted?

Policymakers considering such negotiations have difficult decisions to make. Given that the diplomatic effort is costly, when is the best time to gather international support and try to make a deal? If a previously truculent proliferant suddenly becomes willing to negotiate, is this real, or just a tactic intended to buy time for its nuclear program to succeed? Is there a point beyond which further negotiations should be discarded in favor of unilateral solutions? Most topically, is the current deal with Iran merely a smokescreen?

We try to answer these questions by analyzing and testing a formal model in which two states bargain over disputed issues, while one (the "proliferant") potentially invests in and makes progress toward acquiring nuclear weapons that, once deployed, would increase its bargaining power, and the other (the "US," though other states have played this role) imperfectly observes its investment and progress. The US may initiate costly conflict—attacking an enemy proliferant, or abandoning an allied one—in reaction to the proliferant's efforts. Because weapons development and observation occur over time, the model enables us to understand and make predictions about the sources of empirical variation in behavior across both countries and time. It also allows us to analyze the possibilities for policy-makers to shift the underlying factors and thereby improve outcomes.

In the absence of a deal, the proliferant invests in a nuclear weapons program in the hopes that acquiring the weapons will enable it to extract concessions from the US. An eventual outcome of costly conflict or proliferation can only be avoided if the two sides agree to a deal. In such a deal, the US makes concessions in exchange for the proliferant ending its pursuit of nuclear weapons. We find that the viability of a deal is intimately linked to the willingness of the US to eventually resort to costly conflict in its absence. If the US is willing to attack or abandon the proliferant, then it will also be willing to make more concessions in pursuit of a deal that would avert the costs of conflict. If the US would not attack or abandon the proliferant, then it also would not concede enough to render a deal viable.

We demonstrate that there are essentially two points in time at which a nonproliferation deal is most likely to be viable, referred to as "early" and "late." One occurs when a state's nuclear program or intent is first suspected and at an early stage. The US cannot credibly threaten to attack or abandon the proliferant at this point, so that the penalty for reneging on a deal is modest, making compliance harder to enforce. However, the proliferant will be less tempted to renege, since its rudimentary program would take a long time to succeed, so that it is relatively easily "bought off." The second opportunity for a deal occurs when the US believes a state's nuclear program is close enough to fruition that the US would rather attack or abandon it than tolerate its continued progress. At this point, the US threat is credible, so that the punishment for cheating on the deal is severe, but the temptation for the proliferant to renege and rush to complete its program is higher. At any point between these two, uncertainty about the extent of the program's progress renders the US threat to attack or abandon it not credible, but the temptation for the proliferant to cheat might be high due to the unseen progress of its efforts, so that a deal is less likely to be viable.

The possibility of a late deal, agreed when the proliferant's program is thought to be nearing success, carries two important implications. First, a proliferant's sudden willingness to negotiate, even after years of pursuing nuclear weapons and dissembling about it, may be genuine, in the sense that the proliferant expects that a deal to end its program that it would willingly comply with has now become possible. Second, and more insidiously, the proliferant's earlier expectations about US attitudes toward a late deal may undermine the possibility of an earlier deal and encourage proliferation. If pursuing a weapons program is expected to lead to a late deal, rather than attack or abandonment by the US, a proliferant will be less inclined to accept a deal at earlier points and more inclined to pursue a program in the first place. Thus, while a late deal like the one with North Korea in 1994 offers the hope of eliminating the possibility of proliferation and war in that case, it may encourage other states to continue a nuclear weapons program, knowing that they can hold out for a better deal once the program has advanced.

This provides an explanation for the puzzling phenomenon of states that possess the technological prerequisites for nuclear weapons, but have chosen not to build them (e.g., Japan). If a state is not going to actually build nuclear weapons, and reap the security benefits therefrom, why would it pursue the technology in the first place? The model implies that nuclear latency can arise from (possibly covert or implicit) deals between the US (or other relevant interlocutors) and other states aimed at avoiding proliferation. These states benefit from nuclear latency because it forces the US to offer greater concessions in exchange for their restraint.

To test the theory, we gather data on all meaningful nonproliferation deals in the historical record. This data reveals some surprising patterns. Of the 30 states that were judged capable of pursuing a nuclear weapons program, 27 (90%) agreed to a nonproliferation deal at some point, and 23 (85%) of those states made an early deal. Of the ten states that had the opportunity to make a late deal, 6 (60%) did so. Overall, a deal is made in 74% of the instances that our theory identifies as opportune times for a deal. Because deals are usually made, we focus our testing on our theory's predictions about the timing of, concessions made in, and implicit threat behind these deals, as opposed to whether a deal occurs.

We find substantial support for the theory. First, 29 out of the 32 deals in the dataset occur either early in a proliferant's program, when it is known to be rudimentary or not yet begun, or late, once it is estimated to be nearing success, and not in between. Second, all six late deals coincide with a newly-credible US threat to attack or abandon the proliferant in the absence of a deal. Third, four of the five instances of proliferants that made both early and late deals received more generous concessions in the latter.

Ours is, to our knowledge, the first analysis of all known nonproliferation deals. Some previous formal work has pointed out the theoretical possibility of bilateral nonproliferation deals, but not analyzed them in detail or studied them empirically (Bas and Coe, 2016; Debs and Monteiro, 2014; Spaniel, 2015).[1] Both these studies and the closely-related formal literature on conventional arming focus on determining the conditions under which arming/proliferation or war to prevent it will occur, with the latter typically setting aside the potential for avoiding both through a deal.[2] We contribute to this body of knowledge by explicitly incorporating the possibility for such deals and analyzing their viability, timing, and content.

Other works formally analyze the incentives behind the nonproliferation regime and its enforcement, but do not consider those deals made after the regime was established (Coe and Vaynman, 2015; Verdier, 2008). A rich empirical literature studies particular later deals but not the whole set (Braut-Hegghammer, 2008; Budjeryn, 2016; Coutto, 2014; Miller, 2014; Reardon, 2010; Volpe, 2015). Our theory generalizes these previous studies, offering a set of empirical implications that should hold for all nonproliferation deals, and largely do.

Our model is among the first to treat states' arming dynamically—that is, new military capabilities are acquired over time rather than immediately—something we show is essential

[1]Debs and Monteiro (2014) shows that its key comparative statics about preventive war are robust to the possibility of deals, and that war may not be avoidable even if the players are very patient. Spaniel (2015) demonstrates that changes in the cost of preventive war may rule out any deal, and uses this to explain why the US did not make a nonproliferation deal with the Soviet Union.

[2]The formal literature on conventional arming is reviewed in Fearon (2011).

to understanding the history of deal-making over proliferation.[3] Acquiring a consequential new military capability often requires a substantial, but uncertain, length of time for research, development, and construction before the capability can be deployed. Consequently, a state that is attempting to monitor another's arming may be uncertain not only about whether a new capability is being sought, but also about precisely how soon this capability will be ready, and thus when it is no longer to safe to put off the costs either of conflict or of the concessions needed to support a deal.

Previous models take arming to be static—capabilities are acquired immediately rather than over time—and so cannot speak to the unfolding of these interactions over time (Baliga and Sjöström, 2008; Benson and Wen, 2011; Debs and Monteiro, 2014; Feaver and Niou, 1996; Jackson and Morelli, 2009; Meirowitz and Sartori, 2008; Powell, 1993).[4] In models of nuclear proliferation, this approach typically leads to one state randomizing over whether to seek nuclear weapons, so that the key uncertainty driving behavior is over the existence of a program, while the other state randomizes over whether to attack to stop it (Baliga and Sjöström, 2008; Benson and Wen, 2011; Debs and Monteiro, 2014). However, in many cases there is little empirical evidence of uncertainty over a program's existence—as opposed to its progress—at any time when attack might be seriously considered (Bas and Coe, 2016). Moreover, these models produce only probabilistic predictions for what should happen in most empirical cases.

By contrast, in our model the key uncertainty is over the progress of a program, which more plausibly arises from technological trial-and-error and imperfect intelligence-gathering: neither weapons programs nor preventive attack are randomly chosen.[5] Our theory holds that program progress and intelligence estimates of it explain why deals occur at some times but not others. As a result, given the actual and estimated progress of a program, our model yields deterministic predictions for all empirical cases, which we show are born out in most.

[3] A working paper by Fearon (2011) also does so, but is focused on quantitative arms racing rather than the development of qualitatively new capabilities. Bas and Coe (2012) studies proliferation in a dynamic context, but does not actually analyze arming since proliferation is taken to be exogenous.

[4] A partial exception is Spaniel (2015), which takes arming to be static but allows variation over time due to exogenous changes in the cost of preventive attack.

[5] More technically, previous models typically yield the behavior of empirical interest only in mixed-strategy equilibria, while our model produces the interesting behavior in pure-strategy equilibria.

Model Setup

Two states, A (the "US," referred to as feminine) and B (the "proliferant," masculine), bargain over revisions to a prior division of a composite of disputed issues, represented by the unit interval.[6] In the first of infinitely many discrete periods of time, A first chooses whether or not to initiate costly conflict. If A does so, the game ends and A receives a fraction of the contested stake corresponding to her power relative to B, while B receives the rest. Each player pays a positive cost of conflict, c_A and c_B respectively, in this and all future periods.

"Costly conflict" is intended to represent the most cost-effective option available to A for unilaterally responding to B's program. If B is an adversary, this may be a full-on invasion, such as the US war with Iraq in 2003, or instead a limited attack intended to destroy key nuclear facilities, as with Israel's attacks on reactors in Iraq and Syria. If instead B is an ally, this may entail ending the alliance, as the US threatened to do with South Korea, Taiwan, and others. Which of these is chosen will depend on the costs of each option and its anticipated effectiveness at responding to B's pursuit of nuclear weapons, but this choice is not the focus of our analysis.

Attack or abandonment are obviously very different responses empirically, but from the point of view of our model, they share all the relevant strategic features. First, each imposes serious costs on both sides. In the case of attack, death and destruction will result. In the case of abandonment, the two sides will suffer the loss of the benefits of alliance—perhaps especially the deterrence of an outside enemy. Second, the outcome of attack or abandonment both depend on the balance of power between the two sides. If the proliferant is an enemy, this balance corresponds to the likely outcome of a war between the two. If he is instead an ally, the balance is often determined by how each side would do without the alliance in a war against their common enemy.[7] Finally, each response might prevent or at least delay the proliferant's acquisition of nuclear weapons, directly in the case of attack and indirectly in the case of abandonment, which would expose the proliferant to attack by an enemy.

If A chooses not to initiate costly conflict, then she must offer to B a disposition of the

[6]It may seem that the principal disputed issue in interactions like that between the US and Iran is precisely the latter's possible nuclear weapons program. However, this dispute arises only because there are underlying contested issues—such as influence over other states in the region—whose settlement would be affected by Iran's acquisition of nuclear weapons.

[7]Our representation will be valid so long as abandonment will result in some partition of the alliance's total value between the two sides, less the joint benefit created by the alliance itself.

contested issues. If B rejects the offer, costly conflict results, ending the game with the same outcome as above. If he accepts the offer, the revision is implemented immediately and the associated payoffs are realized.

This take-it-or-leave-it bargaining protocol offers a simple way to model the imposition of economic and political sanctions, which is often undertaken in response to states' nuclear programs. Sanctions reduce the value B receives from international commerce and political influence, and so are akin to A making an offer that is less generous to B than the status quo. For simplicity, we ignore any cost incurred by A in imposing sanctions.[8]

Peaceful acceptance by B of A's offer is followed by an opportunity for B to invest in developing nuclear weapons (i.e., to start a program or continue an extant one). To simplify the analysis, we assume that B's development effort is all or nothing—the choice to pursue nuclear weapons is binary.[9] We also abstract away from any direct (budgetary) cost of B's investment, taking this to be negligible.[10]

B must master a series of technological prerequisites before he can actually deploy nuclear weapons. For simplicity of presentation, we assume there are only two prerequisites, which we take to be the production of fissile material in sufficient quantity, and the manufacture of viable weapons. Thus, there is a first or "rudimentary" stage of development where B has mastered neither, a second or "advanced" stage where B has mastered the production of fissile material but not the manufacture of weapons, and a third stage n where B has mastered both and is assumed to possess nuclear weapons.[11] B begins the game in the first stage.

The overcoming of these hurdles is partly a result of trial-and-error, so that the time at which B will master one and then the next cannot be perfectly predicted by either player. If B begins a round in the second stage and chooses to invest in that round, then he remains

[8]Incorporating a cost makes a deal more likely to be viable, since avoiding this cost would increase the surplus available to support a deal.

[9]It can be shown that, if A's ability to monitor the size of B's investment is low enough, then B will never choose an intermediate level of investment in equilibrium. It seems empirically plausible to assume that it would be very hard for A to observe the size of B's investment, as opposed to observing whether a program was underway and whether the program had made tangible progress, and thus that B would only choose all-out effort, or none at all.

[10]Incorporating such a cost has the same effects as a cost for sanctions, with one obvious exception: if the program cost is high enough, then B will not start a program and A will never resort to costly conflict, with or without a deal.

[11]More realistically, there may be more than two potentially observable stages to a nuclear weapons program. We will see later that the key is simply that there is a late-enough stage at which A would resort to costly conflict in the absence of a deal, and earlier stages in which A would not.

at the second stage with probability $1 - \lambda$, and advances to acquiring nuclear weapons in that round with probability λ. If B begins a round in the first stage and invests, then he remains at the first stage with probability $1 - \epsilon$, advances to the second stage in that round with probability ϵ, and advances all the way to acquiring nuclear weapons in that round with probability $\epsilon\lambda$, so that it is possible to master both stages in a single round.[12]

This representation of the weapons development process is the central analytical innovation of this model; many of our results flow from it. It has two virtues. First, it is the simplest possible representation of the empirical fact that the development of any complex technology is both progressive and stochastic. Second, it is the simplest way to generate a quintessential feature of the empirical interactions we are interested in: the US's intense focus on estimates of when a proliferant will get nuclear weapons. Although B's chances of advancing to a given stage depend only on his current stage and his decision to continue trying, as time goes by, his probability of acquiring nuclear weapons will increase, and in the absence of contradicting intelligence, A's *estimate* of his time to acquiring nuclear weapons will decrease.

If B's development effort succeeds and he acquires nuclear weapons, this immediately becomes common knowledge (e.g., because of an easily observable test detonation). The balance of power shifts in the next period, from p to $p_n < p$, in B's favor. If B is an enemy of A, this assumption means B would do better with nuclear weapons than without in a war with A.[13] If instead B is A's ally, this assumption means that B would do better with nuclear weapons than without if the alliance ended. This assumption—that nuclear weapons increase B's power—is crucial to all that follows. If they do not, then B has no incentive to acquire them, A has no incentive to prevent this, and there is no reason for a deal. Despite the ongoing debate over this assumption, most US policymakers seem to believe it (Gavin, 2012), and there is some evidence that states seeking nuclear weapons do also (Brands and Palkki, 2011; Narang, 2014).

The first period ends after B's progress or lack thereof is determined. The next period,

[12]As time goes by and the proliferant remains at a given stage, the probability of mastering that stage would, more realistically, increase. Allowing for this would not change the results, which only depend on the fact that the program's success becomes more likely over time.

[13]Powell (2015) develops a model in which the risk of nuclear escalation causes a state facing a nuclear-armed opponent to bring less power to bear in a war, lowering the former's probability of winning—just as we assume here. Note that this effect occurs even if the war is over a limited stake, for which neither side would be willing to use nuclear weapons. Nuclear weapons might also alter the costs of conflict; incorporating this would not qualitatively change the results.

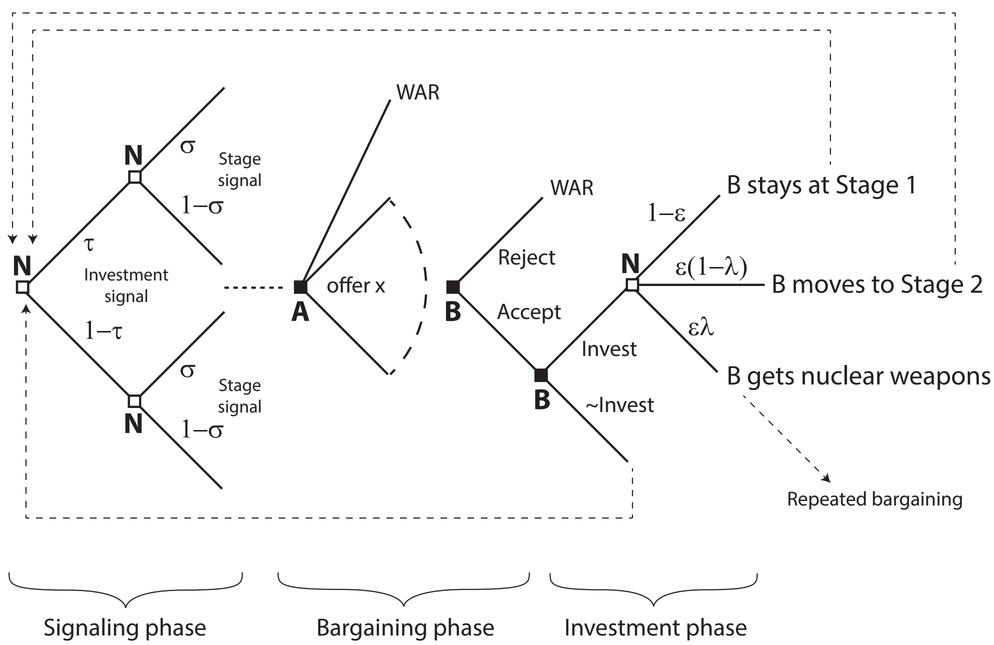

Figure 1: A single period of the game

and every subsequent period, differs in structure from the first only in that it begins with the possible reception of new intelligence by A. This takes the form of two signals, which are assumed to be common knowledge.[14] The first signal, of the program's existence, indicates whether B invested in the last period or not. If B did invest, then with probability τ A receives a signal that he did, and with probability $1-\tau$ A receives a signal that he did not. If B did not invest, then A receives a signal that he did not with probability 1. Thus, A's intelligence on B's investment is noisy, but for simplicity there are no false positives.[15] The second signal, of the program's progress, indicates the current stage of B's program. A will receive a true signal of B's current stage with probability σ, and an uninformative (i.e., "null") signal with probability $1-\sigma$. Thus, A's intelligence on B's progress is spotty, but accurate.[16] Figure 1 illustrates one such period.

Each player's per-period payoff is assumed to be linear in her or his share of the value of the contested issues, while future payoffs are discounted by a factor $\delta < 1$ per period. Players' preferences and all the exogenous parameters of the game are common knowledge.[17]

What Happens in the Absence of a Deal

The driving force in the interaction between the two sides is that, if B acquires nuclear weapons, he can use the resulting increase in his bargaining power to extract more generous concessions from A than he would otherwise receive. Thus, B has an incentive to pursue nuclear weapons and A has an incentive to prevent this. There are only two ways to stop B's program: A can initiate costly conflict against B when she feels the time is right, or negotiate a deal in which she offers B concessions in exchange for him not pursuing nuclear weapons. Any such deal must be enforced by A's threat to retract concessions and possibly eventually resort to costly conflict if B reneges. So, before we determine whether any deal is viable, we must specify what would happen in the absence of a deal, or if a deal fails—

[14] In the equilibria studied here, A never has any incentive to conceal these signals. Empirically, the US has strong incentives to credibly reveal its intelligence on the program in order to build international support for action against the proliferant.

[15] False positive signals of investment may lead A to end a deal she mistakenly believed B had cheated on. The anticipation of this possibility lowers the value of a deal for both, making it less likely to be viable.

[16] A low probability of false second-stage signals would make costly conflict more likely in the absence of a deal, increasing the surplus available to support a deal. A low probability of false first-stage signals has the opposite effect.

[17] The online appendix contains a detailed description of the differences between our model and others of proliferation.

what we term a "no-deal" equilibrium. Our brief description here summarizes the formal analysis presented in the online appendix, focusing on the results necessary to understand the subsequent analysis of deals.

In the absence of a deal, A makes offers just generous enough to avoid B's rejection, as more generous offers would only be worthwhile under a deal, to avoid proliferation. In response, B invests in a program in the hopes of eventually getting nuclear weapons and forcing A to offer more in recognition of the altered balance of power.

In the meantime, A faces a tradeoff between the risk of proliferation and the costs of attacking or abandoning B. Early on, B's program is unlikely to succeed any time soon, so that A is content to watch and wait, putting off the costs of conflict.[18] As time passes, B's program is increasingly likely to have made progress, and without new intelligence that it has not, A's estimate of the time until it succeeds will shorten.[19] Thus, the risk of imminent proliferation grows, and A's tradeoff begins to shift in favor of costly conflict. If the sophistication of B's program is low enough, so that even in the penultimate stage of development, its success is still distant, and the shift in power that proliferation would effect is small enough relative to the costs of conflict, then A would simply tolerate B's program indefinitely, and eventual proliferation is inevitable. Otherwise, A's estimate of the time until B has weapons will eventually grow short enough that A would rather act than tolerate any further risk of proliferation.[20] Generally, the greater the effect of proliferation on the balance of power, the lower the costs of conflict, and the higher the sophistication of B's program, the sooner A will resort to conflict.

Given that A would be willing to initiate costly conflict at some point, chance determines whether the game ends in proliferation or conflict. With luck, B's program will advance

[18]Of course, there are empirical instances of preventive attacks undertaken or seriously considered relatively early in a program. Bas and Coe (2016) shows that most such instances occur during an ongoing war between the proliferant and attacker, when the costs of attack are temporarily greatly reduced. Because a nonproliferation deal is unlikely to be agreed between two sides already at war, we ignore this qualification to our picture of what happens in the absence of a deal.

[19]In our model, A may be uncertain only about the existence and progress of B's program (i.e., which stage it is at). In reality, A may also be uncertain of its sophistication (i.e., how quickly it is expected to master each stage). If so, then as time passes, A's estimate of the time until proliferation occurs may rise or fall, but we believe equilibrium behavior would be qualitatively the same: A waits to act until she is confident enough that proliferation is near.

[20]This conclusion holds even if A's action does not permanently end B's program. A limited attack on the program's facilities, for instance, might only delay the program, but the closer it is to success, the more facilities there are to be rebuilt after a strike and so the larger the delay A can impose by attacking, again leading her to postpone action until she estimates its success is near.

quickly enough to succeed before A elects to act. Otherwise, A will grow confident enough that proliferation is imminent and act before it occurs. The danger inherent in the no-deal equilibrium—that it may soon lead to costly conflict—creates incentives for the two sides to instead agree to a deal that would avert this danger, as we will see next.

The Viability of Possible Deals

The two sides can avoid proliferation or costly conflict only if they make a deal, in which A offers more generous concessions to B's interests than it would otherwise make, and in exchange B refrains from initiating (or continuing) a nuclear weapons program. If A reneged by making less than the agreed concessions to B, or B was caught pursuing nuclear weapons, the two sides would revert to the no-deal equilibrium, with its attendant risks of proliferation and conflict. If a deal is viable, it enables A to commit to concessions that make nonproliferation worthwhile for B. This in turn enables B to commit not to seek nuclear weapons, thus rationalizing A's concessions. If no deal is viable, then there is no way for B to assure A of his non-investment. A must therefore assume B will seek nuclear weapons and B must do so to have any chance of better offers, potentially leading to costly conflict as described in the previous section. Here, we investigate whether and *when* a viable deal exists.[21]

No Conflict, No Deal

Our first result says that, if going without a deal poses no risk of eventual costly conflict, then no deal will be viable.[22]

Proposition 1. *If a given subgame of the game does not have a no-deal equilibrium featuring a positive probability of costly conflict, then it also does not have a deal equilibrium.*

To understand this result, notice that in any nonproliferation deal, two things must be true. First, the value B gets from a deal must be at least a little higher than what he would get without the deal, because otherwise he has no reason not to cheat on the deal—that is, to invest in a weapons program—in the hopes of doing even better once he has nuclear weapons. Second, the value A gets from a deal must be no less than the value she would

[21] A subgame has a viable deal whenever it has a Perfect Bayesian Equilibrium in which, along the equilibrium path, B never (again) invests in a nuclear weapons program.

[22] Proofs for the propositions appear in the online appendix.

get without a deal, because otherwise she would rather not accept the deal in the first place. Now, the only way it can simultaneously be true that A gets at least the same value, and B gets a higher value, with a deal than without it, is if going without a deal means doing something that destroys part of the game's total value. In our model, the only thing that destroys value is costly conflict. If no equilibrium features attack or abandonment, then there is no surplus value with which A can encourage B not to invest while still leaving herself at least as well off.[23]

Remarkably, this does not mean that it is impossible for A to secure nonproliferation without attacking or abandoning B. In fact, there are deals A could offer that would leave both sides better off than costly conflict (whether such conflict was possible in equilibrium or not) *and* that would ensure B's compliance. The problem is that the generous offers to B involved in such a deal actually leave A worse off than the stingier offers she would make in the absence of a deal, even though this would lead to B's eventual proliferation. Thus, *sometimes a deal, while feasible, is not worth the cost for A.* Any offers generous enough to win B's compliance would demand too much from A, and any offers A would actually be willing to make would only lead to B's cheating.

What this result does mean is that the viability of any possible deal is closely linked to the willingness of A to resort to costly conflict in its absence. If force or abandonment is "taken off the table," then no deal will be viable, and any deal that is agreed upon will soon unravel, either because A refuses to offer the agreed concessions, or because B covertly initiates (or restarts) a program. So, in subsequent sections, we make the following assumption.

Assumption 1. *If A knows that B's program is advanced, then in the absence of a deal, A will immediately initiate costly conflict.*

This assumption implies that, even if A is uncertain of the progress of B's program, but confident enough that it is advanced (i.e., it has mastered fissile material production), then A will attack or abandon B in the absence of a deal. When A is confident enough that B's program is advanced to resort to costly conflict unless a deal is made, we will say that "A believes B's program is nearing success."[24]

[23]Recall that our model ignores the budgetary costs of a program and the efficiency costs of sanctions incurred while it is ongoing. Theoretically, these costs might create enough surplus to support a deal, even in the absence of any possibility of attack or abandonment. Empirically, however, these costs are typically very small relative to those of attack or abandonment. Because the surplus they create is so much smaller, by themselves they would suffice to render a deal viable only under a much narrower set of circumstances than would the costs of attack or abandonment.

[24]This belief is formally defined in the online appendix by the threshold at which A is just confident

Because, as we will document in the empirical section, most proliferants make a deal at some point, we focus in the rest of this section on the timing, threats, and concessions involved in a deal rather than the question of whether a deal ever occurs. However, the appendix contains a full discussion and formal analysis of the conditions for deal viability.

Two Best Times for a Deal

From start to finish, a program to develop nuclear weapons can take decades. The times at which a deal might be made can be divided into three periods, which we refer to as early, middle, and late. These times are defined in terms of A's *estimate* of the progress of B's program, which may or may not correspond to the program's actual progress. An "early" deal is one agreed before B has even begun a program or once it has begun but while A is sure it is rudimentary. A "middle" deal is one made when A has become uncertain of the program's progress, but does not yet believe it is nearing success. A "late" deal is one negotiated once A has come to believe that B's program is nearing success. As we will see next, the prospects for a viable deal are not uniform among these three periods.

Proposition 2. *The conditions for a deal equilibrium are least restrictive either for an early or for a late deal.*[25]

This result says that the best chance for agreeing to a deal occurs either early, when A is certain B's program is rudimentary, or late, when she believes it is nearing success. Finding a viable deal in between—in the middle period that, as we show subsequently, makes up most years of most programs—will be harder.[26]

To understand the intuition for this result, observe that two factors that together govern the viability of a deal will change as time passes. First, B's temptation to cheat on a potential deal will grow as his program makes progress. Starting from the advanced stage of progress, successful acquisition of nuclear weapons is more likely in the next period, and

enough that the program is advanced to resort to conflict. Assumption 1 guarantees that such a threshold exists.

[25]To simplify the analysis, we assume that, if a viable deal is agreed while the stage of B's program is uncertain, then it will not be altered if A later learns the program's stage. In principle, the US might renege on a deal designed to satisfy a B with an advanced program if the US later found out that the program was actually rudimentary. In practice, the US has arguably upheld its late deals even when it later found that the program in question was not as advanced as it had feared, perhaps for reputational reasons.

[26]This is not an artifact of our assumption that a program has only two stages before acquisition. The proof of this proposition given in the online appendix generalizes quite readily to any number of stages.

so he has more to gain from reneging on an agreement. Thus, if A believes there is some positive probability that B's program has reached this stage, the only way for her to assure nonproliferation is to make concessions to B that would be sufficient to induce him to comply with the deal *even if his program was advanced.* It is not enough to offer concessions that would only satisfy a B whose program remained at the rudimentary stage: if his program had made unobserved progress, then he would cheat on the deal. Thus, the temptation to cheat that A's concessions must be sized to overcome is the same as long as there is some probability that B's program is advanced, or equivalently, as long as A is not sure that B's program remains rudimentary.

Second, however, the severity of the punishment that A can credibly threaten to enforce a deal rises as she becomes increasingly worried that B's program is advanced. If a deal is agreed early, A will estimate that B's program likely remains rudimentary, and so will elect to wait and see rather than attack or abandon B immediately if she catches him cheating, a relatively mild punishment for B. The later a deal is agreed, the more suspicious A will be that B's program is nearing success, and the quicker she will resort to conflict if she detects cheating. Because conflict will happen sooner after cheating is detected, the punishment for cheating is more severe. Once A believes that his program is nearing success, she can threaten immediate attack or abandonment, which is the most severe punishment possible for B.

Combining these two observations, the temptation to cheat that A's concessions must overcome is the same as long as there is some probability that B's program is advanced, but the severity of the punishment for B cheating increases the more confident A is of this. Thus, once enough time has passed that A is no longer sure B's program remains rudimentary, a deal is most likely to be viable when A is confident enough that B's program is advanced to justify immediate resort to conflict in the face of detected cheating.

By contrast, when A is sure B's program is rudimentary, his temptation to cheat is known to be low, but the punishment she can credibly threaten is also less severe. Depending on the parameter values, the single best opportunity for a deal may be in this early period, or instead in the late period, when A believes B's program is nearing success. A deal may be viable only in the early period, only in the late period, in both, or in neither, in which case no deal will ever be viable. Regardless, a middle deal will always be harder to agree, because in the middle period A must overcome a potentially high temptation to cheat without the benefit of a credible threat of immediate conflict.

This result carries an important implication: even if no deal was agreed upon or complied with in earlier years, once a proliferant's program is believed to be nearing success, a newly viable deal may arise. Put another way, a proliferant may decline a deal, cheat on an agreed deal, or even refuse to negotiate seriously in the early and middle years of its program, but eventually become genuinely willing to negotiate, expecting that there is now an agreement that would be complied with by both sides. It cannot be assumed that this apparent change in amenability to a deal is a deception, intended to lull the US into protracted negotiations while the proliferant moves toward completing his weapons program.

For instance, one might presume that Iran's agreement to temporary restrictions on its program is simply a smokescreen. After all, the US and others have been negotiating with Iran for a long time, and no deal was forthcoming, and Iran has already cheated on an earlier deal implicit in the NPT. The current deal only came after increasingly open discussion of attacking Iran in the US and Israel, so Iran may only be trying to buy time. But the model provides another interpretation: it is precisely the increasing willingness of the US and Israel to attack, motivated by their fear that Iran's program is nearing completion, that makes a deal potentially viable. Thus, Iran's change of negotiating posture may in fact be sincere.

Proposition 2 thus carries two observable implications:

Hypothesis 1. *Deals should occur either very early in a program, when it is known to be rudimentary, or once it is believed to be nearing success—not in between.*

Hypothesis 2. *A late deal should coincide with a newly-credible threat to attack or abandon the proliferant in its absence.*

We turn next to considering the relationship between the viability of a late deal and that of an early deal.

Anticipating a Better Deal

It turns out that a viable late deal must be more generous than any viable early deal. This later generosity can pose a serious obstacle to the viability of any earlier deal, and can even encourage a proliferant to pursue a program in the first place, increasing the probability of proliferation.

Proposition 3. *Suppose that both an early and a late deal are in equilibrium. Then no early deal is renegotiation-proof, and if ϵ is low enough, any viable late deal must feature larger concessions to B than any viable early deal.*

Imagine the following scenario. When an enemy proliferant's program is rudimentary or has not yet begun (equivalent to low ϵ), it agrees with the US to an early deal, enforced by the threat of reverting to the no-deal equilibrium, that is viable under this threat. However, after the deal is agreed, the US catches the proliferant with a covert weapons program. The two sides revert to the no-deal equilibrium, and at some later point, the US estimates that proliferation is near enough to merit the use of force to prevent it. The equilibrium US strategy calls for an immediate attack.

If a viable late deal exists, the US could offer it instead of attacking. The surplus from avoiding war means there is a deal that both sides would strictly prefer to sticking with the no-deal equilibrium. Why would the US go to war, given that there is a way to avoid both war and proliferation that is better for both sides? Faced with this situation, the US would be sorely tempted to make a new deal rather than to attack. From the proliferant's perspective, this new deal would be even better than the early one, as the US would have to make more generous concessions in order to overcome the proliferant's higher temptation to cheat if its program is advanced.

Hypothesis 3. *A late deal should be more generous than an early deal.*

Because a late deal will be more generous, the proliferant will be severely tempted to cheat on an early deal if it anticipates that the US would later deal rather than attack when the latter was called for. Pursuing nuclear weapons would no longer be a risky bet that the proliferant would get them before being discovered and attacked. Instead, it would be a sure thing: eventually, either the proliferant would get nuclear weapons, or it would be offered a better deal. Either way, it would end up better off than under the early deal. This anticipation creates another motive for a proliferant to invest in a program in the first place, in addition to the prospect of eventually getting the weapons.[27] Some actual proliferants—such as Iran or North Korea—might have pursued nuclear weapons partly out of the hope that it would eventually bring forth a more generous "bribe" from outside powers in exchange for stopping an advanced program.

[27] If we assume that in equilibrium a late deal rather than costly conflict will occur, then by Proposition 1 no early deal will be viable and the equilibrium will always feature B investing and either proliferation (if B's program succeeds before A comes to believe it is nearing success) or a late deal (otherwise). Thus the viability of an early deal turns on whether cheating is expected to lead to conflict or instead to a late deal.

Empirical Tests

We proceed to test the hypotheses derived from our model. Our universe of cases is the set of all states for which there is ever the potential for a meaningful nonproliferation deal. Theoretically, this means that each state must possess, or have been perceived by a potential attacker or abandoner as possessing, an interest in nuclear weapons and the capability to pursue them. Operationally, we include in this set every state that has ever pursued nuclear weapons or explored doing so, as coded by Singh and Way (2004) (SW), or possessed a nuclear weapons program, as coded by Jo and Gartzke (2007) (JG).[28] We also include an additional set of states which, while coded as never exploring, pursuing, or possessing a nuclear weapons program by SW and JG, were at some point judged capable of and potentially interested in developing nuclear weapons, as coded by Coe and Vaynman (2015) (CV). Finally, we include the three successor states to the USSR that were left in possession of Soviet strategic nuclear weapons and associated infrastructure. These states did not have operational control over the Soviet weapons or the ability to quickly produce more, but they had the option of pursuing both and so were potential proliferants. This definition excludes those states that, even having chosen to do so, could not plausibly have acquired nuclear weapons, such as Afghanistan or Cameroon, as well as those states that never showed any serious interest in doing so, such as Mexico or Thailand.[29]

From this set, we then drop the United States, WWII-era Germany and Japan, the Soviet Union, the United Kingdom, France, China, and India. The first three states pursued nuclear weapons while engaged in total war, so that there was little chance of a nonproliferation deal. The dominant view in the literature holds that the Soviet program could never be credibly threatened with attack (by the US or any other state).[30] The UK and France were safe from Soviet preventive attack due to the anticipation that this would escalate to a large war with the US with prohibitive costs for the USSR, while China was similarly safe from US attack due to also-large costs and its alliance with the USSR. These three states were also safe from abandonment by their allied superpower over their nuclear programs, because, in the early

[28]Following Montgomery and Sagan (2009), we use both datasets to ensure that all of our tests are robust to the coding disagreements between them. We updated the SW dataset to the latest version available at Way's website, dated June 12, 2012.

[29]Instead including such states in our tests would strengthen the evidence for H1.

[30]See Bas and Coe (2012) for a review of this literature, but also for the possibility that the US actually might have attacked had it not underestimated the progress of the Soviet program. Treating the Soviet case in this way, the Baruch Plan can be conceived of as a US attempt at a deal with the USSR when its program was still thought to be in an early stage.

years of the nuclear era, both superpowers saw advantages to the spread of nuclear weapons to their allies (Coe and Vaynman, 2015). Subsequent changes in how the superpowers conceived of proliferation led them to threaten later allies suspected of pursuing nuclear weapons with suspensions of aid and threats of abandonment, which would have exposed those allies to attack by the opposing side (Coe and Vaynman, 2015). India was the only exception to this rule (Bas and Coe, 2016, 25). Because these states could not be credibly threatened with attack or abandonment, Proposition 1 implies there was never any potential for a deal, and indeed these states never agreed to any deal.

The remaining states are listed in Table 1. The first group of states never explored, pursued, or possessed a nuclear weapons program but were judged able to do so (from CV); the second group explored but did not pursue (from SW); the third and fourth groups pursued a program and agreed to one or more deals (respectively). The fifth group inherited nuclear capabilities, and the last group got nuclear weapons. The period of each state's exploration (e) or pursuit (p) of a program is from SW.

Table 2 lists every nuclear nonproliferation deal made by these states, up to 2007.[31] Though our model is set in a bilateral context, we treat the participation of any state in our case universe in the multilateral Treaty on the Non-proliferation of Nuclear Weapons (NPT) as a deal to which the model applies. The NPT can be viewed as an efficient substitute for a large number of bilateral nonproliferation deals with each of these states (Verdier, 2008), in which each agreed not to seek nuclear weapons in exchange for concessions such as access to peaceful nuclear technology and assistance with it, principally enforced by the US or USSR (Coe and Vaynman, 2015). We treat subsequent multilateral deals, such as the US/UK deal with Libya in 2003, similarly.[32]

This data reveals some surprising patterns. Of the 30 states that were judged capable of pursuing a nuclear weapons program, 27 (90%) agreed to a nonproliferation deal at some point. Of the 29 states that had the opportunity to agree to an early deal (operationalized below), 23 (79%) did so. Of the ten states whose programs were ever believed to be nearing success and thus had the opportunity to make a late deal, six (60%) did so. In aggregate, a deal is made in 74% of the instances that our theory identifies as opportune times for a

[31] This cutoff excludes only the 2015 agreement with Iran. The recency of this case makes it difficult to be confident in the information needed to test our hypotheses, but we will discuss it in the conclusion.

[32] Deals that were intended only to prevent overt nuclear deployment (made with Israel, Pakistan, and South Africa) do not meet our definition (Rabinowitz, 2014; Rabinowitz and Miller, 2015). We also exclude unilateral abandonment of nuclear weapons (by South Africa) and deals to constrain vertical proliferation (made by the superpowers).

State and Period	H1: deal timing early	btwn	late	H2: late deal ⇒ threat	H3: late > early deal
Belgium	✓			—	—
Canada	✓			—	—
Czechoslovakia	✓			—	—
Italy	✓			—	—
Japan	✓			—	—
Poland	✓			—	—
W. Germany	✓			—	—
Algeria, 1983–(e)	✓			—	—
Indonesia, 1965–67(e)	✓			—	—
Romania, 1985–90(e)	✓			—	—
Sweden, 1946–69(e)	mixed			—	—
Switzerland, 1946–70(e)	✓			—	—
Yugoslavia, 1954–65(e), 74–88(e)	✓			—	—
Argentina, 1968–77(e), 78–90(p)		X		—	—
Australia, 1956–60(e), 61–73(p)	mixed			—	—
Brazil, 1953–77(e), 78–90(p)		X		—	—
Egypt, 1960–64(e), 65–74(p)	✓			—	—
Iraq, 1976–82(e), 83–95(p)	✓			—	—
Syria, 2000–(p)	✓			—	—
Iran, 1976–84(e), 85–(p)	✓		✓	✓	✓
Libya, 1970–2003(p)	✓		✓	✓	✓
N. Korea, 1965–79(e), 80–05(p)	✓		✓	✓	✓
S. Korea, 1959–69(e), 70–78(p)	✓		✓	✓	✓
Taiwan, 1967–77(p), 87–88(e)	✓		✓	✓	X
Belarus	✓			—	—
Kazakhstan	✓			—	—
Ukraine	—	—	✓	✓	—
Israel, 1949–57(e), 58–68(p)				—	—
Pakistan, 1972–86(p)				—	—
S. Africa, 1969–73(e), 74–78(p)				—	—

Table 1: States Judged Capable of Pursuing Nuclear Weapons

State	Deal	Year	State	Deal	Year
Algeria	NPT	1995	Libya	NPT	1968
Argentina	Guadalajara	1991	Libya	Deal w/ US/UK	2003
Australia	NPT	1970	N. Korea	NPT	1985
Belgium	NPT	1968	N. Korea	Agreed Framework	1994
Belarus	Lisbon	1994	Poland	NPT	1968
Brazil	Guadalajara	1991	Romania	NPT	1968
Canada	NPT	1968	S. Korea	NPT	1968
Czechoslovakia	NPT	1968	S. Korea	Deal w/ US	1976–81
Egypt	NPT	1968	Sweden	NPT	1968
Indonesia	NPT	1970	Switzerland	NPT	1969
Iran	NPT	1968	Syria	NPT	1968
Iran	EU-3 Deal	2003	Taiwan	NPT	1968
Iraq	NPT	1968	Taiwan	Deal w/ US	1977–78
Italy	NPT	1969	Ukraine	Lisbon	1994
Japan	NPT	1970	W. Germany	NPT	1969
Kazakhstan	Lisbon	1994	Yugoslavia	NPT	1968

Table 2: Nonproliferation Deals

deal.[33] Because deals are usually made, we focus our testing on our theory's predictions about the timing of, concessions made in, and implicit threat behind these deals, as opposed to whether a deal occurs.[34]

H1: The Timing of Deals

Hypothesis 1 predicts that nonproliferation deals should occur early in a program, when it is known to be rudimentary, or late, once it is believed to be nearing success—not in between. To test this, we need to specify exactly what is meant by these two times. Any deal that is agreed in between these two specified times is then evidence against the theory. Because the choice of these specifications is somewhat arbitrary, the danger is that we end up labelling very few of the state-years in the data as in between, making it hard to falsify the prediction. To avoid this, we adopt what we view as stringent specifications, which result in most years of states' nuclear weapons programs being treated as in between.

We use two alternative specifications of "early in a program, when it is known to be rudimentary." In the first, we include in this period those state-years coded by SW as not pursuing a program or possessing nuclear weapons. In these state-years, a state's leaders cannot have made a political decision to acquire nuclear weapons, and the state cannot have developed any nuclear technology useful only for military application (Singh and Way, 2004, 866). In the second, we include those state-years coded by JG as not possessing a program or nuclear weapons. In these state-years, the highest decision-maker cannot have authorized a nuclear weapons program, and the state's nuclear activities cannot have been "seen to increase noticeably."[35] To each set, we add those state-years coded as respectively pursuing (in SW) or possessing (in JG) a nuclear weapons program in which the state in question is not yet operating and has not yet begun constructing a facility capable of producing fissile material, using the comprehensive dataset on such facilities reported in Fuhrmann and Tkach (2015). In sum, a program is coded as early if it either has not yet started, or has started but has not yet even begun to build the facilities necessary to actually produce fissile material. This criterion results in 23% (46 of 204) of program-years in which a state is coded

[33]Approximately one-third of early deals and one-third of late deals are cheated upon. Cheating might arise from an exogenous change in the parameters governing a deal's viability. Introducing this possibility would not qualitatively alter our results as long as the probability of cheating is not too high.

[34]Coe and Vaynman (2017) provides an explanation for why late deals are less likely to be made than early ones: the former require more intrusive inspections to assure the detection of cheating, and so are more dangerous for the inspected side.

[35]This description is taken from the data notes to Jo and Gartzke (2007).

as pursuing a program (by SW) being labeled as early.

Next, we specify "late, once it is believed to be nearing success" to mean that, in such a state-year, the state's program is contemporarily estimated by the most likely attacker or abandoner to be within four years of acquiring nuclear weapons.[36] To determine which state-years meet this criterion, we draw on the compilation of US government intelligence estimates assembled by Montgomery and Mount (2014), augmented by the additional estimates collected in Bas and Coe (2016).[37] This criterion results in 15% (31 of 204) of program-years being labeled as late.

The second column of Table 1 contains the results. The sub-columns correspond to the three possible times at which a deal could occur: early, when a program is known to be rudimentary; late, when a program is believed to be nearing success; and any time between these two. Each deal in the data appears in a subcolumn according to our coding of the time in which it occurred. A ✓ indicates the timing of the deal supports H1, an X indicates it does not, and "mixed" is used when it supports H1 under one specification but not under the other.

Of the 32 deals in the data, 23 occur when a program is known to be rudimentary and 6 occur when it is believed to be nearing success, while only 3 occur in between these two times. Thus, 91% of deals are timed as H1 expects, with the association between program-years with a deal and our two times statistically highly significant ($p = .0000$ in a one-tailed exact test). The only cases that do not clearly support H1 are Argentina, Brazil, and either Australia (under SW) or Sweden (under JG).[38] Argentina and Brazil made a deal with each other to abandon their nuclear weapons programs in 1991, when both

[36]Instead setting the threshold at five years does not alter the results; a threshold of more than five years seems implausibly long to claim that success is imminent. Setting the threshold to three (or fewer) years significantly weakens the support for this hypothesis, but seems inappropriate given that intelligence estimates of when a program will succeed are often specified as ranges of two to three years.

[37]These cover all cases except Belarus, Kazakhstan, and Ukraine. We found no evidence that the US or Russia estimated that Belarus or Kazakhstan might be able to establish operational control over their inherited nuclear weapons, but they estimated that Ukraine could do so within less than two years (Budjeryn, 2016, 156). Moreover, the nuclear infrastructure and technological expertise present in Belarus and Kazakhstan were substantially inferior to that of Ukraine (Budjeryn, 2016, 54–58). We thus code Belarus and Kazakhstan as never estimated to be nearing success, but Ukraine as being so for 1991–1994.

[38]Fuhrmann and Tkach (2015) codes North Korea as operating a commercial-scale reprocessing facility beginning in 1983, but the source cited for this fact, Albright and Brannan (2007), contains no information about when this facility was built. Oberdorfer (2001, 250) claims the US photographed the facility in an early stage of construction in March 1986, suggesting it was begun that year, so we code the North Korean program as rudimentary in 1985, when it signed the NPT.

programs were neither rudimentary (both operated pilot-scale enrichment facilities during the 80s) nor nearing success. Their deal instead seemed to arise from a change in preferences toward their nuclear programs caused by regime change in both states and a long process of rapprochement in their bilateral relations (Coutto, 2014).[39] Finally, Australia and Sweden had only a single laboratory-scale enrichment or reprocessing (respectively) facility under construction or operating when each signed the NPT. Thus, this hypothesis is strongly supported by the historical record.

H2: Late Deals and Threat Credibility

Hypothesis 2 predicts that, when late nonproliferation deals occur, they should coincide with a newly-credible threat to attack or abandon the proliferant in the absence of a deal. To test this, we must determine whether, at the time the late deal was made, the relevant attacker/abandoner had just become willing to initiate costly conflict over the proliferant's program in the deal's absence. This is difficult to measure directly, so we resort to the proxy of whether that state began issuing threats of attack/abandonment, in public or in private but high-level official communication, just before the deal was made.

In the six late deals—with Iran, Libya, North Korea, South Korea, Taiwan, and Ukraine—the US is the primary interlocutor, with the relevant threat being attack in the first three cases and abandonment in the last three. We find that H2 is supported in all six cases, as shown in the second-from-right column of Table 1. Of course, newly-credible threats might also have occurred in program-years where a late deal was not made. However, even if as much as 48% of all program-years in our data featured such threats, the association we found between late deals and these threats would still be significant at the conventional .05 level. Given that such a high frequency of threats seems implausible, we conclude that the support for our hypothesis is statistically significant.

For brevity, we discuss here just Iran and South Korea, with the details of the other cases relegated to the online appendix. In October 2003, Iran agreed to a deal with the EU-3 (France, Germany, and the UK) that was understood by both sides as a means to avert a US and/or Isreali attack on Iran (Volpe, 2015, 215–217). While US-Iran relations had been hostile since the 1979 revolution, through the 1990s the US response to Iran's nuclear program was largely limited to imposing sanctions and pressuring other countries

[39] One could also argue that neither would ever have been subject to a credible threat of attack or abandonment, in which case both should be excluded as, by Proposition 1, there is no potential for a nonproliferation deal driven by the possibility of costly conflict.

not to assist the program (Volpe, 2015, 202–207). Tensions between the US and Iran began to rise in early 2002 and reached a fever pitch after the US successfully invaded Iraq and overthrew its regime, with Iran fearing that it would be next (Corera (2006, ch. 7), Volpe (2015, 212–214), Kerr (2016, 41)). In mid-October 2003, Israel made explicit threats to strike Iran's nuclear facilities with US help, while the United Kingdom publicly refused to rule out attack (presumably in coalition with the US) (Nuclear Threat Initiative, 2011, 185). Consistent with H2, the deal was made immediately thereafter.

Deal-making with South Korea over its nuclear weapons program began soon after the program was discovered, in 1975, and continued until 1981. South Korea implemented major constraints on its program at three points: January 1976, when it canceled its purchase of a plutonium reprocessing facility from France; December 1976, when it ended its efforts to indigenously develop a plutonium-producing reactor and a heavy-water production facility; and early 1981, when it appears to have finally abandoned the option to eventually produce nuclear weapons (Reardon (2010, 226–236), Siler (1998, 75–78)). Immediately preceding each of these three points, and at no earlier point, the US explicitly threatened to end its alliance with South Korea if the program continued unabated (Reardon (2010, 231), Oberdorfer (2001, 71), Siler (1998, 75–76)).

H3: Late vs. Early Deals

Hypothesis 3 predicts that deals agreed when the proliferant's program is believed to be nearing success should feature more generous concessions than those agreed when it is known to be rudimentary. Because many other factors also affect the concessions a proliferant will receive—such as the cost of attacking or abandoning that state and its technological sophistication—we limit our test to those states that made both types of deals. By comparing the concessions each state received under its early deal to those received under its late deal, we can reduce the variation in these other factors.[40] We find that H3 is supported in four of the five cases, as shown in the rightmost column of Table 1. For brevity, we discuss two supportive cases as examples, and also the one case that does not support the theory.

In their early deals under the NPT, Iran, Libya, South Korea, and Taiwan were offered no concessions other than those written into the treaty and open to any member.[41] However, in

[40] We believe this is the most appropriate way to test this hypothesis, but the results are qualitatively unaltered if we instead compare the average concessions across all early deals to the average for late deals.

[41] More precisely, we found no indication in the literature that any of these states received any special concessions in exchange for joining the NPT.

exchange for North Korea joining the NPT, the USSR did promise to build it nuclear power plants, though without a construction timetable or payment plan (Zhebin, 2000, 32–33).

Iran, Libya, South Korea, and North Korea received several additional major concessions under their late deals. For example, Libya's deal allowed US companies (on which Libya's oil industry depended) to invest in Libya, normalized diplomatic relations, and included public assurances of security from the US and UK, with the latter even agreeing to help Libya strengthen its military (Jakobsen, 2012, 502–503).[42] South Korea received access to US reprocessing and increased technology transfers, strong public statements of support for the alliance and threats against North Korea, recognition of a coup, an upgrade of its alliance to the level of NATO, an end to US withdrawal of forces, and increased military aid (Reardon (2010, 231–232, 235–236), Siler (1998, 77)).

Taiwan's late deal appears inconsistent with our hypothesis. We could find no mention in the literature of any concessions Taiwan received from the US during the relevant time period of 1975–79. Although the 1979 Taiwan Relations Act (TRA) required the US to provide arms to Taiwan sufficient to defend itself and declared that threats to Taiwan would be a "grave concern," this act replaced the 1954 Mutual Defense Treaty, which some have argued provided a stronger, less ambiguous security guarantee than the TRA (Hersman and Peters (2006, 543, 545), Bush (2009)). Thus, to conclude that the TRA constituted a major concession to Taiwan, one would have to argue that in the counterfactual absence of Taiwan's nuclear program, the US would have lessened its commitment to Taiwan even more.[43] Considering the well-known tilt of US policy in favor of China over Taiwan in the 1970s, this argument cannot be ruled out. However, in the absence of clear evidence for this argument, we count this case as inconsistent with our hypothesis.

[42] These were large concessions by the US. The sanctions had debilitating effects on Libya's economy and regime stability, and might have induced Libyans to overthrow Gaddafi, a reviled dictator directly responsible for the deaths of hundreds of American and allied civilians in terrorist attacks (Bowen, 2006, pp. 54–55). Thus, the deal required the US to forgive a hated enemy and to cease undermining his regime. Gaddafi's explicit demand that the US "drop its goal of regime change" as part of the deal was strongly resisted within the Bush administration (Hirsh, 2005, p. 30).

[43] Segal (1998) and Solingen (2007, 112) seem to imply this in asserting that Taiwan used its program to force the US to affirm its commitments.

Broader Implications

We conclude with a discussion of the implications of our results for the study of arming generally and nuclear proliferation specifically, and for policy-making on nuclear issues.

For bargaining theorists, arming is mysterious for exactly the reason war is: why do rational actors employ costly measures (whether war or arming) to wage their disputes? We cannot know the answer if we do not understand why arms control—such as the nonproliferation deals we analyzed here—sometimes fails to be viable. Previous formal analyses of arming do not address this puzzle, because they generally do not consider arms control at all. Ignoring the possibility of deals such as those analyzed here may undermine the validity of these studies' results. For example, we find in the appendix that lower costs of war make war more likely in the no-deal (arming) equilibrium. But when the possibility of arms control is taken into account, lowering the costs of war may render a deal viable and thus *prevent* war. Determining which of this literature's conclusions are robust to the consideration of arms control is an important subject for future research.

For scholars of nuclear proliferation, our theory suggests a new understanding of nuclear latency: the phenomenon of states that possess the technological prerequisites for nuclear weapons, but have chosen not to build them. Fuhrmann and Tkach (2015) shows that more than thirty countries have reached nuclear latency, but only ten ever built nuclear weapons. Mehta and Whitlark (Forthcoming) characterize existing work on latency as advancing two competing views. One takes latency as a provocation that undermines a state's security and prosperity by increasing the risk of sanctions and military conflict. The other argues that latency deters aggression and provides leverage against allies and adversaries. Our theory synthesizes these views: moving toward latency should lead to sanctions and raise the risk of attack. However, once latency is achieved, the leverage it affords can lead to a nonproliferation deal that is generous enough to outweigh the costs incurred along the way. Thus, these states rationally move toward the capability to build nuclear weapons because this might force the US to offer greater concessions in exchange for their agreement not to go further.

Future research might profitably examine why some deals subsequently fail. Of the thirty states judged capable of pursuing nuclear weapons (from Table 1), ten are known to have at least explored a nuclear weapons program after making an early deal, and two also cheated on later deals. One possible explanation for cheating is an unexpected change in the chances of a program's success, such as might have occurred when Iran and North Korea gained access to the Khan proliferation network. Another is that the cost to the US of attack or of the concessions made under a deal might increase over time, leading the US to end a deal, as some argued occurred with the North Korea deal when the Bush administration came to power. If this possibility is anticipated, US assurances of concessions under a deal may not be credible, undermining any deal's viability.

Our theory also surfaces a more subtle explanation for deal failure: the viability of early deals is closely linked to expectations about whether cheating will lead to the risk of costly conflict or instead to a more generous deal. Late deals, such as those agreed with North Korea and more recently Iran, might undermine the nonproliferation regime by shifting states' expectations about how cheating will be dealt with. Have some states been encouraged by other states' late deals to cheat? Does this influence US responses to cheating? Examining how these expectations are formed and influenced is an important task for future research.

For policy-makers, our theory highlights a dilemma that arises when the US must respond to a proliferant's cheating on a nonproliferation deal. On the one hand, the death and destruction caused by preventive attack argue for giving a cheater every chance to agree to a new deal. On the other hand, allowing a proliferant to extract additional concessions by virtue of progress in its illicit weapons program can undermine deals with other potential proliferants. Going to war in response to the latter concern might or might not be the right thing for the US to do. Even if it is the right thing, the US government is still placed in the unenviable position of explaining to domestic and international audiences why it undertook a violent, costly resolution of the crisis, when one that would avoid violence was available.

Finally, our theory helps to illuminate the most recent deal-making with Iran. In 2005, as the US military became increasingly enmired in Iraq and new intelligence estimates put Iran's program further from success than previously thought, the credibility of the US threat to attack Iran in the absence of a deal was severely eroded.[44] Consistent with our theory, the 2003 deal with the EU-3 unraveled late in 2005.[45] When in 2008 Iran's program was once again estimated to be nearing success, US and Israeli threats were made and, in the absence of a new deal, led to what is suspected to have been a campaign in 2009–2011 to sabotage Iran's program through cyberwarfare and assassinations of key scientists.[46] Iran subsequently agreed to serious constraints on its nuclear program in November 2013 and

[44]National Intelligence Council (2007) discusses the timeline of the 2005 estimates.

[45]As we discussed previously, it was understood by all sides that this deal would restrain the US from attacking.

[46]For the revised estimate, see National Intelligence Council (2007). For US and Israeli threats, see the entries on June 11, 2008 and January 11, May 17 and 19, and July 5, 7, and 16, 2009 in Nuclear Threat Initiative (2011). On the sabotage campaign, see Broad, Markoff and Sanger (2011).

July 2015, which were implemented starting in early 2014 and have continued to this writing (Kerr, 2016). In exchange, UN and some EU and US sanctions on Iran will be dropped, the conventional weapons embargo will be ended, and Iran will retain the most commercially valuable aspects of its program, especially its ability to produce low-enriched uranium for power plants (Gordon and Sanger, 2015). These are major concessions by the US and its partners that will enable Iran to build its regional power, and were protested as such by Israel and Arab states (Gordon and Sanger, 2015). Our theory makes clear why large concessions are necessary to support such a deal, agreed when Iran was estimated to be able to build a bomb in a very short time. But it also suggests that such a deal, made after so many years of deception and recrimination, may nonetheless be sincere, and lasting.

Supplemental Information

Online appendix available at `http://www.andrewjcoe.com` and at the International Studies Quarterly data archive.

References

Albright, David and Paul Brannan. 2007. "The North Korean Plutonium Stock, February 2007." http://www.isis-online.org/publications/dprk/DPRKplutoniumFEB.pdf.

Baliga, Sandeep and Tomas Sjöström. 2008. "Strategic Ambiguity and Arms Proliferation." *Journal of Political Economy* 116(6):1023–1057.

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.

Benson, Brett and Quan Wen. 2011. A Bargaining Model of Nuclear Weapons Development and Disarmament. In *Causes and Consequences of Nuclear Proliferation*, ed. Robert Rauchhaus, Matthew Kroenig and Erik Gartzke. Taylor and Francis pp. 45–62.

Bowen, Wyn Q. 2006. "Libya and Nuclear Proliferation: Stepping Back from the Brink." *The Adelphi Papers* 46(380):7–103.

Brands, Hal and David Palkki. 2011. "Saddam, Israel, and the Bomb: Nuclear Alarmism Justified?" *International Security* 36(1):133–166.

Braut-Hegghammer, Målfrid. 2008. "Libya's Nuclear Turnaround: Perspectives from Tripoli." *The Middle East Journal* 62(1):55–72.

Broad, William J., John Markoff and David E. Sanger. 2011. "Israeli Test on Worm Called Crucial in Iran Nuclear Delay." *The New York Times* .

Budjeryn, Mariana. 2016. The Power of the NPT: International Norms and Nuclear Disarmament of Belarus, Kazakhstan, and Ukraine, 1990–1994 PhD thesis Central European University.

Bush, Richard C. 2009. "Thoughts on the Taiwan Relations Act." *China Times* .

Coe, Andrew J. and Jane Vaynman. 2015. "Collusion and the Nuclear Nonproliferation Regime." *Journal of Politics* 77(4):983–997.

Coe, Andrew J. and Jane Vaynman. 2017. "The Tragedy of Arming." Working Paper.

Corera, Gordon. 2006. *Shopping for Bombs: Nuclear Proliferation, Global Insecurity, and the Rise and Fall of the AQ Khan Network*. Oxford University Press.

Coutto, Tatiana. 2014. “An International History of the Brazilian-Argentine Rapprochement.” *International History Review* 36(2):302–323.

Debs, Alexandre and Nuno P. Monteiro. 2014. “Known Unknowns: Power Shifts, Uncertainty, and War.” *International Organization* 68(1):1–31.

Fearon, James D. 2011. “Arming and Arms Races.” Paper presented at the 2010 Annual Meetings of the American Political Science Association, Washington, DC.

Feaver, Peter D. and Emerson M. S. Niou. 1996. “Managing Nuclear Proliferation: Condemn, Strike, or Assist?” *International Studies Quarterly* 40(2):209–233.

Fuhrmann, Matthew and Benjamin Tkach. 2015. “Almost Nuclear: Introducing the Nuclear Latency Dataset.” *Conflict Management and Peace Science* 32(4):443–461.

Gavin, Francis J. 2012. “Politics, History and the Ivory Tower-Policy Gap in the Nuclear Proliferation Debate.” *Journal of Strategic Studies* 35(4):573–600.

Gordon, Michael R. and David E. Sanger. 2015. “Deal Reached on Iran Nuclear Program; Limits on Fuel Would Lessen With Time.” *The New York Times* .

Hersman, Rebecca K. C. and Robert Peters. 2006. “Nuclear U-Turns: Learning from South Korean and Taiwanese Rollback.” *The Nonproliferation Review* 13(3):539–553.

Hirsh, Michael. 2005. “Bolton’s British Problem.” *Newsweek* 145(18):30.

Jackson, Matthew O. and Massimo Morelli. 2009. “Strategic Militarization, Deterrence and Wars.” *Quarterly Journal of Political Science* 4(4):279–313.

Jakobsen, Peter Viggo. 2012. “Reinterpreting Libya’s WMD Turnaround: Bridging the Carrot-Coercion Divide.” *Journal of Strategic Studies* 35(4):489–512.

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

Kerr, Paul K. 2016. “Iran’s Nuclear Program: Status.” Congressional Research Service.

Mehta, Rupal N. and Rachel E. Whitlark. Forthcoming. "The Benefits and Burdens of Nuclear Latency." *International Studies Quarterly* .

Meirowitz, Adam and Anne E. Sartori. 2008. "Strategic Uncertainty As a Cause of War." *Quarterly Journal of Political Science* 3(4):327–352.

Miller, Nicholas L. 2014. "The Secret Success of Nonproliferation Sanctions." *International Organization* 68(4):913–944.

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

Montgomery, Alexander H. and Scott D. Sagan. 2009. "The Perils of Predicting Proliferation." *Journal of Conflict Resolution* 53(2):302–328.

Narang, Vipin. 2014. *Nuclear Strategy in the Modern Era: Regional Powers and International Conflict*. Princeton University Press.

National Intelligence Council. 2007. "Iran: Nuclear Intentions and Capabilities." http://graphics8.nytimes.com/packages/pdf/international/20071203_release.pdf.

Nuclear Threat Initiative. 2011. "Iran Nuclear Chronology." http://www.nti.org/media/pdfs/iran_nuclear.pdf?_=1316542527.

Oberdorfer, Dan. 2001. *The Two Koreas: A Contemporary History*. Basic Books.

Powell, Robert. 1993. "Guns, Butter, and Anarchy." *American Political Science Review* 87(1):115–132.

Powell, Robert. 2015. "Nuclear Brinkmanship, Limited War, and Military Power." *International Organization* 69(3):589–626.

Rabinowitz, Or. 2014. *Bargaining on Nuclear Tests: Washington and Its Cold War Deals*. Oxford University Press.

Rabinowitz, Or and Nicholas L. Miller. 2015. "The Last Line of Defense: US Nonproliferation Policy toward Israel, South Africa, and Pakistan." *International Security* 40(1):47–86.

Reardon, Robert J. 2010. Nuclear Bargaining: Using Carrots and Sticks in Nuclear Counter-Proliferation PhD thesis Massachusetts Institute of Technology.

Segal, Gerald. 1998. "Taiwan's Nuclear Card." *Wall Street Journal* .

Siler, Michael J. 1998. "US Nuclear Nonproliferation Policy in the Northeast Asian Region during the Cold War: The South Korean Case." *East Asian Studies* 16(3-4):41–86.

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

Solingen, Etel. 2007. *Nuclear Logics: Contrasting Paths in East Asia and the Middle East.* Princeton University Press.

Spaniel, William. 2015. "Arms Negotiation, War Exhaustion, and the Credibility of Preventive War." *International Interactions* 41(5):832–856.

Verdier, Daniel. 2008. "Multilateralism, Bilateralism, and Exclusion in the Nuclear Proliferation Regime." *International Organization* 62(3):439–476.

Volpe, Tristan A. 2015. Proliferation Persuasion: Coercive Bargaining with Nuclear Technology PhD thesis The George Washington University.

Zhebin, Alexander. 2000. A Political History of Soviet-North Korean Nuclear Cooperation. In *The North Korean Nuclear Program: Security, Strategy, and New Perspectives from Russia*, ed. James Clay Moltz and Alexandre Y. Mansourov. Routledge.