International Studies Quarterly (2018) **62**, 606–617

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

Muhammet A. Bas
New York University in Abu Dhabi

and

Andrew J. Coe
University of Southern California

We develop a theory of when a deal can be made to stop a state pursuing nuclear weapons, thereby avoiding proliferation or conflict to prevent it. We show a deal can only be made if costly conflict would occur in its absence. Deals are most likely to be made early, when the state's nuclear program is rudimentary, or late, when it is believed to be nearing success, but not in between. A late deal is credibly enforced by more severe punishment than an early one—immediate conflict rather than merely sanctions—and yet must be more generous to the state. If the state anticipates that a late deal would be offered, it will refuse an early deal in favor of continuing its program to secure the more generous late deal. We test and find support for these predictions against the historical record of deal-making over states' nuclear programs.

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 these negotiations succeed, sometimes they fail. Some deals endure, while others prove short-lived. Sometimes a potential proliferant refuses negotiations for years, only to make a deal just as the risk of conflict seems highest. What explains these very different outcomes?

Policymakers considering such negotiations face difficult decisions. Diplomatic effort costs time, resources, and political capital. So when is the best time to gather international support and try to make a deal? If a potential proliferant suddenly becomes willing to negotiate, is it sincere, or just trying to buy time for its nuclear program to succeed? Is there a point beyond which the US should eschew negotiations in favor of unilateral solutions?

We try to answer these questions by analyzing and testing a formal model. In it, two states bargain over disputed issues. One (the "proliferant") potentially invests in, and makes progress toward, acquiring nuclear weapons that, once deployed, would increase its bargaining power. 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. This will eventually lead to either costly conflict or proliferation. Alternatively, the US might make some concessions in exchange for the proliferant ending its pursuit of nuclear weapons. Whether the two sides can reach such a deal depends on 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* prove willing to make more concessions in pursuit of a deal to avoid a costly conflict. However, if the US would not, in fact, attack or abandon the proliferant, then it also would not be willing to make sufficient concessions to render a deal viable.

The model implies that there are essentially two crucial points in time at which the sides are most likely to reach a successful agreement: *early*, when the US initially suspects the existence of a nuclear program and that program is at an early stage of development; and *late*, when the US believes the program is close enough to fruition that the US would rather attack or abandon the state than tolerate its program's continued progress.

In the early stages of proliferation, the US cannot credibly threaten to attack or abandon the proliferant. This makes the penalty for reneging on a deal modest and, in consequence, compliance harder to enforce. However, the proliferant faces little temptation to renege, since its rudimentary program would take a long time to succeed. This renders it relatively easy to "buy off" the proliferant. In the late stages of proliferation, the US threat to attack or abandon the proliferant becomes credible. This makes the penalty for reneging on a deal more severe. But because its program is nearly complete, the proliferant is also more tempted to renege and rush to complete its program. At any point *between* these two extremes, uncertainty about the extent of the program's progress renders the US threat to attack or abandon not credible. Moreover, the temptation for the proliferant to cheat might be high due to the unseen progress of its effort. Thus, a deal is less likely to prove viable.

The possibility of a late deal, agreed when the US believes the proliferant's program is close to success, carries

Muhammet A. Bas is Associate Professor of Political Science at New York University in Abu Dhabi.

Andrew J. Coe is Assistant Professor of International Relations at the University of Southern California.

Authors' note: We are grateful to Robert Powell, Dustin Tingley, and Jane Vaynman for comments on earlier versions of this paper, and to Malfrid 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.

Bas, Muhammet A., and Andrew J. Coe. (2018) Give Peace a (Second) Chance: A Theory of Nonproliferation Deals. *International Studies Quarterly*, doi: 10.1093/isq/sqy015

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. This is because the proliferant would now face more severe punishment if caught cheating on a deal and can also expect to receive more favorable concessions for halting its program. 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 thus encourage proliferation. If it expects that pursuing a weapons program will 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, as well as more inclined to pursue a program in the first place. Thus, while a late deal—such as the US-North Korean Agreed Framework of 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, as it signals to them that they can hold out for a better deal once the program has advanced.

This provides a new explanation for the phenomenon of states, such as Japan, that possess the technological prerequisites for nuclear weapons but have chosen not to build them (Fuhrmann and Tkach 2015; Mehta and Whitlark 2017a, 2017b). If a state is not going to 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. These data reveal some surprising patterns. Of the 30 states that were judged capable of pursuing a nuclear weapons program, 27 (90 percent) agreed to a nonproliferation deal at some point, and 23 (85 percent) of those states made an early deal. Of the ten states that had the opportunity to make a late deal, 6 (60 percent) did so. Overall, a deal was made in 74 percent of the instances that our theory identifies as opportune times for a deal. Because most proliferation attempts end in deals, we focus our testing on our expectations about the timing of, concessions made in, and implicit threat behind these agreements, as opposed to whether one occurs at all.

We find substantial support for the theory. First, 29 of the 32 deals in the dataset occur either early in a state's program—when it is known to be rudimentary or not yet begun—or late—once it is estimated to be nearing success. Only three deals happen in the period between these two points. Second, all six late deals coincide with a newly credible US threat to attack or abandon the state. Third, four of the five instances of states that made both early and late deals received more generous concessions in their later agreements.

Ours is, to our knowledge, the first analysis of all known nonproliferation deals. Some previous formal work points out the theoretical possibility of bilateral nonproliferation deals but does not analyze them in detail or study them empirically (see Bas and Coe 2016; Debs and Monteiro 2014; Spaniel 2015).[1] These studies, and the closely related formal literature on building conventional military forces, focus on determining the conditions under which proliferation or arming, or war to prevent it, will occur, with the latter typically setting aside the potential for avoiding both through a deal.[2] We explicitly incorporate the possibility for such deals and analyze 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 (Verdier 2008; Coe and Vaynman 2015; Gheorghe 2017). A rich empirical literature studies specific deals, but not the entire set (Braut-Hegghammer 2008; Reardon 2010; Coutto 2014; Miller 2014; Budjeryn 2015, 2016; Volpe 2015). Our theory generalizes insights from these previous studies. It offers a set of empirical implications that should hold for all nonproliferation deals.

Our model is among the first to treat states' arming as a dynamic process—that is, to incorporate the fact that new military capabilities are acquired over time. We show that a focus on temporality is essential 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. 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. Thus, it faces uncertainty about when it will no longer be safe to put off the costs either of conflict or of the concessions needed to produce a viable deal.

Previous models take arming to be static—they assume that capabilities are acquired immediately—and so cannot speak to the unfolding of these interactions over time (Powell 1993; Feaver and Niou 1996; Baliga and Sjöström 2008; Meirowitz and Sartori 2008; Jackson and Morelli 2009; Benson and Wen 2011; Debs and Monteiro 2014).[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, many cases involve little empirical evidence of uncertainty over the existence of a program—as opposed to its progress—at any time when a state might seriously consider an attack (Bas and Coe 2016, 672–73). 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 borne out in most.

[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).

[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 a set of disputed issues, represented by the unit interval.[6] In the first of infinitely many discrete periods of time, A first chooses whether to initiate costly conflict. If A initiates, 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.

We intend "costly conflict" 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 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, conflict may entail ending the alliance, as the US threatened to do with South Korea, Taiwan, and others. A's choice 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 obviously differ empirically, but they share all the strategic features relevant to our model. First, each imposes serious costs on both sides. An attack results in death and destruction. Abandonment necessitates the loss of the benefits of alliance—perhaps especially the deterrence of an outside enemy. Second, the outcomes 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. Attack accomplishes this directly; abandonment indirectly by exposing 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 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 economic and political sanctions that are often imposed 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]

If B peacefully accepts A's offer, then B can invest or not in developing nuclear weapons (that is, 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 disregard 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 identify as 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.

Overcoming these hurdles is partly a result of trial-and-error. This means that neither player can perfectly predict the time at which B will master one and then the next. If B begins a round in the second stage and chooses to invest in that round, then he remains 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 $1 - \varepsilon$, 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 $\varepsilon\lambda$, so that it is possible to master both stages in a single round.[12]

This representation of the weapons development process is our central analytical innovation; many of our results flow from it. It has two virtues. First, it is the simplest way to incorporate 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 state will get nuclear weapons. While 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 increases. In the absence of contradicting intelligence, A's estimate of how long it will be until B acquires 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 does better with nuclear weapons than without in a war with A.[13] If instead B is A's ally, it means that B would do better with nuclear weapons than without if the alliance

[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 is valid so long as abandonment results in some partition of the alliance's total value between the two sides, less the joint benefit created by the alliance itself.

[8] Subsequently, we will discuss the robustness of our results to allowing sanctions to be costly.

[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 this is true: 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.

[10] Later, we will consider what happens when this assumption is relaxed.

[11] More realistically, nuclear weapons programs may progress through more than two potentially observable stages. 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.

[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. This lowers the former's probability of winning—just as we assume here. 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.

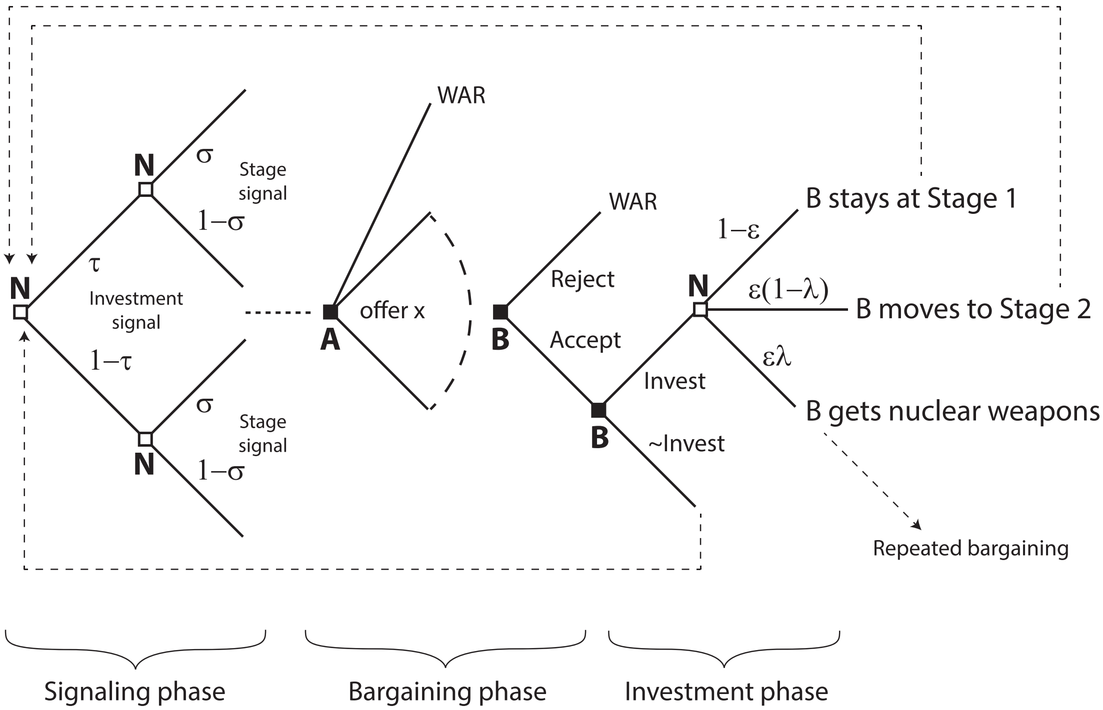

Figure 1. A single period of the game

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). Some evidence suggests that states seeking nuclear weapons also believe it (Brands and Palkki 2011; Narang 2014).

The first period ends after chance determines *B*'s progress or lack thereof. The next period, and every subsequent period, differs in structure from the first only in that it begins with *A* possibly receiving new intelligence. This takes the form of two signals, which we assume to be common knowledge.[14] The first signal concerns the program's existence: it indicates whether *B* invested in the last period or not. If *B* invested, 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 concerns the program's progress: it indicates the current stage of *B*'s program. *A* receives a true signal of *B*'s current stage with probability σ and an uninformative (that is, "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.

We assume each player's per-period payoff is 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*. 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 *B* agrees not to pursue nuclear weapons. Each side's willingness to comply with a deal depends on what would happen in its absence. So, to determine whether any deal is viable, we must first specify what would happen if no deal were made, or if a deal fails—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.

[14] In the equilibria studied here, *A* has no 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, undermining its viability.

[16] A low probability of false second-stage signals would make costly conflict more likely in the absence of a deal. This increases 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.

As a result, *A* is content to watch and wait in order to put off the costs of conflict.[18] As time passes, *B*'s program is increasingly likely to have made progress. Without new intelligence that it has not, *A*'s estimate of the time until it succeeds shortens.[19] Thus, the risk of imminent proliferation grows, and *A*'s tradeoff begins to shift in favor of costly conflict. If *B*'s program is slow enough, so that its success is unlikely even in the penultimate stage of development, and the shift in power from *B* acquiring nuclear weapons is small enough relative to the costs of conflict, then *A* would simply tolerate *B*'s program indefinitely. In this case, eventual proliferation is inevitable. Otherwise, *A*'s estimate of the time until *B* has weapons may eventually grow short enough that *A* would rather initiate conflict than tolerate any further risk of proliferation.[20]

If *A* is 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 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.

The Viability of Possible Deals

The two sides can avoid proliferation or costly conflict only if they make a deal. Under a deal, *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 noninvestment. *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 foregoing a deal poses no risk of eventual costly conflict, then no deal will be made.[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 viable deal, two things must be true. First, *B* must receive at least a little higher value from the deal than he would get without it. 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, *A* cannot receive any less value from a deal than she would get without it, 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 is jointly costly for the two sides. In our model, the only source of such costs is 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 result does not imply 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*. Offers generous enough to win *B*'s compliance simply demand too much from *A*, and offers *A* would actually be willing to make would only lead to *B*'s cheating.

Instead, this result means 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. Any deal that is agreed 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, so long as *A* is confident enough that *B*'s program is advanced (in that it has mastered fissile material production), then even if *A* is still a little uncertain about it, she 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]

[18] Of course, in some empirical cases, preventive attacks were undertaken 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. In reality, *A* may also be uncertain of its speed (that is, 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. We nonetheless 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. The closer a program is to success, the more facilities there are to be rebuilt after a strike. This implies that *A* can impose a longer delay by attacking when the program is closer to success, again encouraging her to postpone action until she estimates success is near.

[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.

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

We will document in the empirical section that most proliferants make a deal at some point. In light of this, 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. We divide the times at which a deal might be made into three periods: 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. The prospects for a viable deal vary 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—is harder.[26]

To understand the intuition for this result, observe that two factors, that together govern the viability of a deal, change as time passes. First, *B*'s temptation to cheat on a potential deal grows as his program makes progress. Starting from the advanced stage of progress, his program is more likely to succeed in the next period. He therefore has more to gain from reneging on an agreement, so that more generous concessions will be required to assure his compliance. If *A* believes there is some chance *B*'s program is advanced, she can only be assured of nonproliferation if she offers these more generous concessions to *B*. It is not enough to offer concessions that would only satisfy a *B* whose program remained rudimentary: if his program had made unobserved progress, then he would cheat on the deal. Thus, 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, or equivalently, as long as *A* is not sure that *B*'s program remains rudimentary.

Second, the severity of the penalty for reneging 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 know that *B*'s program is rudimentary. If she catches *B* cheating, she will therefore elect to wait and see rather than attack or abandon *B* immediately, opting only to retract her concessions in the meantime. This is 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 so 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*.

We can combine these two observations as follows. As long as there is some probability that *B*'s program is advanced, *A*'s concessions must be generous enough to overcome *B*'s potentially high temptation to cheat. But the more confident *A* is that B's program is advanced, the more severe a punishment *A* can credibly threaten in response to catching *B* cheating. Thus, when *A* is not sure *B*'s program remains rudimentary, a deal is most likely to occur once *A* is confident enough that *B*'s program has advanced to immediately attack or abandon *B* if she detects cheating.

By contrast, consider the situation when *A* is sure *B*'s program is rudimentary. *A* knows *B*'s temptation to cheat is low, but the severity of the punishment she can credibly threaten is also low. Depending on the parameter values, the single best opportunity for a deal may come 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 means that, even if no deal was agreed or complied with in earlier years, a viable deal may arise in later years, once a proliferant's program is believed to be nearing success. 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 a weapons program.

For instance, one might presume that Iran's agreement to temporary restrictions on its program is insincere. After all, the US and others had already negotiated with Iran for a long time, and no deal was forthcoming. Iran has already cheated on an earlier deal implicit in the Treaty on the Nonproliferation of Nuclear Weapons (NPT). The current deal came only after increasingly open discussion of attacking Iran in the US and Israel, so it seems plausible that Iran may only be playing for 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 yields two observable implications:

H1: *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.*

H2: *A late deal should coincide with a newly credible threat to attack or abandon the proliferant in its absence.*

[25] 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 proliferant 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 discovered 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.

Anticipating a Better Deal

It turns out that a viable late deal must concede more to the proliferant than any viable early deal. This later generosity poses a serious obstacle to the viability of any earlier deal, and can even encourage a proliferant to pursue a program in the first place, thereby 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. An enemy with a rudimentary or nascent program agrees with the US to an early deal. The threat of reverting to the no-deal equilibrium renders the deal viable. However, after making the deal, 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 guarantees 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 mutually beneficial way to avoid both war and proliferation? Faced with this situation, the US would be sorely tempted to make a new deal rather than to attack. From the enemy's perspective, this new deal would be even better than the early one, as the US would have to concede more to overcome the enemy's higher temptation to cheat if its program is advanced.

H3: *A late deal should be more generous than an early deal.*

Because a late deal will be more generous, the proliferant is severely tempted to cheat on an early deal if it anticipates that the US would later deal rather than attack. Pursuing nuclear weapons would no longer be a risky bet on getting 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 be 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.

Empirical Tests

We proceed to test our hypotheses. Our universe of cases is the set of all states which could ever have made a meaningful nonproliferation deal. This means that each state must possess, or be suspected of possessing, an interest in nuclear weapons and the capability to pursue them. This set includes every state that 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 additional states that, 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 these weapons or the ability to quickly produce more, but they had the option of pursuing both and so were potential proliferants. We exclude the states that could not plausibly have acquired nuclear weapons, such as Afghanistan or Cameroon, as well as the states that showed no 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, ruling out any 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] Their alliance with a superpower protected the UK, France, and China from preventive attack by the opposing superpower due to the anticipation that this would escalate to a large, prohibitively costly war between the superpowers. These three states were also not threatened with abandonment over their nuclear programs, because, in the early years of the nuclear era, both superpowers saw advantages to the spread of nuclear weapons to their allies (Coe and Vaynman 2015, 989). 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 no 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 NPT as a deal to which the model applies. The NPT can be viewed as an efficient

[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.

[28] Following Montgomery and Sagan (2009), we use both datasets to ensure that all of our tests are robust to 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.

[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.

Table 1. States judged capable of pursuing nuclear weapons

State and period	*H1: deal timing*			*H2: late deal ⇒ threat*	*H3: late > early deal*
	Early	*Middle*	*Late*		
Belgium	√			—	—
Canada	√			—	—
Czechoslovakia	√			—	—
Italy	√			—	—
Japan	√			—	—
Poland	√			—	—
W. Germany	√			—	—
Algeria, 1983–(e)	√			—	—
Indonesia, 1965–1967(e)	√			—	—
Romania, 1985–1990(e)	√			—	—
Sweden, 1946–1969(e)	mixed			—	—
Switzerland, 1946–1970(e)	√			—	—
Yugoslavia, 1954–1965(e), 1974–1988(e)	√			—	—
Argentina, 1968–1977(e), 1978–1990(p)		X		—	—
Australia, 1956–1960(e), 1961–1973(p)	mixed			—	—
Brazil, 1953–1977(e), 1978–1990(p)		X		—	—
Egypt, 1960–1964(e), 1965–1974(p)	√			—	—
Iraq, 1976–1982(e), 1983–1995(p)	√			—	—
Syria, 2000–(p)	√			—	—
Iran, 1976–1984(e), 1985–(p)	√		√	√	√
Libya, 1970–2003(p)	√		√	√	√
N. Korea, 1965–1979(e), 1980–1905(p)	√		√	√	√
S. Korea, 1959–1969(e), 1970–1978(p)	√		√	√	√
Taiwan, 1967–1977(p), 1987–1988(e)	√		√	√	X
Belarus	√			—	—
Kazakhstan	√			—	—
Ukraine	—	—	√	√	—
Israel, 1949–1957(e), 1958–1968(p)				—	—
Pakistan, 1972–1986(p)				—	—
S. Africa, 1969–1973(e), 1974–1978(p)				—	—

Table 2. Nonproliferation deals

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	1976–1981
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	1977–1978
Italy	NPT	1969	Ukraine	Lisbon	1994
Japan	NPT	1970	W. Germany	NPT	1969
Kazakhstan	Lisbon	1994	Yugoslavia	NPT	1968

substitute for a large number of bilateral nonproliferation deals (Verdier 2008), in which each state 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 judged capable of pursuing a nuclear weapons program, 27 (90 percent) 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 percent) 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 percent) did so. In aggregate, a deal is made in 74 percent of the instances that our theory identifies as opportune times for a 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]

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 the middle. To test this, we need to specify exactly what is meant by these times. Any deal that occurs in the middle is evidence against the theory. Because the choice of these specifications is somewhat arbitrary, the danger is that we end up labeling very few of the state-years in the data as the middle, 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 the middle.

We use two alternative specifications of "early." 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. For this we use 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 percent (46 of 204) of program-years in which a state is coded as pursuing a program (by SW) being labeled as early.

Next, we specify "late" 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 percent (31 of 204) of program-years being labeled as late.

The second column of Table 1 contains the results. The subcolumns correspond to the three possible times at which a deal could occur: early, middle, and late. Each deal in the data appears in a subcolumn according to our coding of the time in which it occurred. A checkmark 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 early and 6 occur late in a program, while only 3 occur in middle. Thus, 91 percent 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 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.

[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).

[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 (2018) 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).

[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 often specify 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 those 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.

[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 would be no potential for a deal.

Late Deals and Threat Credibility

Hypothesis 2 predicts that, when late 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 of the late deal, the relevant interlocutor had just become willing to initiate costly conflict over the proliferant's program. This is difficult to measure directly, so we resort to the proxy of whether that state began issuing threats of attack or abandonment, in public or private but high-level official communication, just before making the deal.

In the six late deals—with Iran, Libya, North Korea, South Korea, Taiwan, and Ukraine—the US is the primary interlocutor. We find that all six support H2, as shown in the second-from-right column of Table 1. Of course, threats might also have occurred in program-years where a late deal was not made. However, even if as much as 48 percent 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 and the EU-3 (France, Germany, and the UK) agreed to a deal that both sides understood as a means to avert a US and/or Israeli attack on Iran (Volpe 2015, 215–17). 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 not to assist the program (Volpe 2015, 202–7). 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, 212–14; Kerr 2016, 41). In mid-October 2003, Israel explicitly threatened 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 began soon after its 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 finally to have abandoned the option to eventually produce nuclear weapons (Reardon 2010, 226–36; 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 (Siler 1998, 75–76; Oberdorfer 2001, 71; Reardon 2010, 231).

Late vs. Early Deals

Hypothesis 3 predicts that late deals should feature more generous concessions than early deals. Because many other factors also affect the concessions a proliferant receives—such as its technological sophistication and the cost of attacking or abandoning it—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 and also the one case that does not support the theory.

Iran, Libya, South Korea, and Taiwan were offered no concessions in their early deals under the NPT, other than those written into the treaty and open to any member.[41] However, in 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. 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, 502–3).[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 (Siler 1998, 77; Reardon 2010, 231–32, 235–36).

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 1975–1979. 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." However, 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.

Broader Implications

For bargaining theorists, arming is mysterious for exactly the reason war is: why do rational actors employ costly measures (whether war or arming) to prosecute their disputes? We cannot know the answer if we do not understand why arms control—such as the nonproliferation deals we analyzed here—sometimes proves unachievable.

[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.

[42] These were large concessions by the US. The sanctions had debilitating effects on Libya's economy, 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, 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, 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.

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 we take the possibility of arms control into account, lowering the costs of war may render a deal viable and thus prevent war. We should determine which of this literature's conclusions are robust when arms control deals are allowed.

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, despite only ten ever building nuclear weapons. Mehta and Whitlark (2017a) characterizes 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. Cheating might stem from 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 (Corera 2006, 92, ch. 3). Alternatively, the cost to the US of an attack, or of the concessions made under a deal, might increase over time, leading either the proliferant or the US to end a deal. We suggest below that this may help to explain why Iran abrogated its 2003 deal with the EU-3. 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 might undermine the nonproliferation regime by shifting these expectations.

This suggests a distinct dynamic from the usual concern in nonproliferation policy circles about the normative precedents set by late deals. For example, the US-Iran deal of 2015 allowed Iran to retain its uranium enrichment capabilities, an unusual concession. Some worry that this undermines the norm against the spread of these capabilities; states can point to the Iran deal to justify their own nuclear programs (Kroenig 2014, 6; Kaplow and Gibbons 2015, 7–10). Our theory provides a more strategic version of this concern. The deal might weaken nonproliferation not only by raising questions of fairness about which states are allowed to have which nuclear capabilities, but also by causing states to believe that cheating on a deal will be met with further deal-making rather than conflict.[44]

Have some states been encouraged by other states' late deals to cheat? Does this influence US responses to cheating? Our theory does not specify how states form their beliefs about the likely response of powerful states to detected cheating on nonproliferation. Future research should examine how these expectations are formed and influenced.

For policy-makers, our theory highlights a dilemma that arises when the US must respond to a state's cheating on a deal. The death and destruction caused by preventive attack argue for giving a cheater every chance to agree to a new deal. However, allowing a cheater to extract additional concessions by virtue of progress in its illicit program can undermine deals with other potential cheaters. 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.[45] Consistent with our theory, the 2003 deal with the EU-3 unraveled late in 2005.[46]

By 2008, the US intelligence community again estimated that Iran's program was nearing success. Both the US and Israel issued threats. The absence of a new deal led to an apparent campaign in 2009–2011 to sabotage Iran's program through cyberwarfare and assassinations of key scientists.[47] Iran subsequently agreed to serious constraints on its nuclear program in November 2013 and July 2015, which were implemented starting in early 2014 and have continued to this writing (Kerr 2016). In exchange, some United Nations, European Union, and US sanctions on Iran were dropped and the conventional weapons embargo was ended. Iran retained the most commercially valuable aspects of its program, especially its ability to produce low-enriched uranium for power plants. These major concessions by the US, and its partners, will enable Iran to build its regional power. Indeed, critics of the deal, including the governments of Israel and a number of Arab states, objected on precisely these grounds (Gordon and Sanger 2015). Our theory makes clear why large concessions are necessary to support such a deal. But it also suggests that such a deal, made after so many years of deception and recrimination, may nonetheless prove effective and durable.

Supplementary Information

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

[44] See Tannenwald (2013) for an overview of the questions of fairness that plague the nuclear nonproliferation regime.

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

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

[47] 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).

References

Albright, David, and Paul Brannan. 2007. "The North Korean Plutonium Stock, February 2007." http://www.isis-online.org/publications/dprk/DPRKplutoniumFEB.pdf. Accessed January 19, 2018.

Baliga, Sandeep, and Tomas Sjöström. 2008. "Strategic Ambiguity and Arms Proliferation." *Journal of Political Economy* 116 (6): 1023–57.

Bas, Muhammet A., and Andrew J. Coe. 2012. "Arms Diffusion and War." *Journal of Conflict Resolution* 56 (4): 651–74.

———. 2016. "A Dynamic Theory of Nuclear Proliferation and Preventive War." *International Organization* 70 (4): 655–85.

Benson, Brett, and Quan Wen. 2011. "A Bargaining Model of Nuclear Weapons Development and Disarmament." In *Causes and Consequences of Nuclear Proliferation*, edited by Robert Rauchhaus, Matthew Kroenig and Erik Gartzke, 45–62. London: Routledge.

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): 133–66.

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*, January 15.

Budjeryn, Mariana. 2015. "The Power of the NPT: International Norms and Ukraine's Nuclear Disarmament." *The Nonproliferation Review* 22 (2): 203–37.

———. 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*, April 21.

Coe, Andrew J., and Jane Vaynman. 2015. "Collusion and the Nuclear Nonproliferation Regime." *Journal of Politics* 77 (4): 983–97.

———. 2018. "The Tragedy of Arming." Paper presented at the 2018 Annual Meeting of the Midwest Political Science Association. Chicago, IL. April 6, 2018.

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

Coutta, Tatiana. 2014. "An International History of the Brazilian-Argentine Rapprochement." *International History Review* 36 (2): 302–23.

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 Meeting of the American Political Science Association*. Washington, DC. September 4, 2010.

Feaver, Peter D., and Emerson M. S. Niou. 1996. "Managing Nuclear Proliferation: Condemn, Strike, Or Assist?" *International Studies Quarterly* 40 (2): 209–33.

Fuhrmann, Matthew, and Benjamin Tkach. 2015. "Almost Nuclear: Introducing the Nuclear Latency Dataset." *Conflict Management and Peace Science* 32 (4): 443–61.

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.

Gheorghe, Eliza. 2017. "Proliferation and the Logic of the Nuclear Market." Paper presented at the 2017 Annual Meeting of the International Studies Association. Baltimore, MD. February 25, 2017.

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*, July 14.

Hersman, Rebecca K. C., and Robert Peters. 2006. "Nuclear U-Turns: Learning from South Korean and Taiwanese Rollback." *The Nonproliferation Review* 13 (3): 539–53.

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): 167–94.

Kaplow, Jeffrey M., and Rebecca Davis Gibbons. 2015. "The Days After a Deal with Iran: Implications for the Nuclear Nonproliferation Regime." Accessed April 2, 2018. https://www.rand.org/content/dam/rand/pubs/perspectives/PE100/PE135/RAND_PE135.pdf.

Kerr, Paul K. 2016. "Iran's Nuclear Program: Status." Congressional Research Service.

Kroenig, Matthew. 2014. "Mitigating the Security Risks Posed By a Near-Nuclear Iran." *The Atlantic Council*, November.

Mehta, Rupal N., and Rachel E. Whitlark. 2017a. "The Benefits and Burdens of Nuclear Latency." *International Studies Quarterly* 61 (3): 517–28.

———. 2017b. "Unpacking the Iranian Nuclear Deal: Nuclear Latency and U.S. Foreign Policy." *The Washington Quarterly* 39 (4): 45–61.

Meirowitz, Adam, and Anne E. Sartori. 2008. "Strategic Uncertainty as a Cause of War." *Quarterly Journal of Political Science* 3 (4): 327–52.

Miller, Nicholas L. 2014. "The Secret Success of Nonproliferation Sanctions." *International Organization* 68 (4): 913–44.

Montgomery, Alexander H., and Adam Mount. 2014. "Misestimation: Explaining US Failures to Predict Nuclear Weapons Programs." *Intelligence and National Security* 29 (3): 357–86.

Montgomery, Alexander H., and Scott D. Sagan. 2009. "The Perils of Predicting Proliferation." *Journal of Conflict Resolution* 53 (2): 302–28.

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

National Intelligence Council. 2007. "Iran: Nuclear Intentions and Capabilities." Accessed January 19, 2018. http://graphics8.nytimes.com/packages/pdf/international/20071203_release.pdf.

Nuclear Threat Initiative. 2011. "Iran Nuclear Chronology." Accessed January 19, 2018. http://www.nti.org/media/pdfs/iran_nuclear.pdf?_=1316542527.

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

Powell, Robert. 1993. "Guns, Butter, and Anarchy." *American Political Science Review* 87 (1): 115–32.

———. 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: 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*, August 5.

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): 859–85.

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

Spaniel, William. 2015. "Arms Negotiation, War Exhaustion, and the Credibility of Preventive War." *International Interactions* 41 (5): 832–56.

Tannenwald, Nina. 2013. "Justice and Fairness in the Nuclear Nonproliferation Regime." *Ethics & International Affairs* 27 (3): 299–317.

Verdier, Daniel. 2008. "Multilateralism, Bilateralism, and Exclusion in the Nuclear Proliferation Regime." *International Organization* 62 (3): 439–76.

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*, edited by James Clay Moltz and Alexandre Y. Mansourov, 27–40. London: Routledge.