A Dynamic Theory of Nuclear Proliferation and Preventive War

Muhammet A. Bas and Andrew J. Coe*

Abstract

We develop a formal model of bargaining between two states, where one can invest in a program to develop nuclear weapons and the other imperfectly observes its efforts and progress over time. In the absence of a nonproliferation deal, the observing state watches the former's program, waiting until proliferation seems imminent to attack. Chance elements—when the program will make progress and when the other state will discover this—determine outcomes. Surprise proliferation, crises over the suspected progress of a nuclear program, and possibly "mistaken" preventive wars arise endogenously from these chance elements. Consistent with the model's predictions and contrary to previous studies, the empirical evidence shows that the progress of a nuclear program and intelligence estimates of it explain the character and outcomes of most interactions between a proliferant and a potential preventive attacker. Counterintuitively, policies intended to reduce proliferation by delaying nuclear programs or improving monitoring capabilities may instead encourage it.

Word Count: 14,000

*We are grateful to James Fearon, Robert Powell, Dustin Tingley, and Jane Vaynman for comments on earlier versions of this paper, and to Målfrid Braut-Hegghammer, Fiona Cunningham, David Palkki, Or Rabinowitz, Uri Sadot, and Joseph Torigian for sharing ongoing research and sources essential to our empirical analysis. 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 him during his fellowship.

Iraq, North Korea, and Syria are historical enemies of the United States and have all pursued nuclear weapons, a technology that could radically shift the balance of military power with the US in their favor. The US launched a decisive war in 2003 in part to prevent Iraq from ever obtaining nuclear weapons. But the US did not stop North Korea from doing so, despite long negotiations and at least one crisis in which war was threatened, in 1994. And the US neither attacked nor even threatened Syria with war as it pursued its nuclear weapons program. Why the radically different outcomes across these cases?

In the cases of Iraq and North Korea, there was also substantial variation in the relationship between each and the US over time. The final outcomes of war and successful proliferation occurred only after a long period of negotiations and threats. During this period, the US intelligence community obsessively monitored each state's nuclear weapons efforts, and sporadic crises arose in which the US prepared itself for war, and sometimes even seemed on the verge of striking, only to end with a fading threat of war and continued negotiation. What explains this drawn-out process and the occurrence of crises?

Now the focus of many commentators on US foreign policy has shifted to Iran.[1]. Will the US (or its ally Israel) attack Iran to halt its nuclear progress? Should the US simply tolerate Iran's efforts? Why doesn't the US attack now, rather than continue to risk Iran acquiring nuclear weapons covertly?

To answer these questions, we analyze a formal model in which two states bargain repeatedly over disputed issues, while one potentially invests in and makes progress toward acquiring nuclear weapons which, once deployed, would increase its bargaining power, and the other imperfectly observes its efforts and progress and decides whether to attack to stop it. 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

[1]See, for instance, Kahl (2012), Kroenig (2012), and Waltz (2012).

shift underlying factors and thereby improve outcomes.

In the absence of a nonproliferation deal, the proliferant invests in a nuclear weapons program in the hope that acquiring the weapons will enable it to extract concessions from the US. The US tries to monitor the progress of this program, and as time passes, becomes increasingly worried that it is nearing completion. If, faced with a program known to be on the verge of success, the US would not attack to prevent it, then eventual proliferation is inevitable. Otherwise, if the US becomes confident enough of a program's imminent fruition, it will attack. Whether the interaction ends in war or proliferation then depends on whether the program succeeds before the US becomes worried enough to attack.

Surprisingly, much of the variation in behavior is, from the point of view of the participants, essentially a result of chance. The specific instantiations of two stochastic elements—a proliferant's halting progress toward acquiring nuclear weapons, and the US's noisy intelligence on its current stage of development—can make the difference between a final outcome of war or peace, prevention or proliferation, and also determine whether the road to this outcome is quick and calm or long and tense. They can render the US surprised by a state's proliferation, or lead to slowly increasing apprehension about a proliferant's nuclear progress, peaking in crises which may end in war or merely a repeat of the cycle. A war may even turn out to be "mistaken," in the sense that the proliferant was not about to acquire nuclear weapons. In short, these variables can easily have as large an impact on the outcome as previously studied factors such as the anticipated costs of war and effects on the balance of power of a state's acquisition of nuclear weapons.

We derive three observable implications from the model, test them against the historical record of proliferation interactions, and find strong support for the model. First, intelligence estimates of nuclear programs, especially after the programs' early years, are focused on assessing the progress, rather than the existence, of these programs. Second, the serious consideration of preventive attack during peacetime is strongly associated with intelligence

estimates that the program in question is nearing completion. Third, the course of intelligence estimates during a nuclear program tends to determine the character and outcome of the interaction between the proliferant and a potential attacker.

Our results also imply that many commonly-advocated policies for suppressing proliferation may have the opposite of the intended consequences. For example, better intelligence can sometimes make proliferation less likely as the US gets better at determining when the proliferant is close to success and preventing it. But under other conditions, it can make proliferation *more* likely because the increased patience of the US in waiting for definitive intelligence gives the proliferant more time for its program to succeed. Increases in the sophistication of a proliferant's program generally *decrease* the probability of proliferation and can *increase* the chance of a mistaken attack because the US more than compensates for the proliferant's sophistication by attacking sooner.

Our model is the first to treat states' arming dynamically, which we show is essential to understanding most historical episodes of attempted proliferation.[2] Acquiring a consequential new military capability such as nuclear weapons 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 safe to put off the costs of preventive attack.

Previous models take arming to be static—capabilities are acquired instantly or in the

[2]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.

very next period—and so cannot speak to the unfolding of these interactions over time.[3] In models of nuclear proliferation, this assumption typically leads in equilibrium 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.[4] However, we find little empirical evidence of uncertainty over a program's existence at any time when preventive attack might be seriously considered. Worse, these models produce no definite predictions for what should happen in most empirical cases. When the costs of attack are low enough relative to the consequences of proliferation, so that preventive attack is potentially worthwhile, these models can predict only that attack *may* occur, since it is chosen at random.

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, even when preventive attack is potentially worthwhile, it occurs in some cases and at some times but not others.[6]

[3]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]Baliga and Sjöström (2008); Benson and Wen (2011); Debs and Monteiro (2014).

[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.

[6]Fuhrmann and Kreps (2010) finds support for the role of the costs of attack and consequences of proliferation in determining whether an attack occurs, but conclude that how close a program is to success does not affect whether it is attacked. However, the proxies used to measure proximity to success are too rough to provide a good test. For example, one proxy is the number of years the program has been ongoing, but this may say less about the proliferant's proximity to success than about its technological sophistication.

As a result, it yields specific predictions for all empirical cases, which we show are born out in most.

In this paper we focus on "no-deal" equilibria in which the two sides do not agree to a nonproliferation deal. A companion paper analyzes the viability of deals in which the proliferant eschews a weapons program in exchange for concessions from the US. The possibility of a deal does not undermine the importance of the results we report here, for two reasons. First, in most empirical episodes of states pursuing nuclear weapons, no deal is ever made, so that it makes sense to focus on what happens in a deal's absence. Second, we show in the companion paper that a deal's viability depends strongly on a parameter—how well the US can observe a proliferant's initiation (or restarting) of a weapons program, so that this cheating can be punished—that has no effect on the no-deal equilibria. Deal equilibria thus exist only under certain conditions, while the no-deal equilibria always exist, so that the results we report will remain valid whether a deal exists or not.

Setup of the Model

We model the interaction between two states, A ("the US," referred to using feminine pronouns) and B ("the proliferant," masculine), as they bargain over revisions to a prior division of a composite of disputed issues, represented by the unit interval.[7] In the first of infinitely many discrete periods of time, A first chooses whether or not to start a war with B. If A attacks, the game ends with a costly lottery. The value of this lottery to each player depends on the balance of military power between them, represented by A's probability of victory in

[7]It may seem that the principal dispute 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.

the war. The winner receives the entire contested stake in this and all future rounds; the loser gets nothing. Regardless of who wins, each player pays a positive cost of war, c_A and c_B respectively, in this and all future periods.

"War" here is intended to represent the most cost-effective option available to A for unilaterally ending or delaying B's program. This may be a full-on invasion, such as the US war with Iraq in 2003, or it may instead be a limited set of strikes intended to destroy key nuclear facilities, as with Israel's attacks on reactors in Iraq and Syria. Which of these is chosen will depend on the costs of each option and its anticipated effectiveness at ending or delaying B's acquisition of nuclear weapons, but this choice is not the focus of our analysis.[8]

If A chooses not to attack, then she must offer to B a disposition of the contested issues. If B rejects the offer, war results, ending the game with the same costly lottery. If he accepts the offer, it 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. We ignore any cost incurred by A in imposing sanctions.[9]

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.[10] We also abstract away from any direct (budgetary) cost of B's

[8]Our results would not change qualitatively if we allowed play to continue after limited strikes, rather than ending the game, so that B could choose to continue its program and A might eventually attack again.

[9]Incorporating a small cost would make preventive attack more likely to occur in the absence of a deal, and earlier, because the costs of sanctions lessen the surplus from peace.

[10]It can be shown that, if A's ability to monitor the size of B's investment is low enough,

investment, taking this to be negligible.[11]

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 "early" stage of development where B has mastered neither, a second or "late" 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.[12] B begins the game in the early stage.

Overcoming 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 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

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.

[11]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 attack.

[12]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 some "late enough" stage at which A would attack, and earlier stages in which A would not.

probability $\epsilon\lambda$, so that it is possible to master both stages in a single round.[13]

This representation of the weapons development process is the central analytical innovation of this paper; 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 B 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 is successful and he acquires nuclear weapons, this immediately becomes common knowledge (e.g., because of an easily observable test detonation).[14] The balance of power shifts in the next period, from p to $p_n < p$, in B's favor.[15]

This assumption—that nuclear weapons improve B's power—is crucial to all that follows.

[13]As time goes by and the proliferant remains at a given stage, the probability of mastering that stage in the next period 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.

[14]Some states (e.g., Pakistan) are thought to have obtained nuclear weapons well before their first test. While a proliferant may delay testing for diplomatic reasons, it will have strong incentives to find some other way to reveal to its enemies that it is now nuclear-armed.

[15]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 war; incorporating this would not qualitatively

If they do not, then B has no incentive to acquire them, A has no reason to prevent this, and neither proliferation nor war occur. Despite the ongoing debate over this assumption, most US policymakers seem to believe it, and there is some evidence that states seeking nuclear weapons do also.[16] Importantly, this assumption means that our model applies only to those states that anticipate gains from possessing nuclear weapons. It sets aside the many real-world states that, for whatever reasons—reliable security guarantee, benign security environment, moral commitment to nonproliferation—see no net benefits from possessing nuclear weapons. These states would not invest in a nuclear weapons program, and so will not be subject to preventive attack to stop such a program.

[Figure 1 about here.]

The first period ends after B's progress or lack of progress is determined. The next period, 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 that are assumed to be common knowledge.[17] 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 $\tau > 1/2$ 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 $\tau' > 1/2$, and a signal that he did with probability $1 - \tau'$. Thus, A's intelligence on B's

change the results.

[16]For recent contributions to this debate, see Beardsley and Asal (2009), Kroenig (2013), and Sechser and Fuhrmann (2013). Gavin (2012) argues that most US policymakers have subscribed to this assumption, and Brands and Palkki (2011) and Narang (2014) provide evidence that some proliferants' leadership did also.

[17]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.

investment is prone to both false positives and false negatives. 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.[18] 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.

Proliferation or War

To characterize the no-deal equilibria of the game, we first examine what happens when A knows that B's program has reached the later stage of progress, so that proliferation is relatively near. This strongly affects what happens earlier in the game, when A is not sure that B's program has reached this stage, so that proliferation may not be near.[19]

A "no-deal equilibrium" is defined to be a Perfect Bayesian Equilibrium in which A never makes a positive offer that is more generous than is necessary to avoid immediate war.[20] Intuitively, the only reason A would ever make such an offer is to induce B not to invest, in the context of a deal trading these concessions for non-investment.

[18] We will show later that allowing for false stage signals would not alter the character of equilibrium behavior.

[19] Proofs of all the propositions appear in the online appendix.

[20] More precisely, it is a PBE in which, at the beginning of every period, either B's continuation value is equal to his war value or A's offer in that period is $q = 1$.

Proliferation Is Near

We start with the "second-stage-known" subgame, when B has not yet acquired nuclear weapons, but is known with certainty by A to be in the second stage—because at some previous point in the game, A received a signal that B was in the second stage. We will show that, in the absence of a deal, B will always invest in a nuclear program, and then analyze the conditions under which A will either tolerate this program or attack to stop it.

Proposition 1. *Suppose that B is in the second stage and this is common knowledge. In any no-deal equilibrium of this subgame, B always invests, given the chance. If $p - c_A + \frac{\delta\lambda}{1-\delta}(p - p_n) < 1 + \frac{\delta\lambda}{1-\delta}(c_A + c_B)$, there is peace and eventual proliferation. Otherwise, there is immediate war and no proliferation.*

To understand this result, consider B's perspective. In the absence of a deal, A will not reward non-investment, or equivalently, penalize investment, so it makes sense for B to invest as long as there is some benefit from acquiring nuclear weapons (recall we assumed away any direct cost of a nuclear program). Once B is nuclear-armed, A will offer B just enough to make him indifferent between war and peace. Before B has nuclear weapons, he is weaker (that is, he expects to do less well in a war), and A will concede even less, because B will still prefer this to war and, in the absence of a deal, there is no reason for A to do otherwise. Thus, B can expect that acquiring nuclear weapons will bring bigger concessions from A, and so it *always* makes sense to invest in the absence of a deal.

Now consider A's perspective. In a hypothetical world without nuclear weapons, A would offer B just enough of a compromise on their disputed issues to make B indifferent between accepting the offer and going to war—there is simply no reason for A to be any more generous than that. From A's perspective, the problem with B pursuing nuclear weapons is that, once B has acquired them, A will have to make a more generous offer to avoid war. In the absence of a deal, A really only has two choices for what to do. First, A could attack B to try to

prevent B from ever getting the weapons. Second, A could tolerate B's nuclear program. If A chooses toleration, then until B gets nuclear weapons, A can offer B even less than she would in the hypothetical non-nuclear world, because she can take advantage of B's expectation that the nuclear program will eventually be successful and lead to larger concessions from A. If instead A attacks, there will be no proliferation and subsequent concessions to B, but A will no longer have the chance to peacefully exploit B's hopes, and will lose the surplus from peace (the avoided costs of war) she would have enjoyed even once B became nuclear-armed.

Proposition 1 specifies this tradeoff for A. The left-hand side of each condition represents the immediate payoff of attacking for A $(p - c_A)$ and the gain from avoiding the change in concessions A would have to make in perpetuity after B had obtained nuclear weapons, weighted by the probability of B's program succeeding in the current period $(\frac{\delta\lambda}{1-\delta}(p - p_n))$. The right-hand side represents, in its first term, the immediate payoff of toleration: 1 is the most A could get in this period if B was willing to accept zero concessions for now to avoid war. Its second term represents the surplus A will enjoy from peace if B's investment this round is successful $(\frac{\delta\lambda}{1-\delta}(c_A + c_B))$. If the left-hand side is smaller, then A is better off tolerating B's program and taking advantage of B's hopes in the meantime. If it is larger, then A is better off attacking to stop it before it is successful.

War is more likely to occur when the effect of proliferation on the balance of power $(p-p_n)$ is higher. Then there is less for A to gain from taking advantage of B while his program is ongoing, and more to lose when it bears fruit. War is also more likely the smaller its costs; there is less surplus to lose from war, and if c_A is smaller there is also less for A to gain from taking advantage of B prior to proliferation. More surprisingly, the effect of the probability that B's program is successful in a given period (λ) is not always to raise the likelihood of attack. One might think that, the more likely B's program is to succeed, the sooner A will expect B to obtain the weapons and extract a better offer, thus making the commitment problem more severe and war more likely. But this happens only if the shift in power from

proliferation exceeds the costs of war ($p - p_n > c_A + c_B$). When instead $p - p_n < c_A + c_B$, the higher likelihood of proliferation is outweighed by the increased advantage A can take of B's more optimistic hopes for success while the program is ongoing, and thus the incentives for war relative to toleration decrease.

To focus the remaining analysis on the most interesting behavior, we will assume that the unique no-deal equilibrium, once A knows that B has reached the second stage, is immediate attack. It is easily shown that, if instead toleration is the no-deal equilibrium in this subgame, then preventive war cannot occur in equilibrium. Intuitively, because the perceived risk of successful proliferation is highest when A knows that B's program has advanced to the second stage, the commitment problem is most severe in this subgame. In earlier subgames, the perceived risk is lower and the commitment problem less severe because B's program might not have reached the second stage. So, if A tolerates B's program knowing that it is in the second stage, then she will also tolerate it earlier in the game, and there is never a preventive attack. In assuming that the no-deal equilibrium is war, we are setting aside cases in which proliferation is inevitable because A is unwilling to attack to stop B's program, even once it is known to be nearing success.[21]

Proliferation May Not Be Near

To analyze the earlier periods of the game, in which B may have reached the second stage but A does not know this with certainty, we first show that, in the absence of a deal, B will always invest, given the chance. This means that, without contrary evidence, A will grow increasingly suspicious that B's program has covertly reached the second stage. Once A is sufficiently convinced that this has happened, A will attack. Whether this happens, whether the war is "mistaken," and the character of the road to war or eventual proliferation will depend on when B's program makes progress and when A learns of this progress, both

[21] We will return to such cases in the empirical section below.

random from the players' perspectives.

Proposition 2. *Suppose that the second-stage-known equilibrium is immediate war. In any no-deal equilibrium of the game, B always invests, given the chance.*

In the absence of an agreement to avoid proliferation and war, A offers the bare minimum to B in each round. B will not get better offers unless and until he can successfully develop nuclear weapons. In any given round, investing in a nuclear program generates a chance that B will get, or at least get closer to acquiring (by mastering the first stage), nuclear weapons in the next round. By contrast, A would not reward B with more generous offers for not investing. Not investing thus only lowers the probability of having nuclear weapons in any future period, and lengthens the time during which B must suffer A's poor offers. For this reason, B should never miss a chance to invest in nuclear weapons production.

Given that B is always investing in his nuclear program, A's estimate of how likely B's program is to have reached the second stage at any particular point in time will depend only on the stage signals A has received up to that point. Since in equilibrium B always invests, signals of whether B is investing are irrelevant to A. Any signal that B has invested only confirms what A already knows; any signal that B hasn't invested must be false.

Proposition 3. *In any no-deal equilibrium, as time goes by without new intelligence on the stage of B's program, A's estimate of the probability that it has mastered the first stage and reached the second increases. If the first stage is easier to master than the second (i.e., $\epsilon \geq \lambda$), A will eventually become almost certain the program has reached the second stage. If instead the second stage is easier, A will never become confident the program has progressed—A's estimate will converge to $\frac{\epsilon-\epsilon\lambda}{\lambda-\epsilon\lambda} < 1$.*

Proposition 3 specifies A's estimate after a given time without a stage signal. To understand this result, consider that in each period without a signal of B's stage, A must weigh two contradictory pieces of evidence. On the one hand, B has not yet gotten nuclear

weapons, which suggests that in the last period his program was in the first stage rather than the second. On the other hand, time has passed and A knows B has been trying, so it is possible his program has mastered the first stage and moved to the second. Which of these weighs most heavily depends on whether it is easier for B to master the first stage or the second (i.e., whether ϵ is greater than λ). If the first stage is easier, then over time, in the absence of an informative signal of B's stage, A will eventually become almost sure that B has reached the second stage: intuitively, it is increasingly likely that B has mastered the easy stage but gotten stuck in the hard one. If instead the second stage is easier, then even after a great deal of time, A will still not be sure which stage B's program is at: he could be stuck at either stage, and the harder it is to master the first relative to the second, the more likely it is he remains at the first. Figure 2 illustrates the change in A's estimate as she waits without new stage intelligence for several different combinations of ϵ and λ. The limit to which A's confidence that B has reached the second stage converges is important for determining what happens in later periods, as we will see.

[Figure 2 about here.]

Before continuing, we assume that, when A knows that B's program is in the first stage, A will not attack immediately. While this behavior is technically possible in equilibrium, we view this as an artifact of our simplification of the process of nuclear weapons development to just two stages. At any known stage of a program, immediate attack is in equilibrium if and only if $p - c_A + \frac{\delta\gamma}{1-\delta}(p - p_n) \geq 1 + \frac{\delta\gamma}{1-\delta}(c_A + c_B)$, where γ is the probability of imminent proliferation (that is, the probability that the program succeeds in the very next period).[22] This condition can be satisfied only if γ is high enough. But empirically, in the early stages of a nuclear weapons program, the probability of imminent proliferation (γ) is approximately zero because there are simply too many remaining prerequisites for states to

[22]This condition is easily derived in the same way as the one given in Proposition 1.

master simultaneously. Thus, the condition for immediate attack in these early stages will not be met. Under this assumption, our results should be interpreted as governing what happens as B's program transitions from the last stage that A would knowingly tolerate to the first stage at which she would attack immediately rather than wait any longer.

Proposition 4. *Suppose that the second-stage-known equilibrium is immediate war. In any no-deal equilibrium of the game, A tolerates B's program unless and until she becomes sufficiently confident that it has reached the second stage, and then attacks.*

B's steady investment confronts A with a tradeoff between the risk that B's program will succeed in the next round and the costs of attacking to prevent it. Early on, B's program is likely to be in the first stage, so that the risk of proliferation is low and A is content to watch and wait, putting off the costs of attack. If A receives a new signal that the program remains in the first stage, then she can assume that the program is not going well and safely wait. If instead A receives a signal that the program has reached the second stage, then she can be certain that proliferation is near and attacks to end the high risk of proliferation. But if A receives no new stage intelligence, she will not know for sure how the program is doing. As time passes, B's program is increasingly likely to have made progress, so that A's estimate of the risk of proliferation grows (by Proposition 3), and A's tradeoff begins to shift in favor of preventive attack.

If her confidence that proliferation is near gets high enough, A will attack even without stage intelligence. It might then turn out that the program remained stuck in the first stage so that, in retrospect, it would have been better for A to wait rather than bear the costs of an unnecessary war. Going to war based on an (uncertain) estimate, rather than reliable information, is rational for A but poses the risk that the war is a "mistake." To be clear, if this happens, A is not mistaken about B's intentions—B is trying to obtain nuclear weapons—but is wrong about the progress he has made and thus the imminence of

his success. A might well regret such a war.[23]

This relatively simple description of the equilibrium conceals the variation in behavior that can occur as it unfolds. This variation is driven entirely by the chance successes of both B's nuclear program and A's intelligence-gathering. As a result of these stochastic elements, the game can end peacefully or violently, and with the occurrence of proliferation or its prevention. The final outcome can be reached quickly or slowly, and relations in the meantime can be pacific or crisis-prone. There are four generic kinds of paths along which the equilibrium might travel, illustrated in Figure 3.

[Figure 3 about here.]

1. **Surprise success**: A tolerates B's investment for the first few years since it is unlikely to be successful soon. But B's program masters both stages of development unexpectedly quickly, and B acquires nuclear weapons. The process is calm and ends quickly in proliferation.

2. **Hard intel of progress**: A is content to tolerate B's program initially, but receives intelligence that the program has advanced to the second stage and immediately attacks to stop the program. The process is quick and seems calm, but ends violently.

3. **Growing suspicion of progress**: Lacking recent intelligence on B's program, A grows increasingly apprehensive about the prospect of its imminent success. A crisis

[23]If false first-stage signals are possible, then a first-stage signal will still decrease A's estimate of the probability that B has reached the second stage, but not all the way to zero. Crises of the kind that occur in the third and fourth paths listed later will still defuse if A receives one or more first-stage signals, but more gradually. And if false second-stage signals are possible, A might need more than one second-stage signal to arouse her suspicions to the point of attacking. Both types of false signal would increase the likelihood of mistaken war because they make attacks without surety that B is in the second-stage more likely.

arises (because the time is near when A would be worried enough to attack without hard evidence) and war occurs. The process is potentially long and ends violently.

4. **Crisis defused**: Lacking recent intelligence, A's suspicions become persuasive and a crisis arises. War is threatened and appears imminent, but the arrival of intelligence that B's program remains in the first stage defuses the crisis, and the process continues. The process is drawn-out, tense, and dangerous.

The stochastic elements of the game (nuclear program progress and the receipt of new intelligence) determine which kind of path is actually observed and are therefore central to explaining the variation we see empirically. Because the exogenous parameters (σ, ϵ, λ, p, p_n, c_A, c_B, δ) are relatively stable over time, the model suggests that much of the over-time variation in behavior observed in empirical episodes is driven by the stochastic elements. This might also be true for variation in behavior *across* proliferants, though the exogenous parameters do change substantially across cases and so may explain some of the cross-country variation. However, as we will see, this dependence on chance does not mean the model cannot be used to generate testable predictions.

Observable Implications

In the model, the occurrence of proliferation or preventive attack is closely related to the contemporaneous intelligence estimates of the potential attacker. We use this connection to derive three observable implications of the theory.

First, in the no-deal equilibrium, the essential issue for a potential attacker deciding what to do is the degree of progress a proliferant's nuclear program has made. In particular, how far off is its anticipated success, or equivalently, how long does the attacker have to act before it's too late? By contrast, there should be little uncertainty about the existence of a nuclear program, since in equilibrium the proliferant always pursues one. This prediction

is importantly different from that yielded by previous models, which emphasize uncertainty about *whether* a state is seeking nuclear weapons at all but assume no uncertainty about progress.[24]

Hypothesis 1. *Intelligence estimates of nuclear programs should focus primarily on the progress of a program and the time at which it will succeed, rather than its existence.*

Next, if a potential attacker would *ever* find it worthwhile to use force to prevent a proliferant from acquiring nuclear weapons, then in equilibrium she will do so only once proliferation is near. At any earlier point, the attacker should bide her time, putting off the costs of attack until the risk of proliferation is higher. From Proposition 1, a state would never be willing to attack unless the shift resulting from proliferation is large enough relative to the costs of preventive attack. This also differs from previous models, in which the decision to attack is taken to be static, so that no prediction about *when* an attack will occur can be made. By contrast, in our model the decision to attack or not is made repeatedly over time as the situation faced by the potential attacker changes.

Hypothesis 2. *As long as the effect of proliferation on the balance of power is high enough relative to the costs of preventive attack, preventive attack should occur if and only if the program in question is estimated to be nearing success.*

Finally, given the observed course of a program's progress and a potential attacker's intelligence estimates of it, the model implies that the interaction between proliferant and attacker should match a particular equilibrium path type. That is, the model predicts both the outcome of the interaction (e.g., proliferation or attack?) and its character (e.g., proliferation a surprise? war a mistake?). For example, if a potential attacker estimates for several years that proliferation is not near, but then receives hard intelligence that the proliferant's program has made progress and is nearing completion, the model implies that

[24]Such as Benson and Wen (2011) and Debs and Monteiro (2014).

behavior will match the "hard intel of progress" path type, with peace up until the new intelligence is received, and then attack. Previous models are incapable of making such predictions, since they take the behavior of interest to be a one-time, random choice in a mixed-strategy equilibrium.

Hypothesis 3. *The proliferant/attacker interaction observed in each episode should match the equilibrium path type that corresponds to the observed course of the proliferant's program and the potential attacker's estimates of it.*

Empirical Tests

Our universe of cases comprises all those episodes in which a state pursued a nuclear weapons program. We take these episodes to be composed of all those state-years coded as pursuing nuclear weapons by Singh and Way (2004) (SW) or as having a nuclear weapons program by Jo and Gartzke (2007) (JG).[25] We exclude all those state-years in which a state was coded by SW as only "exploring" a nuclear weapons program, on the grounds that it is implausible that any other state would seriously consider a preventive attack in response to these usually quite tentative efforts.[26] We partition each state's set of years pursuing nuclear weapons into distinct episodes whenever some event occurs that can plausibly be regarded as "restarting" its pursuit of nuclear weapons, such as after a revolution occurs, a deal is agreed, or an

[25]Following Montgomery and Sagan (2009), we use both measures in order 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.

[26]Including these cases would actually strengthen the evidence for Hypothesis 2, because Fuhrmann and Kreps (2010) finds no instances of serious consideration of attack in these state-years, and the programs associated with these cases were almost never estimated to be anywhere close to producing nuclear weapons.

attack destroys the key facilities.

To measure outcomes, we use the first year of nuclear arms possession recorded by SW and JG for episodes that ended in proliferation, and the years in which preventive attacks occurred as documented by Fuhrmann and Kreps (2010) (FK) for those episodes that led to attack, updated to 2007. We also make use of data from FK on the occurrence or absence of "serious consideration of attack" (SCoA), defined essentially as a cabinet-level official deciding that the time is right for an attack. An attack must be seriously considered before it is undertaken, and almost half of SCoAs led to actual attack. Thus, even cases in which attack is seriously considered but not undertaken are relevant to testing our model's predictions about when such an attack is most likely to occur.[27]

Table 1 lists the full set of 27 episodes, with columns for each associated proliferant and the years it sought nuclear weapons, the states that seriously considered preventively attacking its program, and the outcome of the episode.[28] For the outcomes: "P" means proliferation; "I" indicates the episode was interrupted by an unrelated, exogenous event; "SCoA" stands for serious consideration of attack; "D" is for a nonproliferation deal; "A" means a preventive attack; and "O" indicates the episode is still ongoing as of 2007. Question marks are appended to each of the changes we have made to the original data.

[Table 1 about here.]

[27]Empirically, attacks that were seriously considered have been averted in two ways. First, by agreement to a last-minute deal instead, often under the threat of attack in the absence of a deal (North Korea 1994 with the Agreed Framework; arguably Iran 2003/5 and Libya 2003). Second, by the attacker's sudden realization that even though the time is right, the costs would be prohibitive (China 1964, when the US asked the USSR for permission to attack and was denied; also Pakistan 1986/7).

[28]Additional details on the handling of individual cases and construction of various measures can be found in the online appendix.

For data on intelligence, we draw on the remarkable compilation of US government intelligence estimates assembled by Montgomery and Mount (2014) (MM). This compilation covers 18 of our 27 episodes. (We use additional sources, detailed in the appendix, to cover the remaining 9 episodes.) It contains roughly 70 "concrete, verifiable estimates of overall [nuclear] capabilities and intentions" produced by the US intelligence community (10). We also use the appendix to MM, which offers a discussion of these estimates organized by episode and includes citations to many additional sources.

H1: Intelligence Estimates Primarily on Progress, not Existence

To test this hypothesis, we evaluated each of the 69 estimates on our episodes that were identified by MM. We counted the number that explicitly assess the progress of a state's nuclear program toward producing weapons, and also made a separate, more conservative count of how many explicitly offer a number of years until (or expected year at which) the program will succeed. We also counted the number that focus on the existence of a weapons program by answering whether a state intends to seek nuclear weapons, whether its program is oriented toward weapons, or whether a state intends to continue with an earlier nuclear weapon program. We then examined the estimates to ascertain any pattern in when they are concerned mainly with existence vs. progress.

We find that 41 of the 69 estimates deal explicitly with the state of progress of a nuclear program toward successfully producing weapons. Of these 40, 37 offer an explicit expected year of success or number of years to success. We find that 24 of the 69 estimates are concerned with the existence of a weapons program. This evidence supports our hypothesis, in that progress estimates outnumber existence estimates by almost two to one. However, given that one-third of the estimates deal with existence, it is plain that uncertainty about whether a state is pursuing nuclear weapons is also important.

Closer examination reveals a clear pattern in when existence or progress are the main

concern of an estimate. All but 3 of the 24 estimates concerned with existence are made in the early years of an episode, when a state has recently chosen to initiate or restart a nuclear weapons program (sometimes in violation of a nonproliferation agreement), and the US is attempting to assess whether the state is in fact seeking weapons. In the latter half or so of these episodes, the question of existence almost always gives way to confidence that a state is seeking weapons and a focus instead on assessing its progress toward the bomb. The only exceptions are South Africa, where the US remained uncertain of its weapons intent until its program was nearly completed, and Brazil and Argentina, where the last estimate in each episode concludes that a weapons program is being abandoned.

This pattern makes intuitive sense. When a state has just decided—or is in the process of deciding—whether to pursue nuclear weapons (e.g., India in the mid-1960s), or whether to renege on a deal and restart pursuit (e.g., North Korea in the late 1990s), there may be little concrete evidence of this decision, and US estimates turn mostly on tenuous judgments of the state's internal political calculus. However, as the program advances, large funds and staffs are allocated, new facilities are constructed and completed, and foreign materials or expertise are imported. These may all generate signals of nuclear weapons pursuit that remove ambiguity and lead the US to confidently assume that a weapons program is ongoing, so that intelligence estimates later in a program's course are devoted to assessing its progress and expected date of success.

H2: Preventive Attacks Associated with Near-Success Estimates

To test this hypothesis, we first need to determine which episodes meet the condition upon which this hypothesis depends—that the effect of proliferation must be high enough relative to the costs of attack—otherwise attack will never be considered regardless of intelligence estimates. We then use the intelligence estimates on each of these episodes to measure when a program was estimated to be nearing success, and compare this to the instances in which

preventive attack occurred or was at least seriously considered.

We exclude the US, USSR, UK, France, and India from the test set as not meeting this condition. Germany and Japan, the only plausible potential attackers of the US program, were incapable of launching an attack on facilities in the US homeland. The dominant explanation in the literature for the US failure to attack the Soviet program is that this would likely have required a major bombing campaign and escalated to a large war with prohibitive costs.[29] The USSR could reasonably anticipate that any preventive attack on the British or French programs would escalate to a large war with the US, making the costs prohibitive. During the 1960s, changes in how the superpowers dealt with proliferation rendered the British and French experience of pursuing nuclear weapons under the secure protection of a superpower unique: later allies suspected of doing so faced suspensions of aid and threats of abandonment.[30] Moreover, a superpower might plausibly refrain from defending an ally if another state preventively attacked its nuclear program, as evident in US efforts to obtain consent from the USSR for an attack on China's program.[31]

India was the only exception to this rule. The US actually considered giving India nuclear weapons in the 1960s, while the USSR signed a cooperation treaty with India and gave it substantial aid in the 1970s, suggesting that neither disfavored its proliferation enough to attack to stop it.[32] Of India's other potential attackers, Pakistan suffered internal turmoil after military defeats at India's hands in 1965 and, more decisively, in 1971, and so was in

[29]See Bas and Coe (2012) for a review of this literature, but also for the possibility that the US actually might have seriously considered attacking had it not underestimated the progress of the Soviet program. Treating the USSR episode in this way would strengthen the evidence for our hypotheses.

[30]Coe and Vaynman (2015); Miller (2014); Rabinowitz and Miller (2015).

[31]Burr and Richelson (2001).

[32]For US consideration of the value of India having nuclear weapons, see U.S. Department of State (1964, 122–125).

no position to attack its program, while China lacked the capability to attack the program except by invading or using nuclear weapons and thereby starting a large war with prohibitive costs.[33]

Table 2 shows the years of each of the remaining 19 episodes we use in the test. Each row ends with the last opportunity for preventive attack, taken to be the last serious consideration of attack or the acquisition of nuclear weapons, whichever is later. An opening square bracket (i.e., [) marks the beginning of intelligence monitoring, as we measure it. Before this point, a potential attacker is unaware of or not paying attention to a program and so obviously will not consider an attack. Thus, these years cannot reveal anything about the relationship between estimated progress and attacks, and so are excluded from the test.[34] Years in which a likely potential attacker estimated the program to be nearing success—within four years of acquiring nuclear weapons—are in bold.[35] Years in which preventive attack was seriously considered are boxed, while those in which an attack actually occurred are double-boxed.

[Table 2 about here.]

We counted every year after the start of intelligence monitoring in which the associated program was estimated to be nearing success, and the number of these years in which either

[33]China's long-range, conventional strike capabilities in this period were limited to a modest force of bombers based on the Soviet Tu-16. Because China lacked any air-refueling capability at the time, these large, slow, unstealthy planes would have had to fly to their targets in India without fighter escort. See U.S. Central Intelligence Agency (1974), especially pages 23 and 25.

[34]If any of these years are included in the test, our results are strengthened.

[35]Instead setting the threshold at three or at five years leads to qualitatively unaltered results. A threshold of more than five years seems implausibly long to claim that success is imminent, and a threshold of less than three years seems inappropriate given the level of imprecision inherent in these estimates, which often feature a range of two to three years.

an attack or serious consideration of attack occurred, and then did the same for every year in which the program was not thought to be near success. The results are presented in Table 3, with columns for program-years in which a program's success was estimated at more than four years away and at less than four years away.

[Table 3 about here.]

Preventive attacks on nuclear programs are fairly rare, considered seriously in only 36 of the 219 program-years in the data, with only 16 actual attacks resulting. More attacks occur with far-from-success estimates than with near-success estimates, both absolutely (14 to 2) and relative to program-years (.0753 to .0606). This would seem to contradict our theory, though the relative difference is not statistically significant ($p = .5$ in a one-tailed exact test). Including attacks that were only seriously considered, the same number occur with far-from-success estimates as with near-success ones (18), but in relative terms, SCoAs are over five times as likely to occur with near-success estimates (.0968 vs. .545), with the difference statistically highly significant ($p = .0000$). It appears that although serious consideration of attack is much more likely with a near-success estimate, many more of these considerations are resulting in actual attacks under far-from-success than under near-success estimates (14 of 18 versus 2 of 18).

The picture becomes clearer when we separate the data into program-years during which an unrelated war was ongoing between the proliferant and the potential attacker, and program-years in which no such war was occurring. Here, "unrelated" means simply that the war did not start because of the proliferant's nuclear program. The divided data are shown in the middle columns of Table 3: no such wars were ongoing in any of the program-years with near-success estimates, so these data are not divided. Almost all of the attacks (11 of 14) and most SCoAs (12 of 18) under far-from-success estimates took place during ongoing wars. Our model is set in a peacetime context, with bargaining ongoing between the prolifer-

ant and the potential attacker. Thus, strictly speaking, program-years in which an unrelated war is ongoing do not fit the scope conditions of the theory and so should be excluded from the test. Instead, we first test the hypothesis using only the peacetime program-years, and then return to the wartime data.

When we compare only peacetime years, our hypothesis is strongly supported. One more attack occurs under far-from-success estimates than under near-success estimates (3 vs. 2), but attack is more than three times as likely to occur with a near-success estimate (.0190 vs. .0606), though this difference is not quite significant statistically ($p = .16$). The results are even starker with SCoAs, with three-fourths of all peacetime SCoAs taking place with near-success estimates (18 of 24), and a SCoA more than fourteen times as likely to occur with a near-success than with a far-from-success estimate (.0380 vs. .545), with the difference highly significant ($p = .0000$).[36]

Turning to the wartime data, consider how this context would be expected to alter our model's predictions. There might be no bargaining during a war, though that by itself would not qualitatively alter our results. But there is another, more consequential difference: the costs of a preventive attack would be greatly reduced during a war. These costs are, principally: the expected retaliation by the proliferant and possibility of escalation to a large war; violation of the norm against unprovoked attack; and international condemnation or sanctions in response to the attack or the civilian casualties it causes.[37] But since the two states are already at war, a preventive attack would pose no risk of provoking a large war, and it might be hard for the proliferant to impose additional damage, not already being inflicted in the course of the war, in retaliation for the attack.[38] Similarly, there is no norm

[36]Notably, all 6 peacetime SCoAs and 3 peacetime attacks without a near-success estimate were undertaken by Israel, against Iraq and Syria. We will return to these below.

[37]Fuhrmann and Kreps (2010), p. 838.

[38]The unrelated wars in the data include World War II, the Iran-Iraq War, and the Gulf War, all large wars with the first two also total wars, so that the potential for escalation in

against attacking nuclear facilities as part of an ongoing war, and international condemnation or sanctions in response to an attack or the civilian casualties it caused would be unlikely. Moreover, these costs would be greatly reduced *only so long as the war continues.* In the model, if during some period the costs of preventive attack are greatly reduced, but expected to rise back to peacetime levels soon, there will be powerful incentives to strike, regardless of the stage of the program. A war presents a fleeting opportunity for a cheap preventive attack on a nuclear program, and is therefore extremely tempting even if the program is not estimated to be nearing success. Thus, the prevalence of SCoAs during wartime (12 in 28 wartime program-years), and the high frequency with which they result in actual attacks (11 of 12), are explicable by a simple extension of the model.

H3: Episodes Should Fit Corresponding Equilibrium Paths

The second-from-the right column of Table 1 specifies our matching of each of our 27 episodes to a possible equilibrium path type in the model, based on the course of the associated intelligence estimates by a potential attacker. The first group involves those cases, discussed earlier, where there is reason to believe that the costs of preventive attack were high enough relative to the effects of proliferation, so that, in accordance with Proposition 1, it would never be worth it for a potential attacker to use force to delay or halt a program. Since no attack or even serious consideration thereof occurred in these cases, they are assessed in the rightmost column of the table as supporting the theory.

The next group corresponds to the equilibrium path of "surprised by success": the proliferant made rapid progress toward nuclear acquisition, but this was not observed and so was underestimated by the potential attacker until it was too late (or would have been, in the case of Iraq 1982–91, were it not for the Gulf War). As expected, these cases show no serious consideration of attack because the potential attacker thinks it is safe to put off the

retaliation for a preventive attack was quite limited.

costs of attack until the risk of imminent proliferation is higher. This fits the conventional wisdom on Iraq's progress and estimates of it leading up to the Gulf War, though it can be taken as support for the theory only given the uncertain counterfactual that Iraq *would have* gotten the bomb if not for the war. We tentatively also assign North Korea 1995–2006 to this path type, but whether this is actually an accurate interpretation of the limited declassified intelligence on this episode remains to be seen; we therefore code its support for the theory as unclear.

For the episodes in the third group, estimates of the program in question were based on excellent intelligence and proved accurate, and eventually judged that nuclear acquisition was near. As expected in the "hard intel of progress" equilibrium path, these near-success estimates motivated a potential attacker to seriously consider taking preventive action, and so these cases support the theory.

The next group is similar, in that eventually the potential attacker estimated that success was near, but in these episodes the estimates were extrapolated from guesses about how long it would take the program in question to master the various technological steps, rather than from hard intelligence, and so were highly uncertain. As expected in the "growing suspicion" equilibrium path, these estimates almost always led to serious consideration of preventive attack, the one exception being Taiwan: we are unaware of any evidence that China seriously considered attacking Taiwan's program in 1976–77. The intelligence that led Egypt and the USSR to consider preventively attacking Israel's program, and India and Israel to consider attacks on Pakistan's program, proved accurate in retrospect. However, the near-success estimates of both Libya's and Iraq's programs in 2003 proved inaccurate because neither state's program had progressed much at all. Thus, the Iraq War was a mistake, in that proliferation was not likely to occur soon, and an attack on Libya's program would have been, too.

Up to the end of our dataset in 2007, the case of Iran since 1989 appears to match the

"defused crisis" equilibrium path. During the early 2000s, the US and Israel discovered more and more about the assistance Iran had received from the A. Q. Khan proliferation ring, and their estimates of the time until Iran's acquisition of nuclear weapons quickly ratcheted downward. We cannot be certain whether a preventive attack was seriously considered in 2003–05, when these estimates dropped below our threshold of four years or less.[39] But whether a SCoA occurred or not, in the later half of 2005, the US and Israel apparently received new information that led to revised estimates, placing the success of Iran's program in the middle of the next decade.[40] The receipt of new intel suggesting that Iran's program had not progressed as far as observers feared seemed to defuse the incipient crisis, and no attack occurred. Whether attack was seriously considered or not, the rise and fall of a crisis over Iran's program, corresponding to first more threatening and then suddenly more reassuring estimates, supports our theory.

The next group contains episodes that do not match any of the equilibrium path types in our model. These episodes saw preventive attacks occur even when it was clear that the attacked program was still in an early stage of development, contrary to our theory. Most of these attacks occurred during ongoing wars that were not fought over nuclear programs, with the attacker taking advantage of the temporarily-reduced costs of preventive attack to strike early: the US, UK, and Norway attacking Germany's heavy water plant during World War II; Iran attacking the Osirak reactor in Iraq and Iraq attacking the Bushehr reactor in Iran during the Iran-Iraq War; and the US attacking Iraq's nuclear facilities in the Gulf War. However, three attacks (drawn from six serious considerations of attack) occurred even in the absence of an ongoing war: Israel's attacks on Iraq's program in 1979 and 1981, and on Syria's program in 2007.

Our theory cannot explain why Israel chose to attack so early in these two episodes,

[39]Nuclear Threat Initiative (2011).

[40]National Intelligence Council (2007).

but we can offer some possible alternative explanations. First, attacking later is often more difficult militarily, since a more advanced program typically presents a larger number of targets that may be more fortified. While major powers usually have the military capability to attack at any point, smaller powers with more constrained capabilities may fear that an effective strike will become infeasible as the target set grows and is hardened, leading a state like Israel to attack early. Second, attacking later is cheaper politically, since as a program advances, its military intent becomes less plausibly deniable, making it easier to justify an attack and lessen the risk of harsh international reactions. For major powers especially, the declining political costs of striking later ought to outweigh the modestly increasing military costs. By contrast, Israel is already, to some extent, disfavored within the international community, and so may place little weight on the higher political costs of striking early. Finally, we can identify some idiosyncratic factors in each of Israel's three early attacks. The 1979 attack on Iraq's program involved only sabotage and assassination in third countries, and thus was far cheaper than the typical attack involving air strikes or invasion of the targeted country, which might have led to its early occurrence. At the time of the 1981 Osirak strike, Israel apparently estimated that Iraq was no more than five years away from getting nuclear weapons, close to our threshold for nearing success (four years or less).[41] Moreover, Iraq's military was heavily committed to the war with Iran so that the costs of an attack for Israel would be temporarily greatly reduced, as we discussed earlier. Last, considered in isolation, Syria's program was clearly not nearing success when Israel struck its not-yet-operational reactor in 2007. However, Syria's program was receiving extensive assistance from North Korea and possibly also Iran, and this outside help might have greatly shortened Israel's estimate of how long it would take Syria to get the bomb. Our data do not allow us to discern which, if any, of these alternative explanations is right, so further investigation of these anomalous cases would seem warranted.

[41]Sadot (2015), p. 24.

Finally, there is a group of episodes whose outcomes we cannot assess because the associated programs were interrupted early. But given that none of these programs was believed to be nearing success, our theory predicts that no attack or serious consideration of attack should have occurred. The absence of any serious consideration of attack in these program-years thus supports the theory.

In all, the overwhelming majority of the episodes feature the behavior our model predicts, given the course of intelligence estimates in these episodes.

Changing the Parameters of Proliferation

Although the stochastic elements of the game determine the particular path taken through it in equilibrium, the exogenous parameters determine the probabilities of the different paths. Thus, these parameters can still have substantial effects on the *expected* behavior, and so policymakers might be tempted to manipulate them to secure desired objectives. However, we will see that these parameters' effects on the outcomes are quite often counter to initial intuitions, and may sometimes be hard to predict. For brevity, we present only highlights here. The online appendix contains a comprehensive analysis including formal propositions and proofs.

In particular, we are interested in the effects of the proliferation-generated shift in of power ($p - p_n$), the costs of attacking to prevent it ($c_A + c_B$), the technological sophistication of the proliferant (ϵ and λ), and the quality of A's monitoring of the proliferant's progress (σ). All of these vary across cases, and all are potentially policy-manipulable. We focus on the effects these have on four statistical properties of the no-deal equilibrium: the probability of eventual proliferation; the probability of eventual preventive war; the probability of a mistaken war (in the sense that B's program was still in an early stage when A attacked); and the expected time to war or proliferation (the "length" of the game). These properties

seem likely to be the most relevant to policymakers.

Large-enough increases in the effects of proliferation, and large-enough decreases in the costs of preventive attack, reduce the chance of proliferation and the expected length of the interaction and increase the likelihood of war and mistaken war.[42] Generally, large-enough increases in the technological sophistication of the proliferant have the same effects.[43] These changes in the parameters make A less willing to wait before attacking in the absence of reassuring intelligence. Waiting longer before attacking gives A more time to enjoy the surplus from avoiding war, as well as more time for new intelligence on B's program to come in, possibly revealing that it remains at the first stage, so that A needn't attack after all. But it also exposes A to an increasingly large risk that B's program will succeed, forcing A to offer better concessions once B has nuclear weapons. When proliferation has stronger effects or is anticipated to happen sooner (because B's program is more sophisticated), or when attack has lower costs, this tradeoff is tipped in favor of attacking sooner because the risk of proliferation is higher relative to the benefits of delay. The quicker A will resort to attack, the fewer chances there will be for B's program to succeed, so that proliferation is less likely and war more so. Moreover, A will be readier to attack based on her suspicions rather than on hard intelligence, making a mistaken war more likely.

While these effects seem intuitive, they have some counter-intuitive implications for US policy. First, intuition suggests that new missile defense and preemptive strike capabilities would discourage proliferation because they would lessen the value of nuclear weapons for

[42]Smaller changes in these parameters have no effects on the outcomes, because they do not shift the optimal time for A to attack; see the online appendix.

[43]As shown through simulations in the online appendix, there are rare cases in which even a relatively large increase in ϵ or λ instead increases the chance of proliferation and the expected length of the game, and decreases the likelihood of war and mistaken war. Because of these rare cases, we use the modifier "generally" for this result.

proliferants as a counter to US military superiority. However, by rendering proliferation less bad for the US (akin to decreasing $p - p_n$), these capabilities would also lead the US to delay attacking a proliferant's program, raising the probability that the proliferant succeeds in getting nuclear weapons. This would make a potential proliferant more willing to pursue a program in the first place, so that acquiring such capabilities might actually *encourage* proliferation—the opposite of the intended effect. Similarly, any international effort to curtail the consequences of proliferation by reducing the resulting shift in power—say, by stopping the proliferation of effective delivery vehicles such as ballistic missiles—may encourage nuclear proliferation by rendering the US more tolerant of the risk.

Second, better capabilities for preventive attack, such as low-yield nuclear weapons for destroying deeply buried nuclear facilities, intuitively ought to deter states from pursuing nuclear weapons and thereby reduce the need for the US to launch preventive attacks. However, unless these new capabilities suffice to cause proliferants to abandon their programs altogether, they will only make preventive attack more likely because the US becomes less willing to delay war as a result of the lower costs or higher efficacy of preventive attack these capabilities would bring. They would also increase the likelihood of wars occurring when the US was mistaken about the degree of advancement of the target's program.

Finally, foreign assistance with nuclear technology or native technological sophistication may not be advantageous for a proliferant, at least to the extent that these are externally observable. If they increase the ease of a proliferant's development of nuclear weapons enough, they will lead the US (or another such state) only to attack sooner, so that the eventual acquisition of nuclear weapons actually becomes less likely and the proliferant will be left worse off. Paradoxically, even though the proliferant was more sophisticated, a US preventive attack launched without definitive intelligence that the targeted program had reached a late stage of progress is also more likely to be mistaken. In reverse, international efforts to cut off outside help and curtail the supply of nuclear materials and expertise, if

they are effective enough at slowing a proliferant's program, may actually make proliferation more likely as they render the US (or others) willing to wait longer before attacking.

Next, consider an increase in A's intelligence capability for monitoring the progress of B's program. This generally increases A's willingness to wait before attacking and decreases the chance of mistaken war.[44] With better monitoring, waiting is more likely to lead to new intelligence that will reassure A either that her suspicions are wrong (if a new signal reveals B's program has not made progress), or that she is right to attack now (if it instead reveals the program is nearing success). Because A is willing to wait longer before attacking without hard intelligence on B's progress, and more likely at any point to receive definitive intelligence, a mistaken war is more easily avoided.

However, better intelligence capabilities have competing effects on the chances of war vs. proliferation and the length of the interaction depends on the other parameters. On the one hand, better monitoring means that A is more likely to catch B once his program has made progress, shortening the game and making war more likely and proliferation less so. On the other hand, A is also more likely to detect it if B's program hasn't progressed, dispelling A's suspicion that proliferation is imminent and so giving B more time for his program to succeed, which lengthens the game and makes war less likely and proliferation more. For low enough σ and a short-enough game, the latter effect can dominate: B is likely to be in the first stage for most of the game, so that a stage signal to A is likely to buy B more time. If either of these conditions is not met, then B is more likely to reach the second stage and be exposed to detection and subsequent preventive attack, so that the former effect generally dominates.

The implication of these effects is that, although improvements to US intelligence capabilities would unambiguously lessen the likelihood of mistaken wars, they would not necessarily

[44]Once again, simulations reveal rare cases in which increasing σ has the opposite effect; see the online appendix.

reduce the probability of proliferation. If the US was unwilling to wait very long before attacking to halt the program, and had poor intelligence to start, then improvements in monitoring might actually *increase* the probability of proliferation by rendering the US more complacent. Whether improvements will lower the likelihood of proliferation thus depends sensitively on the initial conditions.

Broader Implications

Perhaps our most important result for empirical scholars is the finding that much of the over-time variation, and at least some of the cross-country variation, in proliferation interactions is driven by stochastic elements such as when a proliferant's program will make progress and when this will be observed by its opponent. Together with the small number of cases of nuclear weapons programs in the empirical record, this suggests that there are fundamental limits to our ability to make inferences about the role of exogenous factors, such as the effects of proliferation, the costs of war, the quality of intelligence, and the sophistication of a proliferant's program and its outside assistance. These factors affect only the *expected* outcome in a given case; the actual outcome is still highly variable because of the stochastic elements.

Many statistical studies to date have gotten around the small number of weapons programs by designating the country-year or dyad-year as the unit of analysis, a technique that greatly increases the number of observations and thereby the apparent strength of any patterns in the data.[45] These studies typically assume either that observations are independent over time, or that any unit-specific dependence over time decays rapidly. Our model implies

[45]Montgomery and Sagan (2009). For examples from this burgeoning literature, see Bleek and Lorber (2014); Brown and Kaplow (2014); Fuhrmann (2009); Jo and Gartzke (2007); Kroenig (2009); Singh and Way (2004).

that this approach may be problematic: behavior in the model depends heavily on what has happened in earlier time periods, and for some values of the exogenous parameters, this dependence may be long-lasting. This suggests that statistical analyses of the record can be improved by resorting either to more detailed, theoretically informed modeling of the temporal dependence in these interactions, or to the more conservative approach of treating a country or dyad, rather than a country-year or dyad-year, as the modeled unit of analysis.

Turning to formal theories of arming, our model assumes that the effects on the balance of power are more about whether a state has a particular weapon than about how many it has. This seems in keeping with much of the nuclear weapons literature, but it may be true of other military technologies as well. The focus so far on theoretical models of arming *quantity*, rather than quality, is undoubtedly useful. But in the modern era, the research and development of better weapons may be as important as the conscription of more soldiers and the manufacture of more guns, and raises different theoretical issues. The most salient of these is the fact that development takes time, so that the effect of acquiring the weapons is delayed, and subject to prevention, whether by war or by deal. Unfortunately, studying over-time variations requires the use of infinite-horizon, dynamic games that can be difficult to analyze, rather than the simpler finite-period or repeated games that are more typically used.

For policy-makers, the most important advice to follow from our work is that many commonly-advocated policy initiatives with respect to nuclear proliferation may have counter-intuitive and unintended consequences. Missile defense, preemptive strike capabilities, better intelligence, sabotage of proliferant programs, and diplomatic efforts to cut off outside assistance to a proliferant's program can sometimes improve the prospects for stopping proliferation, but they can also sometimes undermine them. Which occurs depends on whether the mooted policy initiative will be substantially or only modestly effective in shifting the underlying parameters, and also sometimes on the details of the particular case. This means

that each policy aimed at suppressing proliferation can only be evaluated by considering its effects on each potential proliferant the US will face, because a policy that discourages some potential proliferants may encourage others.

References

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.

Beardsley, Kyle, and Victor Asal. 2009. Winning with the Bomb. *Journal of Conflict Resolution* 53 (2):278–301.

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. Taylor and Francis.

Bleek, Philipp C., and Eric B. Lorber. 2014. Security Guarantees and Allied Nuclear Proliferation. *Journal of Conflict Resolution* 58 (3):429–454.

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

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

Burr, William, and Jeffrey T. Richelson. 2001. Whether to 'Strangle the Baby in the Cradle': The United States and the Chinese Nuclear Program, 1960–64. *International Security* 25 (3):54–99.

Coe, Andrew J., and Jane Vaynman. 2015. Collusion and the Nuclear Nonproliferation Regime. *Journal of Politics* Forthcoming.

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. 2009. Spreading Temptation: Proliferation and Peaceful Nuclear Cooperation Agreements. *International Security* 34 (1):7–41.

Fuhrmann, Matthew, and Sarah E. Kreps. 2010. Targeting Nuclear Programs in War and Peace: A Quantitative Empirical Analysis, 1941–2000. *Journal of Conflict Resolution* 54 (6):831–859.

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.

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

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

Kahl, Colin H. 2012. Not Time to Attack Iran. *Foreign Affairs* 91:166–173.

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

———. 2012. Time to Attack Iran. *Foreign Affairs* 91:76–86.

———. 2013. Nuclear Superiority and the Balance of Resolve: Explaining Nuclear Crisis Outcomes. *International Organization* 67 (01):141–171.

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`.

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

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

Rabinowitz, Or, and Nicholas L. Miller. 2015. The Last Line of Defense: US Nonproliferation Policy toward Israel, South Africa, and Pakistan. *International Security* Forthcoming.

Sadot, Uri. 2015. Osirak and the Counter-Proliferation Puzzle. Working Paper.

Sechser, Todd S., and Matthew Fuhrmann. 2013. Crisis Bargaining and Nuclear Blackmail. *International Organization* 67 (1):173–195.

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

U.S. Central Intelligence Agency. 1974. China's Strategic Attack Programs, NIE 13-8-74. Available at `http://www.foia.cia.gov/sites/default/files/document_conversions/89801/DOC_0001086044.pdf`. Accessed 28 February 2016.

U.S. Department of State. 1964. Memorandum of Conversation, November 23, 1964. In *Foreign Relations of the United States, 1964–1968, Volume XI, Arms Control and Disarmament*. U.S. Department of State, Office of the Historian.

Waltz, Kenneth N. 2012. Why Iran Should Get the Bomb. *Foreign Affairs* 91:2–5.

Proliferant and Years	Potential Attacker	Observed Outcome	Category / (Path Type)	Supports Theory
US, 1942–45	—	P	No SCoA ever	Yes
USSR, 1943/45–49	—	P		Yes
UK, 1941/47–52	—	P	(not worth	Yes
France, 1954–60	—	P	attacking)	Yes
India, 1964–65/74	—	P		Yes
Iraq, 1982–91	US	I / P?	No SCoA now	Yes?
N. Korea, 1995–2006	US	P	(surprise)	Unclear
China, 1955/56–64	US/Taiwan	SCoA → P	SCoA	Yes
S. Africa, 1971/74–79	USSR	SCoA → P	(hard intel)	Yes
N. Korea, 1980/82–94	US/S. Korea	SCoA → D		Yes
Israel, 1955/58–66/69	Egypt/USSR	A → P	SCoA	Yes
Taiwan, 1967–76/77	China	I	(growing	No
Pakistan, 1972–87	India/Israel	SCoA → P	suspicion)	Yes
Libya, 1970–2003	US/UK	SCoA? → D		Yes?
Iraq, 1992–2003	US/UK	A		Yes
Iran, 1989–2007	US/Israel	SCoA? → O	SCoA? (defused)	Yes
Germany, 1941–45	US/UK/Norway	A (42–45)	Early	War
Iraq, 1973–81	Iran/Israel	A (79–81)	attack	War/No
Iraq, 1982–1991	US	A (91)		War
Iran, 1984/85–88	Iraq	A (84–88)		War
Syria, 2000–07	Israel	A (2007)		No
Japan, 1943–45	—	I	Interrupted	Yes
Australia, 1961–73	—	I	by	Yes
Egypt, 1965–74	—	I	exogenous	Yes
S. Korea, 1970/71–75/78	—	I	event	Yes
Iran, 1974–78	—	I		Yes
Argentina, 1976/78–90	—	I		Yes
Brazil, 1978–90	—	I		Yes

Table 1: Episodes of **P**roliferation or (**S**erious **C**onsideration **o**f) Preventive **A**ttack

Episode											
Iraq	[82	83	84	85	86	87	88	89	90	91	
N. Korea	[95–96	97	98	99	00	01	02	03	04	05	06
China	55	[56	57	58	59	**60**	**61**	**62**	**63**	**64**	
S. Africa	[71	72	73	74	75	**76**	**77**	**78**	**79**		
N. Korea	80–84	[85	86	87	88	89	90	**91**	**92**	**93**	**94**
Israel	55–57	58	59	[60	61	**62**	**63**	**64**	**65**	**66**	**67**
Taiwan	[67	68	69	70	71	72	73	74	75	**76**	**77**
Pakistan	72–73	[74–78	**79**	80	81	**82**	**83**	**84**	**85**	**86**	**87**
Libya	70–74	[75–94	95	96	97	98	99	00	01	02	**03**
Iraq	[92–93	94	95	96	97	98	99	00	01	02	**03**
Iran	[89–97	98	99	00	01	02	**03**	**04**	**05**	06	07
Germany	[41	42	43	44	45						
Iraq	73	[74	75	76	77	78	79	80	81		
Iran	[84	85	86	87	88						
Syria	00	01	02	03	04	[05	06	07			
Japan	43	44	[45								
Australia	61–63	64	65	66	67	[68	69	70	71	72	73
Egypt	[65	66	67	68	69	70	71	72	73	74	
S. Korea	70	71	72	73	[74	75	76	77	78		
Iran	[74	75	76	77	78						
Argentina	[76–80	81	82	83	84	85	86	87	88	89	90
Brazil	[78–80	81	82	83	84	85	86	87	88	89	90

Table 2: **Near-Success Estimates** and (Serious Consideration of) Attacks

	Estimated Time to Acquisition			
		$>$ 4 years		**$\leq$ 4 years**
	total =	wartime +	peacetime	(all peacetime)
No. of Attacks	14	11	3	2
No. of SCoAs	18	12	6	18
Program-Years	186	28	158	33
Prob. of Attack	.0753	.393	**.0190**	**.0606**
Prob. of SCoA	.0968	.429	**.0380**	**.545**

Table 3: Correlation between Near-Success Estimates and (Serious Consideration of) Preventive Attack

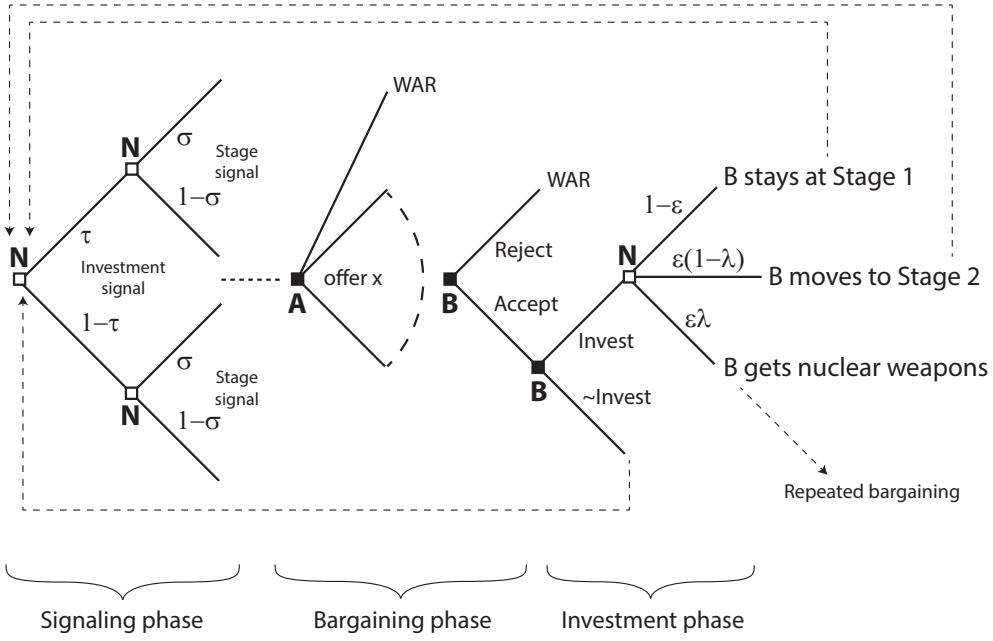

Figure 1: A single period of the game

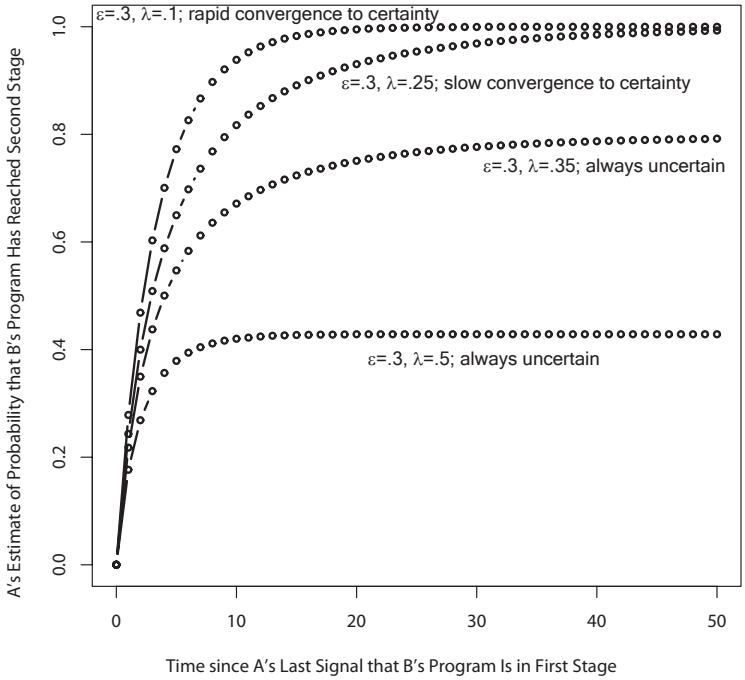

Figure 2: *A* grows more suspicious over time, but may never be certain

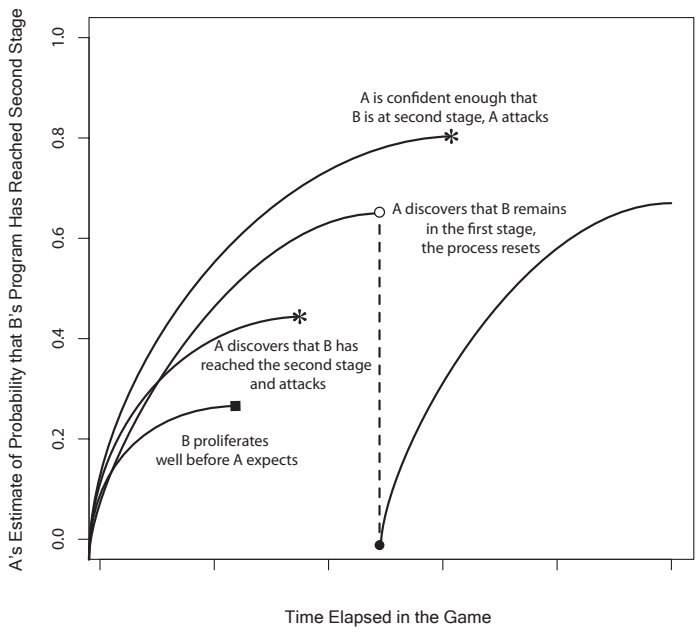

Figure 3: There are many possible paths through the equilibrium