Chapter 8

Preventive Strikes: When States Call the Wrong "Bluff"

On June 7, 1981, fourteen of Israel's best fighter pilots gathered at Etzion air base, near Israel's southernmost point. At 3:55 pm local time, the pilots entered their F-15 and F-16 fighter jets and took off. Their target: Osirak, a nuclear facility on the outskirts of Baghdad. Less than an hour later, the fighters had destroyed the facility. Israel had successfully executed one of the cleanest acts of preventive war in history.

This book's core theoretical model shows how a credible threat of preventive war heavily shifts bargaining power to the declining state. But thus far, the declining state has used that leverage to achieve better *peaceful* outcomes; the threat alone deterred the rising state. But Israel's 1981 attack, dubbed *Operation Opera* demonstrates that rising states do not always successfully internalize a declining state's preventive intentions. Why not?

Operation Opera was neither the first nor last preventive assault.[1] Norway's Vemork Hydroelectric Plant was one of the earliest facilities capable of producing heavy water. On the eve of Nazi Germany's invasion of Norway, French special forces smuggled the plant's entire supply out of the country. When Nazi officials ordered the Vemork plant to produce more heavy water, Allied forces and Norwegian resistance sabotaged the its machinery, dealing a significant blow to Germany's nascent nuclear program.

Similarly, during the early stages of the Iran-Iraq War, Iran executed Operation Scorch Sword. With French assistance, Iraq broke ground the

[1] See Fuhrmann and Kreps 2010 for an exhaustive list of attacks on nuclear facilities.

Osirak nuclear reactor. Osirak had not made much technological progress, but Tehran saw an opportunity to quash a potential nuclear weapons project in its infancy. However, the strike caused virtually no damage to the reactor.[2] Israel's Operation Opera finished the job less than nine months afterward.

Sixteen years later, Israel executed a similar assault known as Operation Orchard. This time, Syria was the target. Not much is known about the attack. The IAEA flagged the target site as a possible nuclear research facility, but Syria claimed it was an ordinary military depot. Either way, the Israeli Air Force leveled it.

While the previous chapter highlighted the role of imperfect information in explaining preventive war, the rising state's nuclear facilities were well-known to each of the declining states in these cases. How, then, can we explain such preventive strikes? The Norwegian heavy water sabotage is intuitive–bargaining had already broken down between the Allies and Nazi Germany, so we should expect miliary investments and the corresponding countermeasures.[3] Iran was similarly at war with Iraq at the time of Operation Scorch Sword. But Israel was not at war with Iraq or Syria prior to those operations. Why couldn't these states bargain their way out of conflict?

Indeed, this should be the expectation. With complete and perfect information, the rising state internalizes the declining state's threat. If that threat is credible, the rising state will not pursue an arms program, lest it needlessly pay a cost only to face preventive war. If not, the declining state offers butter-for-bombs concessions, preempting any need for those weapons and stealing the surplus in the process. The result is always efficient. Yet, in practice, states sometimes pursue weapons programs and sometimes their rivals intervene. Why?

So far, this book has not explored the role of incomplete information in the decision to proliferate.[4] Given past research on uncertainty and interstate coercion (Fearon 1995; Powell 1999, 97-104; Slantchev 2003; Slantchev 2010; Fey and Ramsey 2011), it should not be surprising that adding shifting power

[2]Iraq executed similar strikes on Iran's Bushehr reactor. That said, Bushehr was in its earliest stages, and Iran had abandoned the project at the start of the war (Sick 2001, 133).

[3]This is also why we have not discussed the United States' decision to proliferate in during World War II–butter-for-bombs negotiations are implausible in the middle of a war.

[4]An interaction features imperfect information if one or more actors are unaware of previous actions. In contrast, an interaction features incomplete information if one or more actors do not know the preferences of another actor.

to a model of incomplete information problems lead to inefficient fighting.[5] However, the literature currently ignores the possibility of butter-for-bombs agreements. Thus, those researchers can only analyze how shifting power and incomplete information creates *war*, not inefficient arms building.

This chapter address the discrepancy. In previous chapters, the rising state has always known whether the declining state would prevent if the rising state built. But as the crisis bargaining literature has demonstrated, state resolve is not always transparent and can trigger war. Consequently, we may wonder whether butter-for-bombs bargains remain possible when the rising state doubts the declining state's commitment to non-proliferation.

To analyze how incomplete information sabotages bargained settlements, this chapter explores a model in which the rising state is unsure whether the declining state is a "strong" or "weak" type. Strong types prefer engaging in preventive war to stop the rising state from proliferating. Weak types find preventive war too costly to engage in. The rising state does not know which type it is facing and must make inferences based on the declining state's actions.

The model has a number of compelling features. First, non-proliferation settlements are readily available for all types. Strong declining states can convince the rising state not to proliferate with mere threat of preventive war. Weaker declining states can bribe the rising state with butter-for-bombs offers to remove any incentive to proliferate. Thus, proliferation is not due to the lack of available settlements.

However, incomplete information builds the perfect storm for conflict. If the rising state cannot directly observe whether the declining state is strong or weak, weak declining states have incentive to bluff. That is, if the rising state were to believe that the declining state is strong upon receiving no concessions, then a weak declining state would want to mimic the strong type to convince the rising state not to build. After all, why give costly carrots when the stick is free to leverage? Weak declining states resolve this dilemma by sometimes proposing conciliatory butter-for-bombs settlements and sometimes bluffing as though they were strong. Strong declining states, meanwhile, always offer no concessions.

The bluffing behavior leads the rising state to rationally adopt a dangerous response plan. When the declining state presents butter-for-bombs

[5] Fearon, in his concluding remarks of "Rationalist Explanations for War," states that his causal mechanisms almost certainly work in conjunction with one another.

settlements, the rising state has no incentive to proliferate and therefore does not invest in nuclear weapons. But when the declining state offers no concessions, the rising state is unsure whether it is facing the weak type or strong type. Ultimately, the rising state resolves this dilemma by sometimes investing and sometimes conceding the issue.

Calling the bluff has explosive consequences. When the rising state attempts to proliferate, it does not know whether the declining state is strong or weak. This turns out well for the rising state if the declining state is weak. But strong declining states respond with preventive war. Thus, by sometimes calling the bluff, the rising state will occasionally regret it in retrospect.

The rising state's response has interesting welfare implications for the weak declining state. Intuitively, one might believe that the weak types would benefit from the uncertainty, as they could free-ride off of the strong type's credible threat. However, this incentive to bluff causes the rising state to act more aggressively to avoid exploitation. The end result is that the weak type ends up no better off than had it conceded the issue upfront and offered the necessary butter-for-bombs bribe. On the other hand, the incentive to bluff causes misery for the strong type and the rising state, who both must waste resources to avoid being manipulated.

Empirically, the equilibrium provides an intuitive explanation for preventive nuclear strikes as we see them. States can rationally leave nuclear facilities in the open if they believe their rivals are unlikely to attack them. But if the rivals know they have an easy preventive strike available, they execute those strikes.

The model also indicates this type of preventive war occurs more frequently as the cost to proliferate decreases, which we would expect over time. Although lower capital costs of nuclear weapons obviously means that more states would be willing to acquire a bomb in a vacuum, it also decreases the punishment for calling a perceived bluff wrong. Meanwhile, weaker declining states find engaging in butter-for-bombs deals to be less desirable, which makes them more inclined to bluff. Thus, if the cost of nuclear weapons decreases over time, we should expect more preventive wars to occur *ceteris paribus*.

This chapter has five additional sections. To begin, the next section formally defines the new extension to Chapter 3's baseline model and show the conditions under which the above strategies are optimal. The following two sections provide case studies of Israel's decision to intervene in Iraq and its ongoing strategy against Iran today. We then delve into the comparative statics, focusing on how equilibrium play changes as a function of the cost of nuclear weapons. A brief conclusion ends the chapter.

8.1 Modeling the Bluff

8.1.1 Actions and Information

For the same reasons as last chapter, this chapter focuses on a two-period version of the bargaining problem. In the first period, D offers $x \in [0, 1]$. R sees D's offer and accepts, rejects, or builds. Accepting ends the game and locks in the payoff pair $(x, 1-x)$ for the rest of time for R and D, respectively. Rejecting leads to pre-shift war. As usual, R prevails with probability p_R and D wins with complementary probability. R still pays c_R for fighting. D, in contrast, is either a *strong* type or a *weak* type. The strong type pays costs c_D while the weak type pays c'_D, where $c'_D > c_D$.

Lastly, R can build. If so, it pays a cost $k > 0$. D sees R's decision and chooses whether to prevent, which leads to the same pre-shift war payoffs as before. If D advances to the post-shift stage, it offers $y \in [0, 1]$ and R accepts or rejects that. Accepting locks in the payoff pair $(y, 1 - y)$ for the rest of time. Rejecting begins post-shift war. The payoffs are the same as in the pre-shift war, except R wins with probability $p'_R > p_R$ and D wins with probability $1 - p'_R$.

To incorporate uncertainty, Nature begins the interaction by selecting whether D is strong or weak. Specifically, Nature selects D as strong with probability q and weak with probability $1 - q$. The distribution is common knowledge but only D observes the realized draw. Thus, R must update its beliefs about D throughout the interaction.

8.1.2 Equilibrium

Since this is a sequential game of incomplete information, perfect Bayesian equilibria is the appropriate solution concept. As usual, we begin with the post-shift portion of the interaction:

Lemma 8.1. *Suppose the states reach the post-shift period. Then D offers $y = p'_R - c_R$ and R accepts. R's posterior is irrelevant.*

Proof: Let ψ be R's posterior that D is the defiant type. Consider how R responds to an offer y. With probability ψ, D is defiant, and R earns

$p'_R - c_R$ for the rest of time for rejecting. With complementary probability, D is weak, and R earns $p'_R - c_R$. These are the same payoffs, so R simply earns $p'_R - c_R$ if it rejects. Accepting generates y regardless of D's type. Thus, R accepts if $y \geq p'_R - c_R$ and rejects if $y < p'_R - c_R$.

Note that D's payoff if R accepts is identical regardless of its cost of war. As such, D's optimal acceptable offer equals $p'_R - c_R$. The defiant type prefers making this offer to inducing rejection if $1 - p'_R + c_R > 1 - p'_R - c_D$, which holds. Likewise, the weak type prefers making this offer to inducing rejection if $1 - p'_R + c_R > 1 - p'_R - c'_D$, which also holds. Consequently, both types of D offer $y = p'_R - c_R$ and R accepts. □

Moving forward, this chapter makes couple assumptions about c_D and c'_D to avoid trivial cases:

Assumption 8.1. *Let the strong type's cost of war be sufficiently small, or* $c_D < \delta p'_R - p_R - \delta c_R$.

Assumption 8.2. *Let the weak type's cost of war be sufficiently great, or* $c'_D > \delta(p'_R - p_R) - c_R$.

Combined, these two assumptions ensure that the strong and weak types have sufficiently differing preferences for preventive war. Assumption 8.1's value for c_D implies that the strong type has zero-tolerance for preventive war and will certainly intervene if R builds. Assumption 8.2 requires that the weak type's cost of war pushes it out of the "too hot" parameter space Proposition 3.1 covered.[6]

Thus, with complete information, vastly different outcomes would transpire. R and the weak type of D would negotiate over building weapons out of the shadow of preventive war; the strong type of D would coerce R not to build by credible threat of preventive war. Making these assumptions generates strategic tension in the interaction. R must decipher whether D is strong or weak and makes its investment plans accordingly.

Now to the propositions:

Proposition 8.1. *(Analogous to Proposition 3.2.) If* $p'_R - p_R < \frac{k(1-\delta)}{\delta}$*, both types of D pool in all PBE. They offer* $x = p_R - c_R$ *and R accepts.*

By now, the intuition should be well understood, and the proof follows directly from previous "too cold" scenarios. When $p'_R - p_R < \frac{k(1-\delta)}{\delta}$ R is too

[6] As usual, also maintain the assumption that $k \in (\frac{\delta(p'_R - p_R - c_D - c_R)}{1-\delta}, \frac{\delta p'_R - p_R}{1-\delta} + c_R)$ to rule out the substantively uninteresting cases described in Chapter 3.

impatient for the investment to be worthwhile. As a result, independent of D's cost of preventive war, R will not build. In turn, both types of D propose R's reservation value for pre-shift war. R accepts and the game ends.[7]

The strategic dilemma is far more complex outside of the "too cold" region. Here, strong types prefer to intervene if R invests. Internalizing this, R would not want to invest if it knew of D's resolve. Yet the weak type cannot leverage that same threat. Ordinarily, this would mean that the weak type would offer the necessary butter-for-bombs deal to reach an efficient resolution. But with incomplete information, the weak type might want to mimic the strong type's offer and trick R to not build because of the possibility of preventive war. It is therefore up to R to suss out whether D is weak or strong.

As is typical for this type of bargaining game, perfect Bayesian equilibrium's liberal off the path beliefs admit a large number of equilibria. Some of such equilibria require strange off-the-equilibrium-path inferences. For example, D might need to offer great concessions to R to convince R that it is strong, as R could believe (off the path) that D is weak if it makes any other offers, even if those offers are stingy. To avoid these situations and maintain tractability, this section searches for equilibria in which D offers $x = p_R - c_R$. Note that this is the amount that it would offer in the game with complete information. Thus, equilibria with that quality give the weak type incentive to pretend it is strong by mimicking what the strong type would normally do.

There are two compelling reasons to make this assumption. The first is empirical plausibility. As the Operation Opera case study will later show, Israel stood very firm in the lead up to the strike; it did not offer concessions in a bizarre effort to maintain a reputation for strength. As such, making this assumption ensures that the formal theory traces the empirical application as closely as possible. The second reason is practical. Narrowing the cases simplifies the proof and maintains tractability of the argument.

Given these conditions, R's decision to challenge in equilibrium depends on its prior belief that D is strong:

Proposition 8.2. *If $q > 1 - \frac{k(1-\delta)}{\delta(p'_R - p_R)}$, both types of D pool and offer $x = p_R - c_R$ and R accepts.*

[7]There are multiple PBE here because R could hold any number of beliefs about off the path offers. Regardless of those beliefs, however, it is never in R's interest to build because of the low ratio of return on investment.

For intuition, recall that q represents R's prior belief that D is strong. If R challenges the strong type, D fights and the investment cost goes to waste. Thus, if D is sufficiently likely to be strong, R accepts the minimal offer. Meanwhile, pooling with the strong type satisfies the weak type since it can withhold all concessions by doing so.

To see how to derive the critical value of q, note that R earns $p_R - c_R$ for accepting that offer size. In contrast, if both types of D pool on $p_R - c_R$ and R builds, R earns $p_R - c_R$ with probability q and $(1-\delta)(p_R - c_R) + \delta(p'_R - c_R)$ with probability $1-q$; either way, it pays the investment cost $(1-\delta)k$. Under these conditions, R would want to accept if:

$$p_R - c_R > q(p_R - c_R)(1-q)[(1-\delta)(p_R - c_R) + \delta(p'_R - c_R)] - (1-\delta)k$$

$$q > 1 - \frac{k(1-\delta)}{\delta(p'_R - p_R)}$$

This is the critical point seen in the proposition.

Given this, the weak type cannot profitably deviate. Offering any less than $p_R - c_R$ reveals that D is weak. R then builds, which leads to a payoff less than $1 - p_R + c_R$ for the weak type. The same is true if the weak type offers $x \in (p_R - c_R, p'_R - c_R - \frac{k(1-\delta)}{\delta})$. R accepts if $x \geq p'_R - c_R - \frac{k(1-\delta)}{\delta}$, but the weak type could induce acceptance at a lower price by pooling on the offer $p_R - c_R$.[8] □

Now consider cases where q falls below that critical point. These parameters lead to richer play and bluffing behaviors:

Proposition 8.3. *If* $q < 1 - \frac{k(1-\delta)}{\delta(p'_R - p_R)}$*, the strong type of D offers* $x = p_R - c_R$*; the weak type mixes between offering* $x = p_R - c_R$ *and* $x = p'_R - c_R - \frac{k(1-\delta)}{\delta}$.

[8] While this equilibrium is unique for the assumptions above, it is worth stressing that this is *not* the unique PBE. Indeed, the strong and weak types can pool on *any* demand size on the interval $[p_R - c_R, p_R + c_D]$ as part of a perfect Bayesian equilibrium. However, the PBE on the interval $(p_R - c_R, p_R + c_D]$ require that R believes D is likely to be weak if it offered a smaller amount at least as great as $p_R - c_R$ off the equilibrium path of play. In essence, R must take a surprisingly aggressive offer as a sign of weakness, not strength. (It also allows for wildly different payoffs for D for slight changes to the informational structure of the game. More specifically, recall that with complete information and D strong, D earns $1 - p_R + c_R$. But if we introduce incomplete information and allow D to be weak with some arbitrarily small probability ϵ, D's payoff could drop as low as $1 - p_R - c_D$ in a PBE.) The above assumption on off the path beliefs prevents R from making these types of inferences, and Proposition 8.2 is the unique PBE which survives it.

R accepts $p'_R - c_R - \frac{k(1-\delta)}{\delta}$ *with certainty and mixes between accepting and rejecting* $x = p_R - c_R$.

The appendix gives the exact mixing probabilities.

For intuition, note that the probability D is strong is low. Thus, if the weak type tried to pool with the strong type and offer $x = p_R - c_R$, R would believe that D is very likely to be bluffing, find gambling worthwhile, and therefore build. However, this leaves the weak type regretting it did not offer $x = p'_R - c_R - \frac{k(1-\delta)}{\delta}$, which buys off R and gives the weak type the surplus.

To resolve the dilemma, the weak type sometimes bluffs and sometimes reveals its weakness by proposing butter-for-bombs settlements. But to keep the weak type honest, R must sometimes build in response to receiving $p_R - c_R$. Afterward, the bluffing weak type admits its weakness and allows the power shift to transpire. The strong type, meanwhile, prefers preventive war to the disadvantageous peace in the future and therefore fights.

The bluffing behavior leads to inefficiency. With complete information, the strong type offers $x = p_R - c_R$ while the weak type offers $x = p'_R - c_R - \frac{k(1-\delta)}{\delta}$. In either case, R accepts. Both outcomes are efficient since R never builds and the states never fight a war. On the other hand, in the PBE Proposition 8.3 describes, both war and investment occurs with positive probability. The outcome is therefore inefficient in expectation.

Worse, the weak type's incentive to bluff does not ultimately produce any benefits for it. Although the weak type successfully tricks R into not building with positive probability, it mixes between offering $p_R - c_R$ and $p'_R - c_R - \frac{k(1-\delta)}{\delta}$. Standard indifference conditions then imply that the weak type receives $1 - p'_R + c_R + \frac{k(1-\delta)}{\delta}$ in expectation. Yet that amount is exactly how much it would receive if R knew D was weak from the start. In turn, the strong type and R pay the price for the inefficiency, as their preventive war and building behaviors are costly but necessary to stop the weak type's exploitation.

This chapter's appendix contains a full proof. As before, the equilibrium is not the unique PBE but its strategies are unique for the assumption about off the path beliefs.

Before moving on, a remark about the causal mechanism of war is in order. In the traditional sense, preventive war occurs when a declining state attacks a rising state to forestall an expected change in power over time. Preventive war in the model certainly fits this bill–the strong type interrupts the rising state's investment because it prefers the (slightly) inefficient war today to the efficient but disadvantageous peace of tomorrow.

However, preventive war traditionally understood (Fearon 1995) relies on a commitment problem as the casual mechanism. While preventive incentives motivate war here, credible commitment is not an issue. Consistent with Chadefaux 2011, if the rising state knew the declining state was strong, its desire to save the investment cost would induce it to not build. Instead, incomplete information–as imperfect information before it–drives the result; the rising state's uncertainty causes it to take an optimal risk-return gamble. If the declining state is sufficiently likely to be weak, this gamble induces the strong type to launch an attack.

8.2 Illustrating the Mechanism: Israel, Iraq, and Operation Opera

Studying preventive war in a nuclear context is difficult. As mentioned earlier, there are only four known instances of a military assaults intended to disrupt a nuclear facility.[9] Of these, two (Israel-Iran and Norway-Germany) came in the context of a greater conflict. This makes the quantitative analysis of nuclear prevention an onerous task, especially since strategic randomization injects noise into the data generating process.[10] In turn, studying the usefulness of the model requires comparing it to the historical record in relation to Israel-Iraq and Israel-Syria. However, Operation Orchard (the attack on Syria) occurred in 2007; consequently, much of the historical details remain unknown.[11] In turn, this section focuses on the 1981 Israeli attack on Iraq's Osirak reactor as a case study.

To trace the causal mechanism from the model, three features of the interaction are especially relevant: (1) Iraqi active development of nuclear technology was an open secret between Iraq and Israel (2) Iraq believed

[9]This does not count instances of conflict motivated partially by preventative concerns. Arab states may have instigated the Six-Day War to fight Israel before the latter had full access to nuclear weapons (Evron 1994; Karl 1996; Ginor and Remez 2006). Indeed, the Six-Day War is another possible application of this chapter's model but not discussed here.

[10]Beardsley and Asal 2013 provide a rare quantitative study and focus on how nuclear weapons program impact crisis initiation.

[11]The timing is particularly problematic because the model relies on incomplete information to explain preventive war. As such, to adequately assess the usefulness of the model, we must have an inside perspective on what the actors understood at the time of the conflict. Unfortunately, this type of historical record takes a long time to develop.

that Israel was not likely to intervene if it built the Osirak reactor, and (3) Israel calculated that a preventive strike was preferable to allowing Iraq to continue progressing on the reactor. The first point confirms that imperfect information was not the causal mechanism. The second verifies the role of incomplete information in the interaction. Finally, the third point follows Chapter 4's call to empirically evaluate a model's critical cutpoints.

Analyzing the strategic components of the attack first require an understanding of the players and the timeline of their actions. Thus, this section begins with a historical overview of Operation Opera, aimed at gathering the facts of the attack.[12]

8.2.1 Overview of the Iraqi Nuclear Program and Operation Opera

Iraq's involvement with nuclear technology began in the mid-1950s under King Faisal II as a part of American President Dwight D. Eisenhower's "Atoms for Peace" plan. However, the July 14 Revolution of 1958 saw a military coup overthrow the monarchy. Without American support or sufficient domestic nuclear know-how, the regime turned to the Soviet Union for assistance, ultimately signing an agreement on August 17, 1959. This led to the Soviet installation of a IRT-2000 reactor, which had a limited capacity of a few megawatts and was primarily a research facility (Evron 1994, 25-26).

The Iraq-Soviet Union pact eventually soured. The Ba'ath party rose to power in 1968. With Saddam Hussein in *de facto* control of the government, Baghdad sought a more powerful reactor. The Soviet Union, worried that this would lead to proliferation, rejected Iraqi advances (McCormack 1996, 47). Iraq turned to France, and the parties agreed to terms in 1974.[13] France then installed an Osiris-class reactor a ten miles outside of Baghdad. The site came to be known as *Osirak*, a portmanteau of Osiris and Iraq. France began construction of the project in 1979. It came online the next year, and Iraq started receiving uranium shipments in July of 1980. Unlike the Soviet IRT-2000 reactor, Osirak had a powerful 40 megawatt capacity and could process more than 150 pounds of uranium, enough for three atomic bombs

[12] See Perlmutter, Handel, and Bar-Joseph 2003 (xxix-l) for a more detailed timeline of the events from 1974 to 1981.

[13] Iraq also sought agreements with the United States, Canada, Brazil, Italy, and India (Cordesman 1999, 604). Only negotiations with Italy proved fruitful..

(Cordesman 1999, 605).

With that in mind, France instituted a number of safety protocols designed to ensure that Iraq could not use it to construct nuclear weapons. First, France hoped to alter the terms of its original agreement and send "caramel" fuel instead. Caramel contains lower concentrations of enriched uranium and is therefore ill-suited for nuclear weapons (Reiter 2005, 359) but still useful for nuclear research. Iraq bluntly rejected the switch. France caved, worried that Iraq would cancel its lucrative contracts and go elsewhere (Nakdimon 1987, 103).

Aware of Iraq's potentially duplicitous intentions, France only resupplied Iraq with uranium after receiving the spent uranium from the previous shipments. This meant if Iraq were to cheat on the deal, its nuclear stockpile would be limited. Moreover, cheating without detection would have proven difficult. Iraq deposited its accession to the NPT on October 29, 1969 and thus the IAEA had oversight all declared nuclear activity within Iraq.[14]

Nevertheless, the possibility that Osirak would be a stepping stone to a nuclear weapon greatly worried Iraq's enemies. By 1980, Iraq and Iran were embroiled in a bloody conflict. Just nine days into fighting, the Iranian air force executed *Operation Scorch Sword*, an air raid attempting to destroy the facility. The raid dealt minor structural damage and failed to raze the reactor. Iran did not attempt further action.

Meanwhile, despite the safeguards, Osirak also concerned Israel. The key players within Prime Minister Menachem Begin's cabinet agreed that a nuclear-equipped Iraq was a threat to Israel but differed on their short-term solutions. Hawks urged an attack as a preventative measure. Doves believed that a strike could spawn trigger an anti-Israeli alliance of Arab countries (Perlmutter, Handel, and Bar-Joseph 2003, xl). Moreover, the Osirak reactor was in its infancy, and thus Israel had no need to rush to a decision.

After much debate, the Israeli cabinet approved counter-proliferation measures in October 1980. The strike occurred eight months later, on June 7, 1981. Six F-15As and eight F-16As departed from Etzion Airbase, located toward the southern tip of Israel. The fighters flew through Jordan and Saudi Arabia before heading to Osirak. Once there, the F-16s dropped their payloads and scored numerous hits on the reactor's dome, completely demolishing it. Iraqi resistance proved ineffective; anti-aircraft fire failed to damage any of the planes, and all the pilots returned to Israel safely. Eleven

[14] Iraq officially signed the NPT on July 1, 1968, two weeks before the Ba'ath takeover.

died on the ground–one French technician and ten Iraqi soldiers.

International condemnation was universal. The United Nations Security Council unanimously passed Resolution 487 twelve days later, without veto from the United States. The resolution noted Iraqi compliance to IAEA inspections and its right to peaceful nuclear technology, criticized Israel for not adhering to the NPT, and called for restitution to Iraq. Israel offered no such concessions but compensated the lone French casualty. Ultimately, Israel suffered no real sanctions from the attack. On the contrary, Washington expressed gratitude by the time the Persian Gulf War broke out in 1991, as Saddam Hussein could not deter the United States with a nuclear weapon.

Indeed, Iraq's nuclear program was never the same. Iraq initially vowed to rebuild the site. However, as Proposition 3.1 illustrated, rising states cannot effectively build if their rivals have credible preventive war threats. Israel had credibly demonstrated its willingness and ability to destroy Osirak-like reactors. Consequently, the Iraq nuclear program went underground so as to deny Israel an effective preventive war option. However, as detailed in the last chapter, these methods were ultimately unproductive. By the fall of Saddam Hussein in 2003, Iraq still had no nuclear weapons.

Much of Iraq's nuclear woes came from the international fallout from the attack. Although UNSC Resolution 487 supported Iraq on paper, Saddam found himself without a viable technological partner. France's initial response was to rebuild the facility, but Paris withdrew entirely shortly thereafter. Saudi Arabia offered billions in support, but that money was useless without a seller.

While this overview establishes the players and their actions, it lacks a complete understanding of the parties' motivations and their strategic dilemmas. Even though three decades have passed since Operation Opera, details from within the Iraqi regime remains rudimentary. Still, there is little doubt the Iraqi nuclear program saw technological advances as a prerequisite to a nuclear bomb and not as an end in itself. Claims to the contrary–that Osirak was for research and that true nuclear weapons ambitions–only came after Operation Opera and trace back to an Iraqi nuclear scientist (Cooley 2005, 159). Meanwhile, the consensus in Washington, Tehran, and the IAEA at the time believed that Iraq sought a nuclear arsenal. Conventional wisdom today concurs. Moreover, the types of reactors Iraq sought to purchase from France are consistent with a nuclear weapon being the end game, not nuclear energy (Kirschenbaum 2010, 51; McCormack 1996, 47-49).

Unfortunately, with Saddam Hussein's execution in 2004, we may never

Timeline of the Osirak Reactor and Operation Opera

7/14/1958	Iraqi monarchy overthrown
8/17/1959	Soviet nuclear assistance to Iraq begins
7/17/1968	Ba'ath party comes to power
10/29/1969	Iraq deposits accession to NPT
1969	First reactors come online
8/1974	Iraq turns to France for nuclear assistance
11/18/1975	France agrees to build Osirak
4/9/1979	Original Osirak reactors destroyed in French hanger
7/1980	Uranium shipments to Ira1 begin
9/22/1980	Iran-Iraq War begins
9/30/1980	Iran attacks Osirak, fails to destroy the reactor
10/28/1980	Israeli cabinet votes in favor of attack
6/7/1981	Operation Opera
6/19/1981	UNSC Resolution 487 unanimously condemns Israel
8/2/1990	Iraq invades Kuwait
2/24/1991	Coalition forces invade Iraq

have a complete portrait of Iraq. As a result, the following subsections offer one possible explanation for why Israel and Iraq were unable to come to terms on the eve of Operation Opera, despite the existence of mutually preferable settlements.

8.2.2 The Osirak Reactor: An Open Secret

To separate the role of incomplete information versus the previous chapter's focus on imperfect information, this section must first establish that the presence of the Osirak reactor was an open secret. This is the easiest step to verify. Iraq's nuclear weapons program was a matter of public record at the time. The NPT protected Iraq's right to nuclear technology at the time, and the Osirak facility met all of the IAEA's inspections. While the IAEA did not have complete access to the site, it still reported on Iraq's nuclear technological progress.

Moreover, Iraq believed that Israel was deeply opposed to the nuclear facility. Perlmutter, Handel, and Bar-Joseph (2003, 53-63) detail a number of covert operations aiming to derail Iraq's progression toward a bomb, which we recount here. France scheduled shipment of the original reactors to begin on April 9, 1979. But two days beforehand, there was a mysterious explosion at the hanger which destroyed both reactors. While the French media pinned responsibility on a domestic ecological terrorist group, French intelligence believed Mossad was responsible.[15]

The next operation took place on June 13, 1980. Iraq commissioned Yahya El Meshad, an Egyptian nuclear scientist, to inspect French equipment for delivery to Osirak. French police found him in his hotel room the next day, a victim of a brutally violent murder. Debate over the responsible party persists. Mossad again was the initial presumption, though French intelligence believed the reckless execution suggested otherwise. The alternative theory is that Syrian agents, working on behalf of Russia, had broken into El Meshad's room to obtain documents on the progress of Osirak. These agents bribed a prostitute to keep him occupied, but El Meshad ignored her advances.[16] He then entered the room mid-operation, at which point the agent panicked and killed him.

The final operations occurred two months later, on August 7, 1980. Iraq had coopted Italy to provide chemical separation technology vital to the enrichment process. But on that evening in August, three bombs exploded simultaneously in Italy, two at an Italian company providing the sensitive assistance and one at its general manager's apartment. As usual, Mossad was the likely culprit.

Around this time, the French-Iraqi relationship became a major news story in the Israeli media. Israel's official position was unsurprising: a nuclear Iraq posed an unacceptable threat to Israel's security, and that Israel was prepared to deploy the appropriate countermeasures. Thus, by this point, Iraq understood Israel's knowledge of the program.

What is missing, however, is a clear indication of Israel's resolve. All Iraq had observed at this point was cheap talk to the Israeli press and low-risk, low-cost covert anti-proliferation operations. Whether Israel would actually intervene militarily remained in doubt. And as the next section illustrates, Iraq had good reason to be optimistic.

[15]Given the lax security around the hanger, some suspect French intelligence aided the assailants. According to the theory, elements within France's government did not want to give Saddam Hussein nuclear technology and saw the sabotage as a convenient half-measure that would not harm French-Iraqi relations.

[16]Police questioned the prostitute, Marie-Claude Magal, on July 1. She mysteriously died in a hit-and-run eleven days later.

8.2.3 Iraqi Optimism

The next task is to show that Iraq could have plausibly believed that the construction of Osirak would not result in an Israeli strike. Indeed, despite Israel's efforts to stall Osirak's development, Iraq had many reasons to be optimistic that a full-scale military assault was inconsequential for practical reasons or unlikely for political reasons. In turn, Iraq's decision to push forward on the reactor appears rational.

To start, the Iranian attack a year earlier proved fruitless. Damage to the reactors was superficial, thus raising Iraq's belief that further attacks would be more nuisance than destruction. Moreover, the assault alerted Iraq that such aerial bombardments were possible. To deter further Iranian attacks on the site, Iraq invested $1.7 billion in anti-aircraft defense in the following months (Perlmutter, Handel, and Bar-Joseph 2003, 65). Israel would need superior aim and evasion skills for a military operation to be viable.

Second, Israel faced severe tactical constraints. Unlike Iraq and Iran, Iraq and Israel do not share a border. Thus, any successful attack would have required Israeli jets to travel around 2000 miles round trip and fly over Syrian, Jordanian, or Saudi airspace, increasing the likelihood that Iraq would have an early warning.[17] The distance and foreign airspace created additional problems. Israeli fighter jets could not make the round trip without refueling– but tanker aircrafts needed to refuel over a foreign country, and thus leave them vulnerable to air defenses.[18] Also, if anything went wrong, Israeli pilots would have to set down in an unfriendly location.

Third, Israel had to worry about international backlash. After all, France assisted Iraq in constructing Osirak, and many French workers supervised the facility. Any attack on Osirak would surely result in a French backlash. Thus, preventive war against Iraq was also at least partially a conflict against France.

Moreover, Israel had signed the Camp David Accords with Egypt only two years earlier. An attack on another Arab state risked prompting Egypt to withdraw from the Accords. Perhaps worse, the United States induced Israel to sign the Accords in exchange for a generous stream of aid (Arena

[17]Ultimately, Israel chose the Saudi route. Reports indicate that King Hussein of Jordan saw the planes fly past him shortly after takeoff (Lewin 2007; Nakdimon 1987, 21-22). Communication problems prevented him from contacting Iraq in time.

[18]This continues to be an issue for Israel today with a potential strike on Iran (Raas and Long 2007, 23-27).

and Pechenkina 2012). The U.S. intended its military aid to Israel to be for defensive purposes only (Muller, Castillo, and Morgen 2007, 214). Preventive war would not qualify. The United States therefore had the means and contractual forewarning to punish Israel should it disapprove of a preventive strike.

As it turned out, Iraqi optimism was appreciated within the Israeli cabinet. Only a slight majority consented to what would become Operation Opera, and the opposition was vocal. For reasons we outline in the next section, Prime Minister Menachem Begin ultimately followed the majority. If bargaining broke down here as a result of information problems, we should expect precisely such a close vote within Israel–after all, close votes should be more difficult for foreign entities to anticipate than blowouts.

8.2.4 Preventive War versus Disadvantageous Peace

The remaining question is whether Israel preferred bearing the costs of preventive war to an efficient but disadvantageous peace. A power shift–especially a nuclear power shift–would have left Israel in a particular vulnerable position. Israel is a small country with few large cities. Just a few Iraqi nuclear weapons have the potential to functionally destroy the country (Muller, Castillo, and Morgen 2007, 213). Although the Israel could respond in kind with its own nuclear weapons, the possibility alone would have drastically shifted bargaining power to Iraq.

Moreover, the prevailing wisdom within the cabinet believed that Saddam was belligerent enough to press his luck in the shadow of mutually assured destruction. To that end, in 1978, he announced on public radio that his regime's goal was "an all-out military struggle, aimed at uprooting Zionism from the area" (Perlmutter, Handel, and Bar-Joseph 2003, 39). While the practicality of such a declaration is questionable, this nevertheless concerned Israel. Further exacerbating the problem, Israel grew convinced that Osirak would eventually yield a bomb regardless of IAEA safeguards and French assurances Kirschenbaum (2010, 53).

Dissenters in the cabinet argued that Israel should hold back and wait in the hope that Iraq would yield to other foreign political pressure. But it was unclear how waiting would have benefitted Israel. At best, Israel could hope that Franois Mitterrand, elected President of France on May 10, 1981, would withdraw nuclear assistance. But two weeks later, despite sympathy for Israel's plight, the Mitterrand administration announced it would honor its preexisting commitments to Iraq (Nakdimon 1987, 202-203).

Failing that, Israel understood how distant its policy positions were from Iraq. No third country was willing to step in and buy off Iraq. Iran, the other major rival in the region, was already at war with Iraq. The United States, traditionally the leading anti-proliferation advocate, was only two years removed from the Iranian hostage crisis and sought a better relationship with Saddam Hussein. All other states lacked a large enough stake in the outcome to provide substantial bargaining leverage.

This left Israel as the only state left to initiate a butter-for-bombs deal. But Israel stood firm. And as this chapter's model demonstrates, Iraq could not directly differentiate whether this was because Israel planned a preventive strike or because Israel was bluffing.[19] Waiting would not resolve Iraq's unfounded optimism, as the lack of a butter-for-bombs offer is consistent with both weakness and strength.

Meanwhile, the waiting game only gave Iraq more time to conduct nuclear research. Thus, Israel struck at an opportune moment. According to a former Mossad agent, an Iraqi contact had placed a beacon within the Osirak a few days before the attack, which allowed IAF pilots to score two direct hits to the reactor (Westerby 1998, 25-32). Delay would have given Iraqi counterintelligence more time to uncover the plot, remove the beacon, and reduce the effectiveness of the assault. So Israel pushed forward with Operation Opera and saw a resounding success.

In doing so, Israel banked on a subdued response from the international community. The cabinet was correct–backlash against Operation Opera proved inconsequential. Twelve days after the attack, the United Nations Security Council unanimously approved Resolution 487, which condemned the operation. But this would only amount to a slap on the wrist; although the United States failed to wield its veto power on 487, it promised to do so on any subsequent resolutions to sanction Israel. Washington did punish Israel on its own for using American-supplied aircraft on the mission (Muller, Castillo, and Morgen 2007, 216), but the two-month suspension on sales amounted to a slap on the wrist.

Meanwhile, Israel sent a powerful message to the international community. Paris knew Saddam planned to eventually obtain a nuclear weapon but provided atomic assistance anyway to secure access to Iraq's vast oil re-

[19]While not implicitly included in the model, it should be apparent that Israeli incentives to misrepresent prohibit credible communication from avoiding conflict.

serves. Primier Jacques Chirac successfully quelled opposition from within his own administration at the start of the project (Perlmutter, Handel, and Bar-Joseph 2003, 51). Israel's bold attack bolstered the critics' position. France subsequently withdrew its assistance following Opera.

Militarily, Iraq produced no response. This was foreseeable, as the Iran-Iraq War was ongoing at the time; Iraq could ill afford to initiate a two-front fight. Thus, in pure military costs, Operation Opera only consumed jet fuel and the bombs the planes dropped on the reactor. And as the model shows, intervention is likely when the declining state's true cost of war is low.

Not to be stopped, Iraq multiplied its nuclear technology budget–with Saddam immediately investing $1 billion into new techniques (Albright and Kelley 1995, 56)–and went underground. This is consistent with the information-based mechanism described here. Iraq initially choose to a cheaper method to acquire nuclear know-how under the assumption that Israel would be unlikely to attack. After Operation Opera, Iraq updated its beliefs about Israeli willingness to intervene and strategically altered its path to proliferation.

But the further investment proved futile. Throwing money at the problem could not provide a solution given that indigenous nuclear know-how was insufficient to produce a bomb (Kreps and Fuhrmann 2011, 171-172; Hymans 2012, 84-93). This forced Iraq to look for partners. But Israel believed Operation Opera was worthwhile in part to signal other countries to stay away (Perlmutter, Handel, and Bar-Joseph 2003, 79). The effort succeeded–France pulled out, and no other nuclear power was willing to replace France's technological know-how. Iraq instead had to rely on the A.Q. Khan nuclear black market, and much of the extra spending went toward hiding the program from the eyes of the international community. A decade later, at the start of the Persian Gulf War, Iraq still lacked a bomb despite no further direct intervention on its nuclear program.

Note that the causal logic at work here is not whether Operation Opera was a long-term success. Decades later, this remains a point of debate. Reiter (2005), for example, argues the strike only pushed the Iraqi program underground and contributed to the Iraq War in 2003. While this debate is important for policy prescription, the task here was to show that Israel believed that preventive war was better than the alternatives for the purpose of explaining the causal mechanism. To that end, tracing the Israeli cabinet's decision-making process is sufficient proof.

8.3 Lessons Learned: Israel and Iran Today

Sixteen years after Operation Opera, Israel undertook a similar incursion in Syria. Dubbed Operation Orchard, Israeli fighter jets again played target practice with alleged nuclear facility in Deir ez-Zor. Not much time has passed since, and thus the historical knowledge necessary to check whether the same causal mechanism precipitated the conflict does not yet exist.

Regardless, Israel's successes with Operation Opera and Operation Scorch Sword has delivered an important signal to Israeli antagonists: Israel has the capability and willingness to destroy unprotected nuclear facilities. Rational states ought to learn from the past, understand that their dyadic relationship falls in the "too hot" range from Chapter 3's model, internalize the risk of preventive war, and opt not waste the investment cost of failed proliferation.

That leaves Iran today. Why does Iran continue proliferating despite Israel's tough reputation?

Today's Israel/Iran dilemma is distinct from the previous Israel/Iraq and Israel/Syria incidences in two ways. Indeed, cognizant of Israel's successful forays, Iran has developed critical countermeasures to discourage an assault. First, many key Iranian facilities are underground. Osirak sat on the surface, vulnerable to direct bombings. In contrast, Israeli needs bunker-busting technology to do any serious damage to Iran's program from the air. Israel's alternatives are to bomb the entrances or to stage a ground invasion. The former would merely act as a stalling tactic; the latter would come at significant cost.

Second, Iran's program is highly decentralized. Israel needed to bomb but a single target to massively setback Iraq's program. This will not work against Iran, since Iran can ramp up production at a second facility to make up for lost productivity elsewhere. Again, the goal here is to render an Osirak-like strike ineffective. Israel would instead have to send multiple fleets on bombing runs to deliver a significant setback or fight a more general (and conventional) war against Iran. Either way, Israel risks substantially more casualties.

This places Israel and Iran back into a world with incomplete information and thus uncertainty over the outcome. Publicly, Israel has remained tough, not offering any concessions to Iran. But this behavior occurs when Israel is tough and also when (due to the Iranian countermeasures) the enfeebled Israel bluffs. And there is reason to believe Israel is bluffing–Wikileaks released an Israeli 2005 report confirming that a preventive strike was impossible (Melman 2011).

However, given the incentives to misrepresent, Iran can only discover Israel's true resolve by continuing to work on nuclear weapons. Meanwhile, Israel appears content to sabre rattle in the hope that Iran will back down on its own or let the United States bear the cost of concessions. Whether Israel will actually attack is Israel's private knowledge.

8.4 The Future of Bluffing

Finally, consider the comparative statics of proliferation costs. As time progresses, technology improves and the cost of proliferation correspondingly declines. Recalling Figure 3.5, we should expect fewer states to fall into the "too cold" range and more to actively seek nuclear weapons. In turn, the prevalence of preventive war ought to increase, as more status quo states must decide whether to buy off potential rising states or posture appropriately.

Nevertheless, the model also shows that preventive war is more likely with lower proliferation costs for a subtler reason. Recall that when q (the prior belief that the declining state is strong) is sufficiently high, the declining state extracts the entire surplus from the rising state. The declining state's bargaining leverage comes from two sources. First, a higher probability that the declining state is strong implies that the rising state is less likely to be successful should it attempt proliferation. And second, a lower probability that the declining state is weak means that the investment cost goes to waste more frequently.

But when q is sufficiently low, the rising state is more adventurous, and both states ultimately mix. Note that "sufficiently low" here means that q is less than $1 - \frac{k(1-\delta)}{\delta(p'_R - p_R)}$. As nuclear weapons become increasingly costless, or as k goes to 0, the cutpoint converges at 0, meaning that q is assuredly sufficiently low.[20] This is because the rising state is less fearful of a wasted investment as building becomes increasingly costless. In essence, the rising state becomes more willing to challenge the declining state when the consequences of calling the wrong bluff become irrelevant.

Moreover, the weak type has further incentive to bluff as k decreases. To induce the rising state to accept, it must offer $p'_R - c_R - \frac{k(1-\delta)}{\delta}$, leaving $1 - p'_R + c_R + \frac{k(1-\delta)}{\delta}$ for itself. As k goes to 0, the weak type's share converges

[20] It must be that q is not equal to 1, as that is the complete information case.

to $1-p'_R+c_R$. In the complete information model from Chapter 3, this caused the declining state to steal everything upfront and suffer the consequences of proliferation later on. Here, this incentivizes the weak type to bluff more frequently.

In turn, to combat the declining state's inclination to bluff, the rising state challenges the declining state more frequently upon receiving the stringent $x = p_R - c_R$ offer. To see this, recall that the rising state builds in response to $x = p_R - c_R$ with probability $1 - \frac{(1-\delta)(p_R - p'_R + \frac{k}{\delta})}{\delta(p'_R - p_R)}$. Note that the derivative of this with respect to k equals $\frac{-(1-\delta)}{\delta^2(p'_R - p_R)}$, which is strictly negative. So, as k increases, the probability the rising state challenge decreases; as k decreases, the probability the rising state challenges increases.

Of course, time can effect other factors, which in turn mitigate the negative consequences of cheaper nuclear technology. The spread of continue democracy would lead democratic peace theorists to expect fewer wars anyway. Similarly, increased trade could increase the opportunity cost of war (Gartzke 2007). This would invalidate Assumption 8.1, which requires a strong state's cost of war fall below a critical threshold. If so, otherwise strong states would be compelled to bargain with potential risers and avoid the risks of war altogether. The general obsolescence of war (Mueller 1989) would provide similar incentives.[21]

8.5 Conclusion

This chapter showed how incomplete information can interfere with butter-for-bombs bargaining. If the rising state is unsure whether the declining state is weak or strong, weak declining states have incentive to mimic the strong type's behavior to convince the rising state not to proliferate. However, this causes the rising state to sometimes call the declining state's bluff and invest in nuclear weapons. When the declining state is truly strong, it launches preventive war; when it is weak, it permits the rising state to proliferate unimpeded. The result is inefficient because of the unnecessary proliferation and unnecessary war.

The model contributes to our understanding of how incomplete information promotes or inhibits proliferation behaviors. Referring back to the conventional wisdom from Chapter 1, rising states acquire nuclear weapons if

[21] See Braumoeller 2013 for criticism.

the declining state will not prevent and the gains from bargaining outweigh the investment cost. Incomplete information would then seem to decrease the likelihood of proliferation, since rising states might opt not to proliferate if they are unsure whether a declining state will intervene. However, because butter-for-bombs agreements work, incomplete information ultimately promotes proliferation by sabotaging the bargaining process.

Substantively, the model and case study of Operation Opera provide insight on the preventive strike on Syria and Isarel's ongoing drama with Iran. Only a few years have passed since Operation Orchard. Reliable diplomatic histories are not yet available. Still, the same mechanism gives one plausible explanation–Syria underestimated Israel's tactical advantage, leading to the construction (and ultimately destruction) of the facility. The model also suggests that current tensions with Iran could also be the result of private information.

8.6 Appendix

This appendix proves Proposition 8.3. However, for showing uniqueness of equilibrium strategies given the off the path assumptions, the following lemma will prove invaluable:

Lemma 8.2. *No PBE exist in which D offers* $x > p'_R - c_R - \frac{k(1-\delta)}{\delta}$.

Proof: The proof follows similarly to a portion of the proof for Proposition 3.3. For $x' > p'_R - c_R - \frac{k(1-\delta)}{\delta}$, both types of D can profitably deviate to the midpoint between x' and $p'_R - c_R - \frac{k(1-\delta)}{\delta}$.

Consider R's best response to x'. If R accepts, it earns x', which is greater than its pre-shift war payoff $p_R - c_R$. (Proposition 8.1 covered the situation where this is not the case.) If R builds, then D does not prevent in R's best case scenario. R earns $(1-\delta)x' + \delta(p'_R - c_R) - (1-\delta)k$ for this outcome. This is worse than accepting outright if:

$$x' > (1-\delta)x' + \delta(p'_R - c_R) - (1-\delta)k$$

$$x' > p'_R - c_R - \frac{k(1-\delta)}{\delta}$$

So R accepts any such x'.

Now consider D's payoff for offering any such x'. R accepts and D earns the remainder, or $1 - x'$. But D could profitably deviate to offering $x = \frac{x' + p'_R - c_R - \frac{k(1-\delta)}{\delta}}{2}$. This amount is still strictly greater than $p'_R - c_R - \frac{k(1-\delta)}{\delta}$, so R accepts. But because $\frac{x' + p'_R - c_R - \frac{k(1-\delta)}{\delta}}{2} < p'_R - c_R - \frac{k(1-\delta)}{\delta}$, D keeps strictly more for itself. Thus, offering x' cannot be a part of any PBE. □

Now for the full proof for Proposition 8.3. To begin, consider the full equilibrium strategies and beliefs:

- **Strategies**
  - Strong type of D offers $x = p_R - c_R$
  - Weak type of D offers $x = p_R - c_R$ with probability $\frac{q}{(1 - \frac{k(1-\delta)}{\delta(p'_R - p_R)})(1-q)} - \frac{q}{1-q}$ and offers $x = p'_R - c_R - \frac{k(1-\delta)}{\delta}$ with probability $1 - \frac{q}{(1 - \frac{k(1-\delta)}{\delta(p'_R - p_R)})(1-q)} + \frac{q}{1-q}$
  - Depending on off the path beliefs, R rejects or builds given $x \in [0, p_R - c_R)$
  - R accepts given $x = p_R - c_R$ with probability $\frac{(1-\delta)(p_R - p'_R - \frac{k}{\delta})}{\delta(p'_R - p_R)}$ and builds with probability $1 - \frac{(1-\delta)(p_R - p'_R - \frac{k}{\delta})}{\delta(p'_R - p_R)}$
  - R builds given $x \in (p_R - c_R, p'_R - c_R - \frac{k(1-\delta)}{\delta})$
  - R accepts given $x \in [p'_R - c_R - \frac{k(1-\delta)}{\delta}, 1]$
  - Strong type of D prevents if R builds if $c_D < x_t(1-\delta) + p_R - \delta p'_R + \delta c_R$ and advances otherwise
  - Weak type of D prevents if R builds if $c'_D < x_t(1-\delta) + p_R - \delta p'_R + \delta c_R$ and advances otherwise
- **Beliefs**
  - Given $x \in [0, p_R - c_R)$, R is free to have any belief (off the equilibrium path)
  - Given $x = p_R - c_R$, R believes D is strong with probability $\frac{q}{q + \frac{q}{(1 - \frac{k(1-\delta)}{\delta(p'_R - p_R)})}}$ and is weak with probability $1 - \frac{q}{q + \frac{q}{(1 - \frac{k(1-\delta)}{\delta(p'_R - p_R)})}}$ (via Bayes' rule)

- Given $x \in (p_R - c_R, p'_R - c_R - \frac{k(1-\delta)}{\delta}]$, R believes D is weak with probability 1
- Given $x \in (p'_R - c_R - \frac{k(1-\delta)}{\delta}, 1]$, R is free to have any belief (off the equilibrium path)
- In the second period, R believes D is weak with probability 1

Bayes' rule trivially validates the on the equilibrium path beliefs, so this appendix focuses on proving the rationality of the strategies given the equilibrium beliefs.

First, note that R and the weak type of D mix. Thus, they must be indifferent between their pure strategies at each of these information sets. Let σ_A be R's probability of accepting given that D offers $x = p_R - c_R$. The following derives the weak type's indifference condition:

$$1 - p'_R + c_R + \frac{k(1-\delta)}{\delta} = \sigma_A(1 - p_R + c_R) + (1 - \sigma_A)[(1-\delta)(1 - p_R + c_R) + \delta(1 - p'_R + c_R)]$$

$$\sigma_A = \frac{(1-\delta)(p_R - p'_R + \frac{k}{\delta})}{\delta(p'_R - p_R)}$$

This is R's equilibrium mixed strategy. Note that the dominator is strictly positive. For σ_A to be valid, it must also be that $p_R - p'_R + \frac{k}{\delta}$ so that the numerator is as well. This implies $k > \delta(p'_R - p_R)$, which the assumptions verify.[22]

Likewise, R must be indifferent between accepting and building. Let s be the probability D prevents in response to R building. The following derives R's indifference condition:

$$p_R - c_R = s(p_R - c_R) + (1 - s)[(1-\delta)(p_R - c_R) + \delta(p'_R - c_R)] - (1-\delta)k$$

$$s = 1 - \frac{k(1-\delta)}{\delta(p'_R - p_R)}$$

Note s is not D's strategy but rather a posterior belief that D will prevent. Let σ_B be the probability the weak type bluffs by offering $x = p_R - c_R$. Since

[22] As Proposition 3.3 showed, k must be greater than this to ensure that D prefers making butter-for-bombs offers to taking the entire good upfront.

the strong type always offers $p_R - c_R$ and is the only type to prevent if R builds in response, Bayes' rule gives the value of σ_B which makes R indifferent between accepting and building:

$$s = \frac{q}{q + (1-q)\sigma_B}$$

From before, $s = 1 - \frac{k(1-\delta)}{\delta(p'_R - p_R)}$. Thus, the following substitution gives the solution for σ_B:

$$1 - \frac{k(1-\delta)}{\delta(p'_R - p_R)} = \frac{q}{q + (1-q)\sigma_B}$$

$$\sigma_B = \frac{q}{(1 - \frac{k(1-\delta)}{\delta(p'_R - p_R)})(1-q)} - \frac{q}{1-q}$$

This is D's equilibrium mixed strategy.

Now check for profitable deviations. To begin, consider R's response to receiving $x = p'_R - c_R - \frac{k(1-\delta)}{\delta}$. R believes D is weak with probability 1. If R builds, it earns $(1-\delta)(p'_R - c_R - \frac{k(1-\delta)}{\delta}) + \delta(p'_R - c_R) - (1-\delta)k = p'_R - c_R - \frac{k(1-\delta)}{\delta}$, which is not a profitable deviation. For rejecting to be optimal, it must be that $p'_R - p_R < \frac{k(1-\delta)}{\delta}$, which Proposition 8.1 prohibits. Thus, accepting is optimal.

Next, suppose R receives $x = p_R - c_R$ instead. The indifference condition ensures R earns the same amount for accepting as it earns for building. Note that R also earns $p_R - c_R$ for rejecting as well, so rejecting is not a profitable deviation. In turn, R's equilibrium strategy here is optimal.

Now consider D's strategies. The optimality of the strong and weak type's decision to prevent or advance follows directly from Proposition 3.1. Thus, check whether either type of D can profitably deviate by changing its offer size.

First, consider the weak type of D. In equilibrium, its payoff equals $1 + p'_R + c_R + \frac{k(1-\delta)}{\delta}$.[23] If the weak type offers $x \in [0, p_R - c_R)$, R can have any beliefs in equilibrium. Regardless, accepting cannot be optimal for R, since R could instead reject and earn $p_R - c_R$ instead. If rejecting is optimal, the weak type earns its pre-shift war payoff, which is worse than the butter-for-bombs

[23] This is because it is indifferent between offering $x = p_R - c_R$ and $p'_R - c_R - \frac{k(1-\delta)}{\delta}$. In the latter case, R accepts, thereby granting the weak type a payoff of $1 + p'_R + c_R + \frac{k(1-\delta)}{\delta}$, the equilibrium amount.

offer. If building is optimal, then the proof for Proposition 3.3 shows that D is again better off with the butter-for-bombs offer because $k > \delta(p'_R - c_R)$. So deviating to $x \in [0, p_R - c_R)$ is not profitable.

Next, consider deviations to $x \in (p_R - c_R, p'_R - c_R - \frac{k(1-\delta)}{\delta})$. According to equilibrium beliefs, R believes that D is weak with probability 1. Since R must receive at least $p'_R - c_R - \frac{k(1-\delta)}{\delta}$ to not find investment profitable, it builds. D does not prevent because $k > \frac{\delta(p'_R - p_R + c'_D + c_R)}{(1-\delta)}$. So D earns $(1-\delta)(1-x) + \delta(1 - p'_R + c_R)$. This is not a profitable deviation if:

$$1 - p'_R + c_R + \frac{k(1-\delta)}{\delta} > (1-\delta)(1-x) + \delta(1 - p'_R + c_R)$$

Since $1 - \delta + \delta(1 - p'_R + c_R) > (1-\delta)(1-x) + \delta(1 - p'_R + c_R)$, we may instead prove this by showing:

$$1 - p'_R + c_R + \frac{k(1-\delta)}{\delta} > 1 - \delta + \delta(1 - p'_R + c_R)$$
$$k > \delta(p'_R - c_R)$$

As before, this holds.

Lastly, the weak type could deviate to $x > p'_R - c_R - \frac{k(1-\delta)}{\delta}$. However, this is clearly not optimal–R accepts $x \geq p'_R - c_R - \frac{k(1-\delta)}{\delta}$, so giving any more is an unnecessary concession. Therefore, the weak type has no profitable deviations.

Now consider deviations from the strong type. In equilibrium, the strong type receives a convex combination of $1 - p_R - c_D$ and some amount greater than that. Thus, the strong type earns some amount strictly greater than its war payoff.

First, suppose the strong type offers $x \in [0, p_R - c_R)$. R is free to have any beliefs. Accepting remains suboptimal for R since it could reject and earn $p_R - c_R$ instead. So R either builds or rejects. If R builds, then Assumption 8.1 ensures that the strong type will prevent in response. Thus, regardless of R's decision, D earns its war payoff. But this is strictly worse for the strong type than earning a convex combination of its war payoff and some strictly greater amount. Therefore, the strong type cannot profitably deviate to offering $x \in [0, p_R - c_R)$.

Second, suppose the strong type offers $x \in (p_R - c_R, p'_R - c_R - \frac{k(1-\delta)}{\delta})$. Then R believes D is weak with probability 1 and therefore builds. The strong type then prevents and earns its war payoff. But again, this is strictly worse than offering $p_R - c_R$ and earning a convex combination of its war payoff and some strictly greater amount. Consequently, the strong type cannot profitably deviate to offering $x \in (p_R - c_R, p'_R - c_R - \frac{k(1-\delta)}{\delta}]$.

Finally, suppose the strong type offers $x \in [p'_R - c_R - \frac{k(1-\delta)}{\delta}, 1]$. R believes D is weak with probability 1 if $x = p'_R - c_R - \frac{k(1-\delta)}{\delta}$ and is free to have any belief but accepts regardless. However, the strong type earns strictly less than its war payoff and therefore fares worse than had it offered $x = p_R - c_R$. As such, the strong type has no profitable deviations.

Showing that these are the unique equilibrium strategies given that the strong type offers $x = p_R - c_R$ follows from the above. It is sufficient to show that the weak type cannot offer any values other than $p_R - c_R$ and $p'_R - c_R - \frac{k(1-\delta)}{\delta}$. Lemma 8.2 gives that the weak type must receive $1 - p'_R + c_R + \frac{k(1-\delta)}{\delta}$ in equilibrium. In any other equilibrium, R correctly identifies that D is weak. So if the weak type offers $x \in [0, p_R - c_R)$ or $x \in (p_R - c_R, p'_R - c_R - \frac{k(1-\delta)}{\delta})$, R builds. But the weak type could profitably deviate to $x = p'_R - c_R - \frac{k(1-\delta)}{\delta}$. From Lemma 8.2, the same is true if the weak type offered $x > p'_R - c_R - \frac{k(1-\delta)}{\delta}$. This leaves $p_R - c_R$ and $p'_R - c_R - \frac{k(1-\delta)}{\delta}$ as the only options, as covered in the proof. □