Bargaining over Costly Signals

William Spaniel[*]

July 26, 2021

Abstract

Uncertainty over resolve is a central explanation for war, and costly signaling has become a textbook solution the problem. However, costly signals are an imperfect alternative—they resolve inefficiencies by creating inefficiencies, and I show that the equilibrium signaling inefficiency is often *worse* than the equilibrium war inefficiency if signaling were impossible. Would states prefer to negotiate a fully efficient resolution instead? I develop a model to answer this question. In equilibrium, the uninformed actor often "buys out" the informed party's costly signaling option. This can occur even if the signal would be perfectly informative and come at no direct cost to the uninformed actor. Moreover, where armaments are greatest in traditional costly signaling models, they are the *smallest* when allowing for preemptive bargains.

---

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

1 Introduction

A well-recognized cause of war is uncertainty over an actor's resolve (Fearon 1995). When one actor does not know how much the other cares about the issue at stake, it is liable to demand too much in negotiations, thinking that the other side will back down. Given that uncertainty causes war here, revelation of private information is the solution. Costly signaling has received the greatest attention on this front—sunk costs overcome credibility issues, as only resolved types would sacrifice something of value to obtain a better negotiated outcome (Fearon 1997; Slantchev 2005, 2011; Arena 2013; Tarar 2013; Wolford 2014). In fact, it has become a textbook solution to the problem of asymmetric information (Frieden, Lake and Schultz 2013, 103–104).

Parallel to work on crisis bargaining, researchers have developed a similar inefficiency puzzle regarding armament and mobilizations (Kydd 2000; Rider 2013; Debs and Monteiro 2014; Spaniel 2019). Deploying soldiers to hot zones, building new aircraft, or developing nuclear weapons all leave less funds for productive domestic investment. If states can anticipate what the expected outcome of negotiations would be post-investment, absent any other bargaining frictions, they should reach similar agreements *before* those developments. Arming or mobilizing at that point is unproductive, as the state already has already obtained what it expects to receive by doing so but must still pay the associated cost.

From this perspective, costly signaling is a second-best solution. What makes war tragic is that it is inefficient. If all information were revealed, the parties would agree on a settlement that both prefer to fighting. Costly signaling can reveal that information. However, the solution to one inefficiency is an inefficiency of a different sort—and, as I show later, the inefficiency from costly signals often *exceed* the expected equilibrium inefficiency from war. Thus, a natural question is whether a first-best, wholly efficient solution exists to the asymmetric information dilemma.

To answer that, I develop a model that features preemptive bargaining, before a costly signal takes place. The uninformed side does not know its opponent's underlying resolve, which determines that state's willingness to arm and fight a war. If the parties fail to reach a preemptive bargain, then the play a classic costly signaling game similar to canonical models (Arena 2013; Tarar 2013; Wolford 2014). War occurs if the parties fail to reach an agreement at that point.

In equilibrium, the parties always reach an agreement with positive probability during the preemptive bargaining stage. For many parameter spaces, the parties reach a deal with certainty. This may be surprising given that the uninformed actor could sabotage negotiations and allow costly signaling to fully reveal information at no direct cost to itself. However, the missing intuition is that any costly signal sent represents lost surplus that the uninformed actor could have captured through negotiations. In turn, the uninformed actor indirectly pays the full price of a signal. Consistent with existing models of intelligence (Arena and Wolford 2012), the extra information may not be worth paying for. Breakdown only occurs here when the uninformed state is relatively sure that its opponent is unresolved. Here, paying a premium to induce resolved type to also accept is not worthwhile in expectation. From there, costly signaling helps sort types as usual.

Beyond providing microfoundations for preemptive negotiations, the model generates three other major findings. First, the armament literature's prediction bears out. Holding fixed a set of parameters, the equilibrium of the preemptive bargaining game is strictly more efficient than the equilibrium of the standard costly signaling model. From an armament perspective, the benefits are substantial. For example, under the belief that *maximizes* armaments in a standard costly signaling model, the equilibrium armaments are *zero* in the preemptive bargaining model.

Second, again holding fixed a set of parameters, the equilibrium probability of war in the preemptive bargaining game is weakly lower than the equilibrium of the costly signaling model. More directly, there are circumstances where war would occur under the costly signaling model but where adding preemptive bargaining guarantees the peace. The reason is that preemptive bargaining adds a hurdle to the path of war. When optimal costly signals cannot fully separate types, an unresolved type has an incentive to arm beyond its natural preference to bluff the behavior of a resolved type and obtain a better post-signal deal. Uninformed actors hedge against that possibility by sometimes calling what they know might be a bluff. However, to reach this stage with preemptive bargaining, the uninformed actor must first decide not to "buy out" the unresolved type's bluffing option. Because bluffs and called bluffs generate great inefficiency, the uninformed actor often ensures that the unresolved type accepts the first proposal. This paves the way for the resolved type to separate itself through the signal, allowing for an agreement to arise afterward.

Finally, equilibrium armaments under the preemptive bargaining game have greater variance than in costly signaling game. The proposer always makes an offer that guarantees at least some probability of acceptance—anything else needlessly wastes surplus that it could otherwise siphon off. As a result, an unresolved type's optimal armament quantity is never produced. Instead, the game produces extremes. Either lots of arms are produced—from a resolved state or an unresolved state bluffing—or none are.

Overall, my work qualifies the utility of costly signaling. None of the claims here devalue the theoretical contributions of existing costly signaling models. The ability to signal in my model still transmits information and reduces the probability of war, which is the central theoretical prediction of existing work. Furthermore, much of the work on costly signaling illustrates mechanisms orthogonal to the point I wish to highlight.[1] Rather, the key takeaway is that arming and mobilization hide a deeper strategic problem. Such actions shift the inefficiency puzzle up a step, as the broader literature on arms construction has identified. Surprisingly, the seemingly straightforward explanation—that costly signals ultimately save inefficiency—is often insufficient.

2 Motivating Preemptive Bargaining

As the introduction previewed, the model I develop below adds preemptive negotiations before costly signaling occurs. I pause for a moment here to substantively motivate how this matches empirical crisis negotiations.

U.S.-Iraqi movements following the invasion of Kuwait provide an excellent example. Saddam Hussein's forces entered Kuwait on August 2, 1990 and captured the country within two days. Washington condemned the action. Indeed, on the day of the invasion, the United States pushed through United Nations Security Council Resolution 660. It demanded "that Iraq withdraw immediately and unconditionally all of its forces to the positions in which they were located" on the day before the invasion.

The next five months featured a series of tacit negotiations and public proposals to resolve the crisis. Within a week, Saddam proposed a retreat from Kuwait in exchange

---

[1]As examples, Arena (2013) shows that costly signals do not effectively communicate information about martial effectiveness, Tarar (2013) demonstrates that the uninformed state has a preventive war incentive prior to the signal, and Wolford (2014) examines a tradeoff between greater mobilizations and support from a third party. My model abstracts away from each of those points to clarify my specific mechanism (Paine and Tyson 2020).

for Israeli concessions in Arab and Palestinian areas (Hussein 1990). That same month, National Security Advisor Brent Scowcroft received a more serious offer where Iraq would pull out if it could maintain control of the Rumailah oil field (Royce 1990). By the end of the year, Iraq had moved away from the idea that involving Israeli-Palestinian issues was necessary for movement on Kuwait (Tyler 1991). The United States kept a firmer stance. November's Security Council Resolution 678 still called for a withdrawal and authorized the use of force.

What is notable about this is that the tacit negotiations occurred before the United States had mustered a full signal to Iraq.[2] The American deployment came slowly, with George HW Bush not deciding to deploy offensive capability into the region until November 9. The decision reflected his geopolitical reality. By that time, Bush had become frustrated that Saddam was not taking the underlying threat of war seriously (Stewart 2010, 17). Pulling troops from Europe and increasing the total deployment to 400,000. This came with costs that Bush had hoped to avoid. Reflecting on the decision, Bush wrote that "the troop increase, particularly its size, whipped up a new outcry ... [t]he pundits and congressmen on the morning talk shows and the op-eds averred that I was wrecking my presidency" (Bush and Scowcroft 1998, 396).

Of course, Saddam remained skeptical even after observing the mobilization. Iraq maintained its position in Kuwait and refused to offer further concessions. Saddam bet that American casualty aversion, combined with Iraq's ability to mine the Persian Gulf and destroy oil wells (Woods, Palkki and Stout 2011, 180–182, 194–195), would be enough to hold back the coalition. But Bush was serious, and the war began in February.

The Persian Gulf War is not an isolated incident of this phenomenon. In fact, the lead up to the Iraq War mirrors it. The United States began making demands of Iraq well before the March 2003 war. George W Bush's "axis of evil" speech occurred in January 2002. The Bush administration's threats to Iraq continued throughout the year. Yet they held off on the major troop mobilization to Kuwait until February 2003. This was intentional. During the design process of the invasion, Secretary of Defense Donald Rumsfeld kept pushing for a faster and faster break out time (Dale 2008, 10–13).

---

[2]The first American troops reached Saudi Arabia almost immediately after the invasion. However, Washington's goal here was not to signal U.S. willingness to repel Iraq from Kuwait. Instead, there was concern that Saddam's foothold in Kuwait created a first-strike incentive to attack Saudi Arabia. These troops were designed to be a "speed bump" to counter that incentive (Stewart 2010, 3–8).

By doing so, the U.S. could hold off on suffering any deployment costs until diplomacy fell through. Like before, those negotiations were unsuccessful. The deployment did not fully convince Saddam of the credibility of the threat either, leading to the initiation of war.

Other times, failed initial bargains lead to costly signals that then facilitate a negotiated resolution. The East-West dispute over Cuba is a well-known example. Following Fidel Castro's declaration that Cuba was a socialist republic, Cuban independence became a foreign policy objective for the Soviet Union. However, a tacit understanding between the superpowers was not immediately forthcoming. Instead, the Cuban Missile Crisis had to instead first demonstrate each side's commitment to the issue, beginning with the Soviet nuclear deployment and followed by the Kennedy's administrations preparations for war. But this exchanged paved the way to an agreement that kept the parties from engaging in a real war.

Costly signals also seemed to help resolve the standoff between India and Pakistan in 2001. Tensions rose after the bombing of the Indian Parliament on December 13. India blamed Pakistani-based terrorist groups and mobilized troops to the Line of Control in Kashmir. This decision came with a hefty price tag: £1.1 billion (Slantchev 2011, 230). Recognizing the gravity of the threat, Pakistani President Pervez Musharraf condemned the attack and promised to crack down on Kashmir-focused militant groups within his country. India subsequently canceled its plan to strike Pakistani targets.

Yet other times, preemptive agreements obviate the need of arms programs to signal willingness to fight. After all, "[m]ost leaders would rather prevent a crisis from emerging than pay the costs and take the risks of trying to deter an adversary once the crisis has occurred" (Huth and Russett 1993, 61), with India's £1.1 billion as a prime example. Ukraine's diplomatic saga following independence shows how there can be a way out. As a new country with fresh leadership, the setting for uncertainty and conflict was high in theory (Wolford 2007). In practice, the Ukraine's executive and legislative branches put forth competing platforms on the issue (Garnett 1995, 137–140). Ukraine also had the technical capacity to build nuclear weapons and the historical rivalry with Russia to motivate great defense expenditures.[3] Faced with this uncertainty, the United States and Russia instead bought out Ukraine's option, providing great economic relief

---

[3]Moscow still had command and control over the leftover Soviet missiles on Ukrainian territory (Miller 1993), so this is a comment about a native nuclear program.

and security assurances (Reiss 1995, 100–122).

These examples suggest a need for an additional layer to standard costly signaling models. Rather than limit negotiations to after informed states have communicated information, there are appear to be two phases to crises. First, there is an opportunity to reach a fully efficient agreement, before a state has committed to mobilizing for a crisis or arming to generally deter its opponent. Second, there is a post-armament opportunity to at least stave off the costs of war. I now turn to working toward a model that incorporates both phases.

3   The Costly Signaling Game

Before addressing negotiations over costly signals, I first define a costly signaling game. This is useful for a few reasons. First, it establishes a benchmark to draw comparisons later. In fact, the game I develop here is identical to Arena (2013) with a continuous signaling space, Tarar (2013) without a preventive war decision and with a linear mobilization cost function, and Wolford (2014) without a third party. In other words, this is a classic costly signaling crisis bargaining game. Second, I demonstrate that costly signaling may lead to less efficient outcomes than an analogous game without costly signals. Finally, it forms a subgame of the overall I game I will later analyze, and so it is a necessary step regardless.

The costly signaling game consists of two players, A and B, with some good in dispute. Nature starts by drawing B as an "unresolved" type with probability $q \in [0, 1]$ and as a "resolved" type with probability $1 - q$. B alone observes its type and chooses a military investment $m \geq 0$. After observing the investment, A issues a proposal $x \in [0, 1]$. In the final move, B accepts or rejects.

Payoffs are as follows. I standardize A's value of the good to 1. Meanwhile, B's valuation depends on its type. The unresolved type has a valuation $V > 0$, whereas the resolved type has valuation $V' > V$. If B accepts x, each enjoys its share of the deal. This is x_1 for A. B must also pay for its sunk cost mobilization, leaving $(1 - x_1)V - m$ for the unresolved type and $(1 - x_1)V' - m$ for the resolved type.

War payoffs are more complicated. Let $p(m) \mapsto [0, 1]$ be a twice continuously differentiable function that maps B's investment to A's probability of victory. I assume $\frac{\partial p}{\partial m} < 0$ and $\frac{\partial^2 p}{\partial m^2} > 0$; that is, A's probability of victory decreases (and thus B's increases)

6

in the investment and that investments have diminishing marginal returns. In addition, war costs the respective players $c_i > 0$. All told, A earns $p(m) - c_A$, the unresolved type earns $(1 - p(m))V - c_B - m$, and the resolved type earns $(1 - p(m))V' - m - c_B$.

To facilitate efficiency analysis, and standard to models of crisis bargaining, I exploit the fact that expected utilities are identical across positive affine transformations. As a result, I can scale each type's payoffs by the inverse of its valuation. For example, I can equivalently define the unresolved type's payoffs as $1 - x - \frac{m}{V}$ for a deal and $1 - p(m) - \frac{c_B + m}{V}$ for war. The resolved type's scaled payoffs are analogous. This makes the prize worth 1 to every player and type, so that I can calculate inefficiency as the percentage of that quantity lost to costs.

3.1 The Costly Signaling Equilibrium

Because this is a sequential game with incomplete information, perfect Bayesian equilibrium (PBE) is the appropriate solution concept. Broadly, one of two things can occur in equilibrium. First, the types may wish to separate by arming to different levels. To see how this is possible, consider A's proposal decision. It faces a classic risk-return tradeoff. For any given investment, it could go conservative and offer the resolved type's reservation value to B, choosing to extract a quantity of surplus based on the higher valuation V'. The resolved type is just willing to accept, and so the unresolved type strictly prefers the deal. A's alternative is to go aggressive and offer the unresolved type's reservation value. This increases A's payoff conditional on acceptance but potentially risks war against the resolved type. The optimal strategy under the tradeoff therefore depends on its posterior belief about B's type.

However, the more important inference here is that the resolved type has nothing to hide. It receives an amount equivalent to its war payoff regardless. Therefore, in its investment decision, the resolved type only needs to maximize its tradeoff between power and armament costs. With an objective function of $1 - p(m) - \frac{c_B + m}{V}$, the appendix shows that the resolved type selects the unique solution to:

$$-\frac{\partial p}{\partial m} = \frac{1}{V'} \tag{1}$$

assuming that $-\frac{\partial}{\partial m} p(0) > \frac{1}{V'}$. This condition simply means that the return on the first

bit of investment is positive.[4] Call the optimal amount m^{**}.

The unresolved type's signaling dilemma becomes the following. If it mimics the resolved type's behavior, A may pay concessions commensurate with the resolved type's war payoff. Nevertheless, it must still pay m^{**} in armament costs. This is more than it would ordinarily want to—it is less resolved, and so it internalizes those costs at a higher rate. The alternative is to select some other amount. However, A can infer that the resolved type will select m^{**}. Thus, on any equilibrium path other than m^{**}, it infers that B is unresolved. As such, if the unresolved type does not want to choose m^{**}, it might as well maximize the armament tradeoff from its perspective. Generalizing from Line 1, it should produce the unique solution to:

$$-\frac{\partial p}{\partial m} = \frac{1}{V} \tag{2}$$

conditional on $-\frac{\partial}{\partial m}p(0) > \frac{1}{V}$ being true. I assume this condition holds throughout to avoid corner solutions. Call the optimal amount for the unresolved type m^*. Note further that $m^{**} > m^*$—that is, the optimal production for the resolved type is larger than the optimal production for the unresolved type.[5]

Taking stock, even if A would believe the unresolved type's bluff, it still prefers m^* if:

$$(1 - p(m^*))V - c_B - m^* > \left(1 - p(m^{**}) - \frac{c_B}{V'}\right)V - m^{**}$$

$$\frac{m^{**} - m^*}{V} > \left(p(m^*) + \frac{c_B}{V}\right) - \left(p(m^{**}) + \frac{c_B}{V'}\right) \tag{3}$$

This yields the first claim:

Lemma 1. *Suppose Line 3 holds. The unresolved type selects m^* and the resolved type selects m^{**} in all PBE.[6] With types fully revealed, A makes concessions commensurate with the respective types' war payoffs.*

---

[4]If $-\frac{\partial}{\partial m}p(0) < \frac{1}{V'}$, B chooses $m = 0$.

[5]To see this formally, let $\frac{\partial p}{\partial m} + \frac{1}{V} = 0$ be the implicit function that generates the optimal m. The implicit function theorem says that the derivative of the optimal m with respect to V is $-\frac{\frac{\partial}{\partial V}\left(\frac{\partial p}{\partial m} + \frac{1}{V}\right)}{\frac{\partial}{\partial m}\left(\frac{\partial p}{\partial m} + \frac{1}{V}\right)} = \frac{1}{V^2 \frac{\partial^2 p}{\partial m^2}}$. This is positive because $\frac{\partial^2 p}{\partial m^2} > 0$.

[6]The game has generically unique equilibrium strategies, but equilibrium can sustain multiple off-the-path beliefs.

This is the case where costly signaling "works" as intended. Neither type wants to mimic the other's behavior, and so A learns everything it needs to know from observing the armament decision. With the information problem alleviated, the parties reach an agreement.

The interaction becomes more complicated when Line 3's incentive compatibility condition fails. The appendix covers the details. If A's belief that B is unresolved is low, the unresolved type can pool on m^{**} with the resolved type. Despite knowing of the possible bluff, A still makes a large concession in response. Calling the potential bluff backfires too often to be worthwhile.

In contrast, if A believes B is likely unresolved, pooling on m^{**} backfires—A would see the heavy investment as a likely bluff and call it. But the unresolved type also should not reveal itself by separating at m^*. In response, A would pay off the resolved type after observing m^{**}, thereby inducing the unresolved type to want to bluff. Consequently, the unresolved type semi-separates. Sometimes, it chooses m^*, thereby revealing itself. The remaining time it bluffs on m^{**}. Now skeptical, A responds to m^{**} by occasionally calling the possible bluff. The unresolved type concedes at this point, but the resolved type finds the proposal unacceptable. Letting r represent A's posterior belief B is unresolved, the appendix calculates this threshold as:

$$q > \frac{c_A + \frac{c_B}{V'}}{c_A + \frac{c_B}{V}} \tag{4}$$

Summarizing:

Lemma 2. *Suppose Line 3 fails. If Line 4 also fails, both types select m^{**}, and A makes concessions commensurate with the resolved type's war payoff in all PBE. Peace prevails.*

Lemma 3. *Suppose Line 3 fails. If Line 4 holds, the resolved type selects m^{**} and the unresolved type mixes between m^* and m^{**} in all PBE. After observing m^*, A makes concessions commensurate with the unresolved type's war payoff. Peace prevails. After observing m^{**}, A mixes between concessions commensurate with the unresolved type's and resolved type's war payoffs. In the former case, war occurs with the resolved type. Otherwise, peace prevails.*

Overall, it appears that costly signaling works here. The only pathway to war occurs through Lemma 2's circumstances. Even there, they are narrow. To start, the

incentive compatibility constraint must fail so that the types do not separate. Next, the uninformed state must be wrong in its assessment about its opponent's type. Finally, the uninformed state must choose to call what could be a bluff. Otherwise, the parties agree to a settlement.

3.2 Do Costly Signals Enhance Efficiency?

War is tragic because it is inefficient. Costly signaling is the textbook prescription when uncertainty would otherwise cause war, and the previous lemmata seem to validate the approach. But is the cure worse than the disease? This question is not currently discussed in the literature. I investigate it here. The answer is surprising—equilibrium costly signals are often more inefficient, and sometimes substantially so.

Calculating the net efficiency gains or losses first requires a basis of comparison. Fortunately, the model provides two obvious starting points. First is the identical setup without a costly signaling decision. The second is the same as the original setup but examining the case where $-\frac{\partial}{\partial m}p(0) < \frac{1}{V'}$. This implies that the resolved type's marginal return on investment for the initial military allotment is negative. Because the unresolved type's is also negative under these circumstances, the types endogenously choose not to pursue costly signaling. Thus, the comparison in this latter case is between situations where the cost and military technology are permissive to costly signaling and situations where they are not permissive.

Regardless of the framing point, the equilibrium is the same. When no signals arise—either exogenously or endogenously—Line 4 is all that matters. If it holds, A makes the risky proposal, leading the resolved type to reject. The overall probability of war is therefore $1 - q$. If it fails, A proposes the safe quantity and ensures the peace.

To stack the deck in favor of costly signaling being helpful, consider the outcome of the full costly signaling model under Lemma 1's parameters. Line 3's incentive compatibility constraints hold here, so the signals are fully effective and no war occurs. Even under these rosy circumstances, costly signaling does not always deliver on its efficiency-enhancing promise:

Remark 1. *Suppose Line 3 holds. If Line 4 fails, the inefficiency is strictly greater in the costly signaling game than that of the game without costly signaling. If Line 4 holds, there exists a critical value of q strictly bound below 1 such that the expected inefficiency*

in the costly signaling game is greater than that of the game without costly signaling for any q greater than that critical value.

The problem when Line 4 fails is straightforward. When A suspects B is resolved, it takes a cautious bargaining approach. The natural probability of war equals 0. As such, the realized inefficiency from war is also 0. Adding costly signaling to the mix creates inefficiencies without solving any others. In turn, the inefficiency is always greater in the costly signaling game. Indeed, the total additional inefficiency equals the total value lost to arms production, which equals $\frac{qm^*}{V} + \frac{(1-q)m^{**}}{V'}$.

One might wonder whether this is an unfair comparison. Costly signaling is a prescription to avoid war, and no war would have occurred in the baseline case. However, there are two problems with that perspective. First, states do not selectively choose these types of costly signals when war might otherwise occur. Instead, the motivation is return on investment: higher armaments give states greater coercive power and more concessions at the bargaining table.

Second, the case where war would otherwise occur does not solve the problem. This is the second half of Remark 1. When A suspects B is unresolved, it naturally takes an aggressive approach. Without signals, the equilibrium demand forces the resolved type to fight a war. With the incentive compatibility condition satisfied, however, the costly signals separate types and guarantee the peace. As such, costly signaling halts the inefficiency of war. That is, in the absence of signaling, war would have occurred with probability $1 - q$, and the sides would have collectively suffered $c_A + \frac{c_B}{V'}$.

Unfortunately, this calculation overlooks the inefficiency from signals. Like in the previous case, the unresolved type contributes $\frac{m^*}{V}$ and the resolved type contributes $\frac{m^{**}}{V'}$ in losses. Consequently, the costly signals create more waste than they save if:

$$\frac{qm^*}{V} + \frac{(1-q)m^{**}}{V'} > (1-q)\left(c_A + \frac{c_B}{V'}\right) \qquad (5)$$

Because the right hand side goes to 0 as q goes to 1 and the left hand side remains strictly positive, this *must* hold for a sufficiently large q. As q increases, A becomes increasingly confident that B is unresolved and will accept the smaller concession. The corresponding probability of war decreases. In turn, the inefficiency from fighting becomes near-irrelevant. However, the unresolved type still produces its arms regardless. Consequently, the armament loss eventually exceeds the saved inefficiency from war.
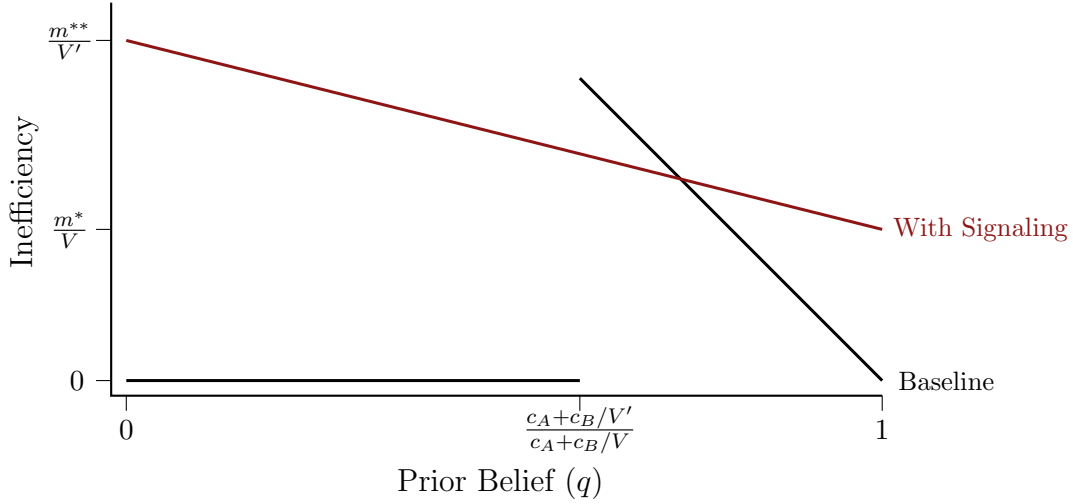
11

Figure 1: Expected equilibrium inefficiencies with and without costly signals.

This is true even if the costly signals are small—e.g., a minor mobilization of resources to a particular theater.

Figure 1 illustrates these results. It plots the inefficiency under the baseline model and with signaling as a function of q. The left half shows that signaling has a massive increase in inefficiency for low priors. On the right side, the baseline briefly shows a higher level of inefficiency. But the signaling model eventually returns as the less efficient setup.

The figure hides a couple of more subtle possibilities. First, it is possible that the loss from costly signaling from is universally larger than the gains made from greater peace. All this requires is that Line 5 holds at the cutpoint on q from Line 4. Minor mobilizations are insufficient for this to hold. Nevertheless, theoretical perspectives and specific historical cases indicate that this condition holds some of the time (Fearon 2018; Koblentz 2018; Coe and Vaynman 2020). For example, from the start of the atomic age to 1996, the United States spent $5.3 *trillion* on deploying nuclear weapons in constant 2020 dollars (Schwartz 2011, 4), despite only using them twice. It is also conceivable that the armament costs exceed the sum theoretical inefficiency of war—i.e., Line 5 holds with the right side converted to $c_A + \frac{c_B}{V'}$.

Second, in the figure, the inefficiency produced attributable to the unresolved type is smaller than the inefficiency produced attributable to the resolved type. However,

it is possible that the inefficiency strictly increases in belief instead. This can arise because, even though the unresolved type chooses a smaller allotment, its per-unit loss is larger.

Remark 1 focused on what happens when incentive compatibility works. Unfortunately, costly signaling does not fare any better when incentive compatibility fails. Because the equilibrium strategies are more complicated, I save this discussion for the appendix. However, the two main points from Remark 1 still hold. First, when q is low, peace prevails in the baseline model. But in the costly signaling model, both types pool on the higher military allocation. In turn, signals remain more inefficient. Second, when q is high, a semi-separating equilibrium prevails. Thus, not only does signaling create inefficiency, it also does not fully solve the war problem. Furthermore, as q approaches 1, the probability of war still goes to 0 in the baseline. But in the costly signaling model, the unresolved type produces some mixture of the high and low allotments. Consequently, a sufficiently high prior belief that B is unresolved yields a greater inefficiency from the signal than would occur due to war.

4   Bargaining over Signals

The central takeaway so far is that costly signals are not the panacea that conventional wisdom suggests they might be. In the best case scenario, they are an imperfect solution, resolving inefficiency by creating a different inefficiency. In the worst case scenario, the inefficiency from signaling is actually worse than the inefficiency that would have occurred in war otherwise.

Of course, in this type of investment game, B also wishes to arm because doing so allows it to extract more in bargaining. But this is still not a satisfying explanation for the signaling inefficiency. States can bargain over armaments as well (Kydd 2000; Rider 2013; Spaniel 2019) and should reach an agreement without any bargaining frictions. In turn, the natural next question is whether a first-best solution exists. The answer is not obvious because uncertainty over resolve means A does not know how B internalizes both the costs of war and the price of arms.

To work toward a solution, a model must give the actors an opportunity to engage in a preemptive bargain. To allow for that, I add a move to the original model. After Nature draws B's type, A now demands $x_1 \in [0, 1]$. B accepts or rejects. If it accepts,

the game ends with the proposed division implemented. If it rejects, the actors play the costly signaling game. The only change here is that I now refer to the post-signal demand as x_2. This makes the overall structure similar to Kydd's (2000) model, but A's uncertainty extends to both how B internalizes the price of armaments and the cost of war. As such, B's signaling decision has downstream consequences for the second round of bargaining.

4.1 Complete Information Equilibrium

I begin with a quick description of the complete information equilibrium to establish a benchmark. From earlier, the parties reach an agreement in the final negotiation phase. Recognizing that the deal becomes more attractive as it invests more, B builds arms until its marginal cost equals its marginal benefit. For an actor with valuation V, that optimal quantity is m^*.

Turning to the preemptive negotiation, A recognizes that B earns $1 - p(m^*) - \frac{c_B + m^*}{V}$ by rejecting the initial proposal. Buying off B at that amount leaves $p(m^*) + \frac{c_B + m^*}{V}$ for A. Failing to reach a deal immediately means that A will eventually come to an agreement at $p(m^*) + \frac{c_B}{V}$. Waiting is strictly worse because it fails to extract the $\frac{m^*}{V}$ value. Consequently, the parties reach an immediate agreement without armaments. This captures the central intuition from Spaniel (2019). Like war, arms are inefficient. As such, settlements exist that leave both sides better off than had armament occurred. Here, A captures that surplus as the proposer of the demand.

In sum, the complete information model is fully efficient. As such, if any inefficiencies arise with incomplete information—through war or arms—the variation in information environments is responsible. But if there are fully efficient equilibria with incomplete information, the previous analysis of the costly signaling game shows that the availability of a preemptive bargain causes the change.

4.2 Incomplete Information Equilibrium

I now turn to the full model to answer this paper's central question. Broadly, the equilibrium depends on whether Line 3's incentive compatibility condition holds. If it does, then regardless of what happens in the preemptive bargaining phase, the parties strike an agreement if they reach post-signal negotiations. In turn, A's risk-return

tradeoff moves exclusively to the preemptive bargain. From the complete information decision, it can buy off the unresolved type by demanding $p(m^*) + \frac{c_B + m^*}{V}$. This induces the resolved type to reject, invest at m^{**}, and reach a deal in the second phase. By analogous reasoning, A's alternative is to demand $p(m^{**}) + \frac{c_B + m^{**}}{V'}$. This induces the resolved type to accept and also buys off the unresolved type in the process. The appendix calculates that A prefers the risky strategy if:

$$q > \frac{\frac{m^{**}}{V'}}{p(m^*) - p(m^{**}) + \frac{c_B + m^*}{V} - \frac{c_B}{V'}} \tag{6}$$

When q exceeds that threshold, A calculates that the probability of facing the unresolved type is high enough that it is worth risking the risk of bargaining failure against the resolved type to capture more of the surplus from the unresolved type. Summarizing:

Proposition 1. *Suppose Line 3 holds. If Line 6 holds, A reaches a preemptive deal with the unresolved type and a post-signal deal with the resolved type in all PBE. If Line 6 fails, A reaches a preemptive deal with both types in all PBE. In either case, war never occurs.*

As anticipated, the inefficiency of costly signals drives A to seek out some surplus. Because incentive compatibility holds here, A could let the signals perfectly screen types for it. But A can also screen types by giving the unresolved type what it expects to receive through the signal. By doing so, A captures the $\frac{m^*}{V}$ value the unresolved type would have otherwise spent on the signal. In some cases, A may forgo screening entirely and buy off both types immediately. This may be surprising given that B pays for the signaling costs. However, the missing intuition is that A indirectly pays these signaling costs through the lost surplus it could have extracted. Learning might not be worth the implicit cost. As a result, a first-best solution to the information problem exists under the right circumstances.

Although coming from a model of costly signals and a costly lottery war, this result has a connection to the war-as-a-process literature (Wagner 2000; Filson and Werner 2002; Slantchev 2003; Powell 2004). In a multi-period war, each round of failed negotiations brings a new battle and further inefficiency. My model features multiple periods of inefficiency, but the sources are split between armaments and conflict. Nevertheless, a similar pattern emerges. If the uninformed actor has sufficient skepticism about its

opponent's resolve, it skims off the unresolved type in its first demand before coming to terms with the more resolved type later.[7]

When the incentive compatibility condition fails, A faces a more difficult strategic challenge. What happens next depends on A's prior belief. Line 4 still drives the decision. If it fails, per Lemma 2, the unresolved type knows it can pool on the resolved type's signal and force A to not call its bluff. Here, A can still buy both types' compliance by offering an amount commensurate with the resolved type's payoff. The alternative is to screen the unresolved type by demanding a smaller amount. Under this risk-return tradeoff, the appendix calculates that A prefers the riskier option if:

$$q > \frac{V}{V'} \tag{7}$$

Again, the threshold on q measures how strong A's belief must be that B is unresolved to justify the riskier proposal. The cutpoint is simple because, upon rejection, A places enough suspicion that B is resolved that it would propose $x_2 = p(m^{**}) + \frac{c_B}{V'}$ in the second phase even if the unresolved type pooled on rejecting x_1. In turn, the amount of surplus A must set its eyes on is either capturing the $\frac{m^{**}}{V}$ armament cost from the unresolved type or resign itself to guaranteeing just $\frac{m^{**}}{V'}$. Summarizing:

Proposition 2. *Suppose Line 3 fails and Line 4 fails. If Line 7 holds, A reaches a preemptive deal with the unresolved type and a post-signal deal with the resolved type in all PBE. If Line 7 fails, A reaches a preemptive deal with both types in all PBE. In either case, war never occurs.*

The implications here are similar to Proposition 1. A's decision does not alter the probability of war, just how much surplus it extracts and risk it takes in the first place. The only difference is how it calculates that balance. The unresolved type holds more leverage here because it knows that A will not call its bluff post-signal. The direct benefit for the unresolved type is a higher payoff should A pursue a risky bargaining strategy. But a subtler second-order consequence gives the unresolved type a

---

[7]In two-period war games, a third type of equilibrium can emerge where the proposer targets the lower value type in both periods, thereby risking war in the second (Spaniel and Bils 2018; Smith and Spaniel 2019). This does not arise here because incentive compatibility holds. As a result, the unresolved type is unwilling to bluff through the first inefficient phase. With the proposer's information problem endogenously solved, there is no risk of war in the second phase. The remaining cases where incentive compatibility fails therefore draw a more direct comparison to these existing models.

further advantage. Securing a deal with the resolved type requires a smaller unnecessary concession to the unresolved type. As a result, A pursues the safer bargaining strategy under a wider range of circumstances.[8] This is consistent with existing understandings of uncertainty premiums (Reed 2003; Arena 2013; Spaniel and Malone 2019). The difference between the resolved type's and the unresolved type's rejection payoffs is smaller when incentive compatibility fails.[9] The proposer becomes less inclined to take chances.

Taking stock, effective costly signals are a double-edged sword. Without preemptive bargaining, they at least guarantee no war inefficiencies. But the very fact that costly signals are effective makes sorting types look more attractive. In turn, avoiding the inefficiency of arms becomes more difficult.

This is not to say that ineffective costly signals are always efficiency improving. The discussion so far has focused on situations where A would not challenge potential bluffs. In the remaining case, A believes B is likely unresolved, corresponding to Lemma 3 if both types pooled on rejecting the initial proposal. The strategic interplay runs deep here. Broadly, A can choose from one of three options. First, it can still propose a quantity commensurate with the resolved type's value. This induces both types to accept.

Second, it can choose the skimming strategy also from Proposition 2. In fact, because the unresolved type is even more frequent than any case from Proposition 2, this strategy is better than the first possibility. As a result, A always incurs some risk within this parameter space.

Indeed, A's third option allows it to pursue an even riskier strategy. Suppose A proposes a quantity designed to pay less than what the unresolved type would earn through a successful bluff. The unresolved type cannot accept with certainty here. If it did, then A would infer B is resolved by observing the rejection and the m^{**} armament. Its post-signal concession would therefore be generous. In turn, the unresolved type would want to deviate to bluffing. However, the unresolved type cannot bluff as a pure strategy either. If it did, A would always call the bluff—the unresolved types are so

---

[8]One can observe this by noting that Line 7's cutpoint is strictly greater than Line 6's cutpoint. See Remark 4 in the appendix.

[9]That is, $\left(1 - p(m^*) - \frac{c_B + m^*}{V}\right) - \left(1 - p(m^{**}) - \frac{c_B + m^{**}}{V'}\right) > \left(1 - p(m^{**}) - \frac{c_B}{V'} - \frac{m^{**}}{V}\right) - \left(1 - p(m^{**}) - \frac{c_B + m^{**}}{V'}\right)$.

likely in this region that the signal is not believable.

The only remaining alternative is for the unresolved type to sometimes accept and sometime reject the lowball proposal. In the latter case, it then mimics the resolved type's armament. Now A is unsure how to respond, and so it sometimes calls the potential bluff. The resolved type rejects when it does.

In total, the last option introduces great inefficiencies: the resolved type always arms and sometimes fights a war, while the unresolved type sometimes arms too. As such, A must be very confident B is unresolved to try this strategy. The appendix calculates that level of confidence as:

$$
q > \frac{\left(\frac{V(V'c_A + c_B)}{c_B(V'-V)}\right)\left(\left(p(m^*) + \frac{c_B + m^*}{V}\right) - \left(p(m^{**}) + \frac{c_B}{V'}\right)\right)}{\left(1 + \frac{V(V'c_A + c_B)}{c_B(V'-V)}\right)\left(\left(p(m^*) + \frac{c_B + m^*}{V}\right) - \left(p(m^{**}) + \frac{c_B}{V'}\right)\right) - \frac{m^{**}}{V}} \tag{8}
$$

Summarizing:

Proposition 3. *Suppose Line 3 fails and Line 4 holds. If Line 8 fails, A reaches a preemptive deal with the unresolved type and a post-signal deal with the resolved type in all PBE. If Line 8 holds, the following occurs in all PBE. The unresolved type sometimes agrees to a preemptive deal. If it does not reach a preemptive deal, both types select m^{**}. Afterward, A mixes between concessions commensurate with the unresolved type's and the resolved types war payoffs. In the former case, war occurs with the resolved type. Otherwise, peace prevails.*

To provide more intuition for what happens when Line 8 holds, A has a deep-seated belief that B is unresolved. It could still fully skim the unresolved type in the first round of negotiations. However, doing so requires buying out the unresolved type's bluffing option. This requires paying concessions commensurate with the resolved type's optimal mobilization level and only withholding some concessions based on the unresolved type's higher cost to arm to that amount. If A is virtually certain B is unresolved, then that concession is not worth granting. Instead, it stands firm with its initial demand. Such a strategy does not yield a clean screen of the unresolved types, as there is still some incentive to bluff. As a result, A still is not certain that B is resolved following rejection and a high level of mobilization. Instead, it sometimes continues to make the small concession. Once more not receiving what it requires, the resolved type fights a war in that case.

5 Discussion and Empirical Implications

Having shown that states may wish to engage in preemptive negotiations to avoid costly signaling outcomes, I now turn to some of the deeper implications of the model. I focus on three points here: an efficiency comparison between the two models, the probability of war, and the levels of armaments observed on the equilibrium path.

5.1 The Efficiency of Preemptive Bargaining

This paper began with a simple premise. Costly signaling theory argues that a state's armament decision communicates information about its resolve, thereby reducing the risk of inefficiency from war. But even if signaling functions as intended, the result is only a second-best outcome: the solution to one inefficiency is a different type of inefficiency. Furthermore, Remark 1 showed that the inefficiency from arms exceeds the inefficiency from war in many cases. Thus, a natural question is whether states can do better if they can bargain over the signals as well.

On the surface, the answer appears to be affirmative. In standard costly signaling games, every interior equilibrium is inefficient.[10] This is because the types produce some level of armaments. No matter what happens in the bargaining phase, the first step of the interaction forecloses the possibility of an efficient equilibrium. In contrast, both Propositions 1 and 2 contain cases with fully efficient equilibria for the preemptive bargaining game.

Fortunately, the surface-level answer hides an even more optimistic result:

Remark 2. *Holding fixed a set of parameters, the equilibrium inefficiency is strictly smaller in the preemptive bargaining model than in the standard costly signaling model.*

A full explanation requires going through many cases. Figure 2 illustrates the logic when incentive compatibility holds, corresponding to Lemma 1 and Proposition 1. The total inefficiency under the costly signaling game is a convex combination of the respective optimal armament productions scaled by the corresponding type's resolve. Across

---

[10]In corner solutions where the marginal value for the first armament is smaller the marginal cost, no arms are constructed in equilibrium. If A believes B is sufficiently likely to be resolved, the corresponding demand guarantees the peace. This is the only way for the equilibrium to be efficient.
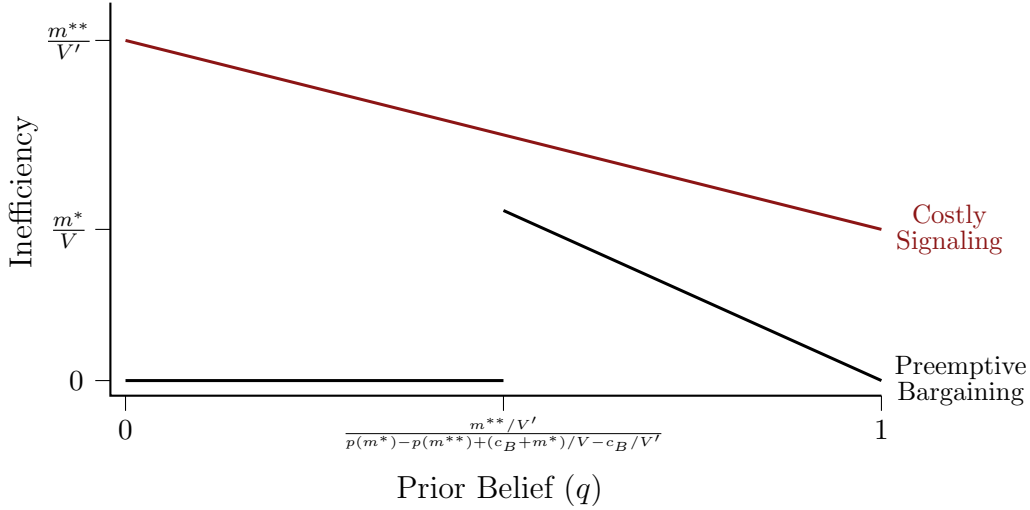
Figure 2: Expected equilibrium inefficiencies with and without costly signals.

the domain, the inefficiency under preemptive bargaining is smaller. For beliefs below the critical cutpoint, A reaches an immediate agreement with both types. As a result, the outcome is fully efficient. But inefficiency is still better above the critical cutpoint. There, the unresolved type never produces arms. Instead, the only inefficiency arises from resolved type's armament. Fortunately, this does not exceed what it produces under the costly signaling model, so the expected efficiency remains strictly less.[11]

From an arms control perspective, it is difficult to understate the improvement here. As Figure 2 suggests, armaments peak in the costly signaling game for lower prior beliefs.[12] This is not a situation where preemptive bargaining allows for some marginal decrease where arms proliferation would have been low anyway. Instead, it shrinks what would have been the largest outlay to *zero*.

Similar results apply without incentive compatibility. When A suspects that B is resolved (Lemma 2 and Proposition 2), no war occurs in either game. However, both types arm to the full m^{**} level in the standard costly signaling model. As such, the

---

[11]The inefficiency in the preemptive bargaining game can be higher or lower than a baseline model the prohibits any type of signaling, depending on where the cutpoints necessary for risky proposals lie and whether the signal is more or less costly than war.

[12]This is slightly different than what Figure 2 depicts. Total armaments in the costly signaling game's equilibrium are $qm^* + (1-q)m^{**}$. Total armaments therefore strictly decrease in q. In contrast, the figure shows inefficiency, which requires weighing each type's production by its valuation. As a result, the inefficiency can either increase or decrease in q.

efficiency problem here is worse. Preemptive bargaining again provides some help. If A is especially convinced that B is resolved, then the equilibrium proposal guarantees immediate acceptance and no inefficiency. Otherwise, it buys off the unresolved type. The only remaining inefficiency is the resolved type's production. But this is still strictly less than the original costly signaling inefficiency. Furthermore, the resolved type is the only one producing the armament, meaning the burden of the m^{**} resources falls on the type that internalizes the costs the least. Even better, that type cannot be especially likely, otherwise A would have brokered an immediate deal with both types.

When A believes B is unresolved, the comparison becomes murkier because of the possibility of war. Nevertheless, the central result still holds here. In the preemptive bargaining game, A buys off the unresolved type at least some portion of the time. Efficiency improves over the costly signaling model by removing those armaments from the tally. And as I show in a moment, the inefficiencies from war are weakly lower as well.

Zooming out, Remark 2 has an important implication for the design of international interactions. The efficiency gains indicate that states would want to structure their interaction according to the preemptive bargaining model and not the standard costly signaling setup. Because both players can effectively veto the initial negotiation, for a deal to work, it must improve A's welfare and at least one type of B's.[13] In fact, in this game, it improves both types' payoffs. For the armament decisions that costly signaling models describe, absent other strategic issues, nothing in principle stops the uncertain actor from making an early interjection into the bubbling crisis.[14] Its leadership can simply approach the rival, propose an efficient alternative, and have both parties benefit from the saved costs of armament.

---

[13]B can veto it directly through its rejection decision; A can indirectly veto it by offering no concessions.

[14]That is, preemptive bargaining is mutually agreeable and effective for publicly observable armaments. Researchers have identified some strategic limitations here. For hidden armaments in the shadow of preventive war and enforcement dilemmas, a variety of problems can arise (Debs and Monteiro 2014; Bas and Coe 2016; Spaniel 2019; Coe and Vaynman 2020). But this inherently runs against the goal of costly signaling decisions, which involve publicly observable actions that communicate information to surrounding actors.

5.2 War and Preemptive Bargaining

The central premise of costly signaling theory is that the price paid communicates information and reduces war. However, the side with private information sends fewer signals in the preemptive bargaining game. One may worry that the greater efficiency from a lack of arms leads to a higher probability of war. Thus, despite Remark 2's positive overall framing, the absence of signals results in more war.

Fortunately, this is not the case either:

Remark 3. *Holding fixed a set of parameters, the equilibrium probability of war is weakly smaller in the preemptive bargaining model than the standard costly signaling model.*

To better understand this result, note that both models share a common pathway to war. Fighting only occurs when Line 3's incentive compatibility constraint fails and the equilibrium is in semi-separating strategies. Within the semi-separating strategies, the equilibrium probability of war is the same. However, the costly signaling model only requires Line 4 to hold. The preemptive bargaining game is more stringent, requiring both Line 4 and Line 8 to be true. The extra condition means that the probability of war in the costly signaling model must be weakly lower than the probability of war in the preemptive bargaining model.

Substantively, the recipe for war under the costly signaling game is as follows. First, the proposer must have a strong prior belief that its opponent is unresolved. Second, this prior belief must be wrong. And third, after observing a large level of armament, the proposer must try to call what it perceives could be a bluff. The recipe for war is similar in the preemptive bargaining game. However, the proposer must start the process even more convinced that its opponent is unresolved. In the belief region that straddles these two thresholds, the probability of war goes from a strictly positive value to zero. This accounts for Remark 3's result.

Why is the probability of war lower despite the lack of costly signal production? The key insight here is that preemptive bargaining does not outright ban costly signals. If one of the types still finds sending the signal to be useful, it can always reject the initial proposal and mobilize. However, the proposer recognizes that relaying that message injects inefficiency and may therefore want to obviate the need to signal. For most of the parameter spaces, this means that the optimal demand is geared less toward avoiding

war and more toward securing the armament surplus. If the proposal is insufficient, then the resolved type can still indicate as much by arming.

The major difference arises in Proposition 3's parameters. There exist circumstances there where the unresolved type would ordinarily want to bluff resolve, knowing that the proposer might call that bluff. As such, the noisy signal fails to generate full information. The proposer may therefore calculate that ending the signaling process altogether is better. In short, the probability of war goes down overall because preemptive bargaining means that one more thing has to go wrong in the causal chain for war to occur.

5.3 Greater Variation in Armament Quantity

The model's final prediction is on the quantity of arms produced. Traditional costly signaling models predict a variety of armaments. When the incentive compatibility condition holds, the types each produce the armament quantity optimal for their valuation. This means a large amount for the resolved type, and a moderate amount for the unresolved type. When incentive compatibility fails, the decision can go in one of two directions. First, both types could produce the large quantity. Second, the resolved type could produce the large quantity, while the unresolved type sometimes chooses the small quantity and sometimes chooses the large quantity.

The preemptive bargaining model's key theoretical finding is that negotiations can supplant the costly signaling decision. But the manner by which that occurs is not exogenously random. Instead, the unresolved type's optimal armament is *never* produced in equilibrium. If the game sees any arming, it is to the resolved type's optimal level. When incentive compatibility holds, the proposer's optimal offer always buys off the unresolved type. It may or may not buy off the resolved type. Either way, though, the equilibrium only admits two armaments on the path: 0 and m^{**}. The same is true when incentive compatibility fails. The proposer may buy off both types, in which case production is 0. It may cleanly screen the types, leading the unresolved type to produce 0 and the resolved type to produce m^{**}. The same principle holds even under Proposition 3's aggressive screening strategy. Instead of mixing between a low armament and a high armament as in the costly signaling game, the unresolved type mixes between accepting and producing m^{**}. The middle ground of revealing its type by choosing m^*

goes away.

To be clear, the overall theory does not generally claim that it is impossible to observe low armaments on an equilibrium path. The proposer always tries buying off the least resolved type. In the context of a two-type game, this means the production quantity associated with the least resolved type is never produced. With more than two types, the proposer's optimal offer may buy off the least resolved type but not a more resolved—but still relatively low valuation—type. This would still lead to some variation in armament levels. But the key insight here is that the preemptive bargain still filters out the least resolved type's optimal production. In short, armaments appear to follow a "go big or go home" type of strategy, though the proposer's screening offer is what generates this process.

6   Conclusion

This paper explored a natural extension to workhorse models of costly signaling. War is tragic because both sides would benefit from a settlement. According to conventional wisdom, costly signaling facilitates such settlements by transmitting information about which of those agreements are mutually acceptable. But this means the prescribed solution to inefficiency is a different kind of inefficiency. Moreover, the signaling inefficiency may be worse than the equilibrium inefficiency from war.

Is a fully efficient solution possible? I show that the answer is yes. Under many circumstances, the proposer calculates that permitting costly signals to sort types is not a worthwhile use of resources. Instead, it makes concessions commensurate with what resolved types expect to receive by signaling, thereby obviating the need to arm or mobilize. At the very least, the proposer buys off unresolved types. The result is still inefficient but *less* inefficient than what would occur under standard costly signaling models.

Overall, the model indicates a need to reorient some policy recommendations. In the presence of uncertainty, costly signals are often suggested as a key step in reducing conflict outcomes (e.g., Walter 1997). To some extent, my model corroborates this idea. Nevertheless, policymakers should recognize signals for what they are: expensive, second-best solutions to a problem. Taking a passive role and allowing other states to signal their resolve may seem tempting. But the benefits are illusory. An uninformed

state is better off extracting at least some of the value of the signal for itself. Accordingly, policymakers should view costly signaling itself as a type of bargaining failure, in line with how researchers view arms races more generally (Kydd 2000; Rider 2013). It may be optimal to allow some portion of types to signal their intense preferences, but states should take a proactive stance and negotiate so that not all do.

References

Arena, Philip. 2013. "Costly signaling, resolve, and martial effectiveness." Manuscript, University of Buffalo.

Arena, Philip and Scott Wolford. 2012. "Arms, intelligence, and war." *International Studies Quarterly* 56(2):351–365.

Bas, Muhammet A and Andrew J Coe. 2016. "A dynamic theory of nuclear proliferation and preventive war." *International Organization* 70(4):655–685.

Bush, George HW and Brent Scowcroft. 1998. *A world transformed.* New York: Knopf.

Coe, Andrew J and Jane Vaynman. 2020. "Why arms control is so rare." *American Political Science Review* 114(2):342–355.

Dale, Catherine. 2008. "Operation Iraqi Freedom: Strategies, approaches, results, and issues for Congress." CRS Report for Congress `https://fas.org/sgp/crs/mideast/RL34387.pdf`.

Debs, Alexandre and Nuno P Monteiro. 2014. "Known unknowns: Power shifts, uncertainty, and war." *International Organization* 68(1):1–31.

Fearon, James D. 1995. "Rationalist explanations for war." *International Organization* 49(3):379–414.

Fearon, James D. 1997. "Signaling foreign policy interests: Tying hands versus sinking costs." *Journal of Conflict Resolution* 41(1):68–90.

Fearon, James D. 2018. "Cooperation, conflict, and the costs of anarchy." *International Organization* 72(3):523–559.

Filson, Darren and Suzanne Werner. 2002. "A bargaining model of war and peace: Anticipating the onset, duration, and outcome of war." *American Journal of Political Science* 46(4):819–837.

Frieden, Jeffry A, David A Lake and Kenneth A Schultz. 2013. *World politics: Interests, interactions, and institutions, second edition.* New York: Norton.

Garnett, Sherman W. 1995. The sources and conduct of Ukrainian nuclear policy: November 1992 to January 1994. In *Nuclear challenge in Russia and the new states of Eurasia*, ed. George Quester. New York: M.E. Sharpe pp. 125–152.

Hussein, Saddam. 1990. "Confrontation in the Gulf; proposals by Iraqi President: Excerpts from his address." New York Times: `https://tinyurl.com/y9jss9bc`.

Huth, Paul and Bruce Russett. 1993. "General deterrence between enduring rivals: Testing three competing models." *American Political Science Review* 87(1):61–73.

Koblentz, Gregory D. 2018. "Saddam versus the inspectors: The impact of regime security on the verification of Iraq's WMD disarmament." *Journal of Strategic Studies* 41(3):372–409.

Kydd, Andrew. 2000. "Arms races and arms control: Modeling the hawk perspective." *American Journal of Political Science* 44(2):228–244.

Miller, Steven E. 1993. "The case against a Ukrainian nuclear deterrent." *Foreign Affairs* 72(3):67–80.

Paine, Jack and Scott A Tyson. 2020. Uses and abuses of formal models in political science. In *The SAGE handbook of political science*, ed. Dirk Berg-Schlosser, Bertrand Badie and Leonardo Morlino. Oxford: Oxford University Press pp. 188–202.

Powell, Robert. 2004. "Bargaining and learning while fighting." *American Journal of Political Science* 48(2):344–361.

Reed, William. 2003. "Information, power, and war." *American Political Science Review* 97(4):633–641.

Reiss, Mitchell. 1995. *Bridled ambition: Why countries constrain their nuclear capabilities.* Baltimore: Johns Hopkins University Press.

Rider, Toby J. 2013. "Uncertainty, salient stakes, and the causes of conventional arms races." *International Studies Quarterly* 57(3):580–591.

Royce, Knut. 1990. "New Iraqi offer to U.S. reported." Tampa Bay Times: `https://tinyurl.com/ycdb656v`.

Schwartz, Stephen I. 2011. *Atomic audit: The costs and consequences of US nuclear weapons since 1940*. Washington, DC: Brookings Institution Press.

Slantchev, Branislav L. 2003. "The principle of convergence in wartime negotiations." *American Political Science Review* 97(4):621–632.

Slantchev, Branislav L. 2005. "Military coercion in interstate crises." *American Political Science Review* 99(4):533–547.

Slantchev, Branislav L. 2011. *Military threats: The costs of coercion and the price of peace*. Cambridge: Cambridge University Press.

Smith, Bradley C and William Spaniel. 2019. "Militarized disputes, uncertainty, and leader tenure." *Journal of Conflict Resolution* 63(5):1222–1252.

Spaniel, William. 2019. *Bargaining over the bomb: The successes and failures of nuclear negotiations*. Cambridge: Cambridge University Press.

Spaniel, William and Iris Malone. 2019. "The uncertainty trade-off: Reexamining opportunity costs and war." *International Studies Quarterly* 63(4):1025–1034.

Spaniel, William and Peter Bils. 2018. "Slow to learn: Bargaining, uncertainty, and the calculus of conquest." *Journal of Conflict Resolution* 62(4):774–796.

Stewart, Richard. 2010. *War in the Persian Gulf: Operations Desert Shield and Desert Storm August 1990–March 1991*. Washington, DC: Center of Military History.

Tarar, Ahmer. 2013. "Military mobilization and commitment problems." *International Interactions* 39(3):343–366.

Tyler, Patrick E. 1991. "Confrontation in the Gulf: Arafat eases stand on Kuwait-Palestinian link." New York Times: `https://tinyurl.com/y8nrd4oh`.

Wagner, R Harrison. 2000. "Bargaining and war." *American Journal of Political Science* 44(3):469–484.

Walter, Barbara F. 1997. "The critical barrier to civil war settlement." *International Organization* 51(3):335–364.

Wolford, Scott. 2007. "The turnover trap: New leaders, reputation, and international conflict." *American Journal of Political Science* 51(4):772–788.

Wolford, Scott. 2014. "Showing restraint, signaling resolve: Coalitions, cooperation, and crisis bargaining." *American Journal of Political Science* 58(1):144–156.

Woods, Kevin M, David D Palkki and Mark E Stout. 2011. *The Saddam Tapes: The inner workings of a tyrant's regime, 1978–2001*. Cambridge: Cambridge University Press.

7 Appendix

7.1 Proof of Lemmata 1–3

I begin with each type's accept/reject decision for a generic mobilization level. The unresolved type is willing to accept a deal if:

$$1 - x - \frac{m}{V} \geq 1 - p(m) - \frac{c_B + m}{V}$$

$$x \geq p(m) + \frac{c_B}{V}$$

Analogously, the resolved type is willing to accept if $x \geq p(m) + \frac{c_B}{V'}$. For the standard reasons, no additional equilibria exist in this subgame where a type rejects with positive probability when indifferent, so I assume acceptance when indifferent throughout the rest of the costly signaling subgame proof.[1]

Let ϕ be A's posterior belief that B is unresolved after observing m. Demanding $x > p(m) + \frac{c_B}{V}$ is not optimal because both types reject; A could profitably deviate to $x = p(m)$. Demanding $x < p(m) + \frac{c_B}{V'}$ is not optimal because both types would accept an unnecessary concession; A could deviate to any value between that x and $p(m) + \frac{c_B}{V'}$, induce both types to accept, and keep strictly more. Finally, demanding $x \in \left(p(m) + \frac{c_B}{V'}, p(m) + \frac{c_B}{V} \right)$ is not optimal. The resolved type rejects and the unresolved type accepts. However, A could deviate to any value between that x and $p(m) + \frac{c_B}{V}$, still have the resolved type reject and the unresolved type accept, and keep strictly more in the latter case while maintaining the same payoff in the former case.

In turn, only two demands can be optimal: $x = p(m) + \frac{c_B}{V}$ and $p(m) + \frac{c_B}{V'}$. The former induces the unresolved type to accept and the resolved type to reject; the latter induces both types to accept. All told, A prefers the risky demand if:

$$\phi \left(p(m) + \frac{c_B}{V} \right) + (1 - \phi)\left(p(m) - c_A \right) > p(m) + \frac{c_B}{V'}$$

$$\phi > \frac{c_A + \frac{c_B}{V'}}{c_A + \frac{c_B}{V}}$$

By analogous argument, A prefers the safe demand if $\phi < \frac{c_A + \frac{c_B}{V'}}{c_A + \frac{c_B}{V}}$ and is indifferent

---

[1]In contrast, rejecting with probability when indifferent is a requirement for equilibrium in a parameter space of the preemptive bargaining game.

if $\phi = \frac{c_A + \frac{c_B}{V'}}{c_A + \frac{c_B}{V}}$.

To begin solving the costly signaling decision, first note that the resolved type's payoff for the subgame for any value m and any belief ϕ equals $1 - p(m) - \frac{c_B + m}{V'}$. The resolved type receives that value directly when A's posterior belief results in the safe proposal; it receives that value indirectly when A's posterior belief results in the risky proposal and it rejects. As a result, for its armament decision, the resolved type has an objective function of:

$$1 - p(m) - \frac{c_B + m}{V'}$$

The first order condition of this is:

$$\frac{\partial}{\partial m} \left(1 - p(m) - \frac{c_B + m}{V'} \right) = 0$$

Rearranging yields Line 1. If such a solution m^{**} exists, it is unique and is the maximum.[2] This is because $\frac{\partial p}{\partial m}$ is strictly increasing and therefore it can only cross $\frac{1}{V'}$ once. No solution exists if $-\frac{\partial}{\partial m} p(0) < \frac{1}{V}$, and therefore it chooses $m = 0$ in this case. As the main text noted, I focus on the interior solution.

In turn, on any equilibrium path, A can only place positive probability on the resolved type after observing m^{**}. As such, on any equilibrium path where the unresolved type chooses anything other than m^{**}, A recognizes that it is the unresolved type with probability 1. The corresponding equilibrium demand is therefore $p(m) + \frac{c_B}{V}$ and the unresolved type accepts. Thus, if A wishes to not choose m^{**}, any on-the-path armament must optimize:

$$1 - p(m) - \frac{c_B + m}{V}$$

The first order condition of this is:

$$\frac{\partial}{\partial m} \left(1 - p(m) - \frac{c_B + m}{V} \right) = 0$$

Rearranging yields Line 2. The same caveats about the unique solution m^* as above apply.

---

[2] The second order condition is $-\frac{\partial^2 p}{\partial m^2} < 0$, which is immediately true.

2

Thus, on the path, the unresolved type can only choose m^* or m^{**}. Even in the best case scenario where A believes B is resolved after observing m^{**}, the unresolved type would not want to choose that quantity if:

$$(1 - p(m^*))V - c_B - m^* > \left(1 - p(m^{**}) - \frac{c_B}{V'}\right)V - m^{**}$$

$$\frac{m^{**} - m^*}{V} > \left(p(m^*) + \frac{c_B}{V}\right) - \left(p(m^{**}) + \frac{c_B}{V'}\right)$$

This proves Lemma 1. For the purposes of demonstrating existence, off-the-path beliefs that B is unresolved with probability 1 after observing anything that is not m^* or m^{**} is sufficient to sustain the equilibrium strategies. These off-the-path beliefs also establishes existence for the remaining cases.[3]

If the incentive compatibility condition fails, the unresolved type must still select either m^* or m^{**}. No equilibria exist where the unresolved type chooses m^* as a pure strategy. If it did, then A believes B is resolved with probability 1 after observing m^{**} and pays the greater concession. But because the incentive compatibility condition fails here, the unresolved type would want to deviate to m^{**}.

Now consider pooling strategies on m^{**}. Note that A must demand $p(m^{**}) + \frac{c_B}{V'}$ if $q < \frac{c_A + \frac{c_B}{V'}}{c_A + \frac{c_B}{V}}$. By the incentive compatibility condition, A prefers this to any other armament level.[4] This proves Lemma 2.

The remaining possibility is a semi-separating strategy where the unresolved type mixes between m^* and m^{**}. This requires indifference between the two choices. The unresolved type earns a guaranteed $1 - p(m^*) - \frac{c_B + m^*}{V}$ for choosing m^*. If it chooses m^{**}, the two feasible payoffs are $1 - p(m^{**}) - \frac{c_B}{V'} + \frac{m^{**}}{V}$ and $1 - p(m^{**}) - \frac{c_B + m^{**}}{V}$. Because incentive compatibility fails, the former quantity is greater than the amount the unresolved type earns from choosing m^*. However, the latter quantity is strictly worse. In turn, a mixture between those two offers yields indifference.

For it to be optimal for A to proposes both $p(m^{**}) + \frac{c_B}{V}$ and $p(m^{**}) + \frac{c_B}{V'}$, A must be indifferent. This requires a posterior belief of $\phi = \frac{c_A + \frac{c_B}{V'}}{c_A + \frac{c_B}{V}}$. Note that this belief is

---

[3]There is some slack in the beliefs, though. As long as an off-the-path belief places sufficiently high weight on the unresolved type, A's proposal still gives the unresolved type its war payoff and thus no desire to deviate.

[4]Again, this uses the belief that A is unresolved for any other signal. The class of PBE still allow for slack in those beliefs but do not permit any other equilibrium strategies.

impossible to induce if $q < \frac{c_A + \frac{c_B}{V'}}{c_A + \frac{c_B}{V}}$. Outside that, letting σ_H represent the probability that the unresolved type chooses the higher m^{**} armament, the unresolved type induces the necessary belief if:

$$\frac{q\sigma_H}{q\sigma_H + 1 - q} = \frac{c_A + \frac{c_B}{V'}}{c_A + \frac{c_B}{V}}$$

$$\sigma_H = \sigma_H^* \equiv \left(\frac{1-q}{q}\right)\left(\frac{V(V'c_A + c_B)}{c_B(V' - V)}\right)$$

Meanwhile, letting σ_G represent the probability A makes the more generous concession after observing m^{**}, the unresolved type is indifferent between the two armament quantities if:

$$\sigma_G\left(1 - p(m^{**}) - \frac{c_B}{V'} - \frac{m^{**}}{V}\right) + (1-\sigma_G)\left(1 - p(m^{**}) - \frac{c_B + m^{**}}{V}\right) = 1 - p(m^*) - \frac{c_B + m^*}{V}$$

$$\sigma_G = \sigma_G^* \equiv \frac{\left(p(m^{**}) + \frac{m^{**}}{V}\right) - \left(p(m^*) + \frac{m^*}{V}\right)}{c_B\left(\frac{1}{V} - \frac{1}{V'}\right)}$$

This proves Lemma 3.

7.2 Costly Signaling Efficiency without Incentive Compatibility

I now show that, when incentive compatibility fails, the efficiency comparison between the costly signaling model and the baseline game without costly signaling is similar to when incentive compatibility holds. The case where Line 4 fails is identical to before. Under that circumstance, the baseline game guarantees peace. Both types pool on m^{**} in the costly signaling game, so the expected inefficiency is $\frac{qm^{**}}{V} + \frac{(1-q)m^{**}}{V'}$. This is strictly greater than 0.

The more complicated case is where Line 4 holds. From the main text, the expected inefficiency from war in equilibrium equals $(1 - q)\left(c_A + \frac{c_B}{V'}\right)$. The expected inefficiency in the costly signaling game is more complicated due to the unresolved type's semi-separating strategy. Working through the expectations, with probability $1 - q$, Nature draws the resolved type. It signals m^{**} with certainty. Then A makes the generous

proposal with probability σ_G^* and therefore provokes war with the resolved type with probability $1 - \sigma_G^*$. Thus, the expected inefficiency from the resolved type's side equals $(1 - q)\left(\frac{m^{**}}{V'} + \sigma_G^*\left(c_A + \frac{c_B}{V'}\right)\right)$.

Meanwhile, Nature draws the unresolved type with probability q. With probability σ_H^*, the unresolved type arms to m^{**}; with probability $1 - \sigma_H^*$, it arms to m^*. Either way, the parties make a deal. As such, the total expected inefficiency attributable to the unresolved type is $q\left(\frac{\sigma_H^* m^{**} + (1 - \sigma_H^*)m^*}{V}\right)$. Thus, the total expected inefficiency equals:

$$q\left(\frac{\sigma_H^* m^{**} + (1 - \sigma_H^*)m^*}{V}\right) + (1 - q)\left(\frac{m^{**}}{V'} + \sigma_G^*\left(c_A + \frac{c_B}{V'}\right)\right) \tag{9}$$

From here, it is straightforward to replicate Remark 1 claim that there exists a critical value of q such that the costly signaling game generates a larger inefficiency. The key here is to recall that the expected inefficiency of the baseline game goes to 0 as q goes to 1. Meanwhile, σ_H^* goes to 0 as q goes to 1. Thus, Line 9 goes to $\frac{m^*}{V}$, which is strictly positive. Noting that Line 9 strictly decreases in q completes the proof.

7.3 Proof of Proposition 1

If any type rejects x_1, they play separating strategies in the following interaction, with the unresolved type choosing m^* and the resolved type choosing m^{**}. A's proposals then reach agreement at those levels of militarization. As such, the unresolved type earns $1 - p(m^*) - \frac{c_B + m^*}{V}$ for rejecting x_1, while the resolved type earns $1 - p(m^{**}) - \frac{c_B + m^{**}}{V'}$. Thus, the unresolved type is willing to accept $1 - x_1 \geq 1 - p(m^*) - \frac{c_B + m^*}{V}$, or $x_1 \leq p(m^*) + \frac{c_B + m^*}{V}$. Likewise, the resolved type is willing to accept $1 - x_1 \geq 1 - p(m^{**}) - \frac{c_B + m^{**}}{V'}$, or $x_1 \leq p(m^{**}) + \frac{c_B + m^{**}}{V'}$.

Now consider A's proposal. If it demands $x_1 < p(m^{**}) + \frac{c_B + m^{**}}{V'}$, both types must accept. None of these are optimal because A could deviate to an x_1 value between that quantity and $p(m^{**}) + \frac{c_B + m^{**}}{V'}$, still have both types accept, and keep strictly more for itself.

If A demands $x_1 > p(m^*) + \frac{c_B + m^*}{V}$, both types must reject. None of these are optimal. Going this route generates a payoff of $q\left(p(m^*) + \frac{c_B}{V}\right) + (1 - q)\left(p(m^{**}) + \frac{c_B}{V'}\right)$ regardless of the specific proposal. If it instead chooses any value $x \in \left(p(m^*) + \frac{c_B}{V}, p(m^*) + \frac{c_B + m^*}{V}\right)$, the unresolved type accepts and the resolved type rejects. This yields an identical payoff in the latter case but a strictly greater payoff in the former case because the alternative

5

amount is strictly greater than $p(m^*) + \frac{c_B}{V}$.

If A demands $x_1 \in \left(p(m^{**}) + \frac{c_B + m^{**}}{V'}, p(m^*) + \frac{c_B + m^*}{V}\right)$, the unresolved type accepts and the resolved type rejects. None of these are optimal because A could deviate to an x_1 value between that quantity and $p(m^*) + \frac{c_B + m^*}{V}$. The unresolved type still accepts, and the resolved type still rejects. In the latter case, A's payoff is identical. In the former case, A earns strictly more. Thus, this is a profitable deviation.

This leaves only two possible demands: $x_1 = p(m^*) + \frac{c_B + m^*}{V}$ and $x_1 = p(m^{**}) + \frac{c_B + m^{**}}{V'}$. The former induces the unresolved type to accept and the resolved type to reject, while the latter induces both to accept. The risky choice generates a greater expected payoff for A if:

$$q\left(p(m^*) + \frac{c_B + m^*}{V}\right) + (1 - q)\left(p(m^{**}) + \frac{c_B}{V'}\right) > p(m^{**}) + \frac{c_B + m^{**}}{V'}$$

$$q > \frac{\frac{m^{**}}{V'}}{p(m^*) - p(m^{**}) + \frac{c_B + m^*}{V} - \frac{c_B}{V'}}$$

By analogous argument, A demands the safe amount if the inequality is flipped. This proves the proposition.

7.4   Proof of Proposition 2

To begin, recall that the resolved type must reject $x_1 > p(m^{**}) + \frac{c_B + m^{**}}{V'}$. No matter how the unresolved type responds to such demands, because $q < \frac{c_A + \frac{c_B}{V'}}{c_A + \frac{c_B}{V}}$, the types of B must pool on m^{**} and the x_2 that follows must be $p(m^{**}) + \frac{c_B}{V}$. Thus, the unresolved type must also reject if:

$$1 - p(m^{**}) - \frac{c_B}{V'} - \frac{m^{**}}{V} > 1 - x_1$$

$$x_1 < p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V} \tag{10}$$

With that logic completed as a preface, no $x_1 < p(m^{**}) + \frac{c_B + m^{**}}{V'}$ can be optimal. Both types accept, making A's payoff equal to that x_1. But A can choose any value between that x_1 and $p(m^{**}) + \frac{c_B + m^{**}}{V'}$, still induce both types to accept, and generate a strictly greater payoff.

No $x_1 > p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V}$ can be optimal either. The resolved type rejects because

this leaves less than its optimal war payoff. Meanwhile, the unresolved type reasons that the pooling equilibrium of the costly signaling game will follow. By the calculation of Line 10, the corresponding proposal afterward generates a greater payoff than accepting. With the types pooling on m^{**} afterward, A's payoff for this outcome equals $p(m^{**}) + \frac{c_B}{V'}$.

Consider instead a deviation to $x_1 \in \left(p(m^{**}) + \frac{c_B + m^{**}}{V'}, p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V} \right)$. The unresolved type must accept in response, while the resolved type must reject. The former case generates a strictly greater payoff than the original strategy. The latter case generates a payoff identical to the original strategy. As a result, this is a profitable deviation.

Finally, no $x_1 \in \left(p(m^{**}) + \frac{c_B + m^{**}}{V'}, p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V} \right)$ is optimal. As just stated, the unresolved type must accept and the resolved type must reject this demand. But consider any deviation to an amount greater than the x_1 under consideration but strictly less than $p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V}$. The unresolved type still accepts and the resolved type still rejects. The payoff in the latter case is identical but the payoff in the former case is strictly greater. The deviation is therefore profitable.

Consequently, there are only two possible optimal demands: $p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V}$ and $p(m^{**}) + \frac{c_B + m^{**}}{V'}$. The former quantity induces the unresolved type to accept and the resolved type to reject. The latter induces both types to accept. Weighing the two, A prefers the risky demand if:

$$ q \left(p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V} \right) + (1 - q) \left(p(m^{**}) + \frac{c_B}{V'} \right) > p(m^{**}) + \frac{c_B + m^{**}}{V'} $$

$$ q > \frac{V}{V'} $$

By analogous argument, A makes the safe demand when the inequality is flipped. This proves the proposition.

The main text noted that this threshold is tighter than when incentive compatibility holds. The following remark restates this, with the proof for it thereafter:

Remark 4. *Suppose Line 3 fails. The threshold on q to make the risky demand is stricter than in the counterfactual world where Line 3 holds.*

Formally, this claim is that:

$$\frac{V}{V'} > \frac{\frac{m^{**}}{V'}}{p(m^*) - p(m^{**}) + \frac{c_B + m^*}{V} - \frac{c_B}{V'}}$$

Rearranging yields:

$$\frac{m^{**} - m^*}{V} < \left(p(m^*) + \frac{c_B}{V} \right) - \left(p(m^{**}) + \frac{c_B}{V'} \right)$$

This is the condition for incentive compatibility to fail, which is true for the case described and thereby completes the proof.

7.5 Proof of Proposition 3

The resolved type's decision rule is the same as under Proposition 2: it accepts $x_1 \geq p(m^{**}) + \frac{c_B + m^{**}}{V'}$ and rejects otherwise. The unresolved type's decision has a wrinkle to it, though. It must still accept $x_1 < p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V}$ because it cannot hope to receive anything more in the second stage. By the same argument, it is optimal for the unresolved type to accept that amount with probability 1. It must also reject $x_1 > p(m^*) + \frac{c_B + m^*}{V}$ because it can do better through the optimal armament and war.

Any $x_1 \in \left(p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V}, p(m^*) + \frac{c_B + m^*}{V} \right]$ is complicated. If the unresolved type accepts as a pure strategy, then A proposes $x_2 = p(m^{**}) + \frac{c_B}{V'}$ following m^{**}. But because such an x_1 is greater than $p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V}$, the unresolved type could profitably deviate to rejecting. Meanwhile, if the unresolved type rejects as a pure strategy and pools on m^{**}, A's prior belief places sufficient weight on the unresolved type that the corresponding demand will be $x_2 = p(m^{**}) + \frac{c_B}{V}$. This is worse than revealing itself by producing m^* instead. But that produces a demand of $x_2 = p(m^*) + \frac{c_B}{V}$. The unresolved type could profitably deviate to accepting because $x_1 < p(m^*) + \frac{c_B + m^*}{V}$. As such, the unresolved type must mix. Moreover, the mixture must be between accepting x_1 and producing m^{**} because the demand leaves more than what the unresolved type would receive by separating itself at m^*.

To induce indifference between x_1 and m^{**}, A's response to m^{**} must generate the same payoff for the unresolved type as keeping $1 - x_1$ would. The only way this can happen is if A mixes x_2 between a lower and a higher concession. But A has a unique optimal demand for any posterior belief except $\phi = \frac{c_A + \frac{c_B}{V'}}{c_A + \frac{c_B}{V}}$. As such, the unresolved type's mixture must generate this belief. Let σ_H be the probability that the unresolved

8

type rejects x_1 and arms to m^{**}. By Bayes' rule, A's posterior belief must be:

$$\frac{q\sigma_H}{q\sigma_H + 1 - q} = \frac{c_A + \frac{c_B}{V'}}{c_A + \frac{c_B}{V}}$$

Solving for this for the mixed strategy yields σ_H^*, just as in the costly signaling model.

And to induce the unresolved type's indifference, A must make the more generous demand with probability σ_G such that:

$$1 - x_1 = \sigma_G \left(1 - p(m^{**}) - \frac{c_B}{V'}\right) + (1 - \sigma_G)\left(1 - p(m^{**}) - \frac{c_B}{V}\right) - \frac{m^{**}}{V}$$

$$\sigma_G = \frac{p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V} - x_1}{c_B\left(\frac{1}{V} - \frac{1}{V'}\right)} \tag{11}$$

To narrow the remaining work, note that given σ_H^*, A is indifferent between demanding $x_2 = p(m^{**}) + \frac{c_B}{V'}$ and $x_2 = p(m^{**}) + \frac{c_B}{V}$. In the former case, both types accept. Consequently, I can write A's expected payoff as a function of x_1 as:

$$q(1 - \sigma_H^*)x_1 + (1 - q(1 - \sigma_H^*))\left(p(m^{**}) + \frac{c_B}{V'}\right)$$

The key inference here is that this utility strictly increases in x_1. As such, no $x_1 \in \left(p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V}, p(m^*) + \frac{c_B + m^*}{V}\right)$ is a part of an equilibrium—A can profitably deviate to a value larger than that x_1 but still within that range.

Taking stock, only three x_1 demands are still possible in equilibrium:

1. $p(m^{**}) + \frac{c_B + m^{**}}{V'}$

2. $p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V}$

3. $p(m^*) + \frac{c_B + m^*}{V}$

In case (1), both types accept immediately. A's payoff therefore equals that amount. In case (2), the unresolved type accepts immediately, followed by the resolved type accepting $x_2 = p(m^{**}) + \frac{c_B}{V'}$. The latter option is better if:

$$q\left(p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V}\right) + (1 - q)\left(p(m^{**}) + \frac{c_B}{V'}\right) > p(m^{**}) + \frac{c_B + m^{**}}{V'}$$

9

$$q > \frac{V}{V'}$$

This is true—the parameter space requires $q > \frac{c_A + \frac{c_B}{V'}}{c_A + \frac{c_B}{V'}}$, which is greater than $\frac{V}{V'}$. As such, the only equilibrium possibilities are (2) and (3). In case (3), the resolved type rejects, while the unresolved type accepts with probability σ_G from Line 11, substituting $x_1 = p(m^*) + \frac{c_B + m^*}{V}$. This works out to be σ_G^* from the original costly signaling model. Comparing the two, A prefers the riskier choice if:

$$q \left(1 - \left(\frac{1-q}{q}\right)\left(\frac{V(V'c_A + c_B)}{c_B(V' - V)}\right)\right)\left(p(m^*) + \frac{c_B + m^*}{V}\right)$$

$$+ \left(1 - q\left(1 - \left(\frac{1-q}{q}\right)\left(\frac{V(V'c_A + c_B)}{c_B(V' - V)}\right)\right)\right)\left(p(m^{**}) + \frac{c_B}{V'}\right)$$

$$> q\left(p(m^{**}) + \frac{c_B}{V'} + \frac{m^{**}}{V}\right) + (1 - q)\left(p(m^{**}) + \frac{c_B}{V'}\right)$$

Simplifying and putting in terms of q yields:

$$q > \frac{\left(\frac{V(V'c_A + c_B)}{c_B(V' - V)}\right)\left(\left(p(m^*) + \frac{c_B + m^*}{V}\right) - \left(p(m^{**}) + \frac{c_B}{V'}\right)\right)}{\left(1 + \frac{V(V'c_A + c_B)}{c_B(V' - V)}\right)\left(\left(p(m^*) + \frac{c_B + m^*}{V}\right) - \left(p(m^{**}) + \frac{c_B}{V'}\right)\right) - \frac{m^{**}}{V}}$$

By analogous argument, A makes the skimming proposal if the inequality is flipped. This proves the proposition.