When Can States Signal with Sunk Costs?

Noam Reich*

March 9th, 2023

Abstract

Canonical models of crisis bargaining maintain that states can convey their willingness to fight by signaling with sunk costs. For such signals to be informative, higher quality challengers must be more willing to invest in them. I argue that this condition seldom holds. States with higher payoffs to fighting have less to gain from settling disputes peacefully, and are therefore less willing to invest in costly signals to obtain a concession. In such cases sunk costs can only be used to reveal information when higher-quality types have a signaling advantage. This can either take the form of (1) differential costs, such that higher quality types pay less to produce sunk cost signals, or (2) index signals, i.e. signals that a lower quality type does not have the technology to imitate. These signaling advantages enable high-quality types to make relatively small investments in costly signaling without fear of imitation.

*I would like to thank Alexandre Debs, Matias Iaryczower, Kathy Ingram, James Morrow, and Kris Ramsay for their helpful discussions and suggestions. Additional thanks to audiences and participants at ISA 2023, Eurasian Peace Science 2023, APSA 2022, MPSA 2022, and SPSA 2021 who provided feedback on the project.

It is well known that states have incentives to misrepresent their strength or resolve in international crises to justify making larger demands and that this can lead to war (Fearon 1995). If a state is unable to verify whether its rival is as strong or resolved as it claims to be, then it may prefer to reject its rival's demands and risk war rather than make an unnecessary concession. Costly signaling theory maintains that the states that are most willing to go to war can overcome this problem by taking costly actions that distinguish them from lower quality types. If high-quality states can issue costly signals that a lower-quality type is unable or unwilling to imitate, then they can distinguish themselves and the receiver is able to infer that the sender is highly motivated to fight upon observing the signal.

Signaling theory presumes that since it is the highest-quality challengers who lose the most due to incentives to misrepresent private information that they should be those motivated to distinguish themselves though costly signaling. However, I argue that when states have private information regarding their strength or resolve, defined here as the costs states pay for fighting, then this does not hold. Intuitively the more attractive a state's outside option, the less it has to gain from settling a dispute peacefully. This suggests that stronger or more resolved states should be willing to devote less resources to signaling than their lower-quality counterparts. Early work managed to show that costly signaling was viable when states had private information regarding their *valuation* of the good or issue under dispute (Fearon 1997, Slantchev 2005). In this case, signaling is informative because states with low valuations would not bother to invest in obtaining the good peacefully or through war. However, this logic does not extend to other commonly recognized forms of uncertainty (Arena 2013, Carroll and Pond 2021).

In this article, I argue that signaling with sunk costs is still viable if stronger or more resolved states have a signaling advantage. This can take one of two forms. First, it is possible that stronger or more resolved states can produce sunk cost signals with less effort. Though weaker or less resolved states are willing to invest more total resource in getting their rivals to agree to a peaceful settlement, signaling becomes viable when these types need to input more effort to imitate a signal sent by a stronger type. Second, sunk-cost signaling is possible when stronger or more resolved states have unique technologies that enable them to issue signals but not weaker types. Originally developed in theoretical biology, such "index signals" are informative because the capacity to send the signal either reveals something about the technology itself or is a broader indicator of the signaler's capacity. Thus such signals derive are informative because weak types do not have the technology to imitate them, not because they require a great deal of effort to produce.

To explore these arguments further I consider an ultimatum crisis bargaining game in which a state with private information regarding its strength can issue a sunk cost signal. Initial analysis of the model demonstrates that sunk cost signals do not allow for increased information transmission and will not be used. To explain this result, I adopt a "budgeting" approach which asks how much a state with a certain level of strength would be willing to invest in sunk costs to "purchase" an increase in the probability that their rival concedes. Signaling requires that higher-quality types have a larger budget to spend on signaling then do lower-quality types for any given reduction in the probability of fighting generated by the signal. If a weaker type would be more willing to invest in sunk costs than a strong type would be, then the signal becomes uninformative. Early results suggest that as the sender's strength or resolve increases their "budget" for costly signaling decreases. This implies that weaker states will be willing to match spending on signaling by stronger states and suggests that signaling is only possible when stronger states are allowed to signal on the cheap.

Another set of results demonstrates that signaling is possible in a crisis bargaining game if strong states have a signaling advantage. When states have either of the signaling advantages described above, then the Fearon (1997) results wherein strong types signal and the Defender always concedes in response can be restored. This result contributes to a growing literature on signaling with differential costs. For example, Wolton (2020) shows that signaling with differential costs can be difficult when the receiver can then respond with a TIOLI offer. Additionally, Reich (forthcoming) shows that states can signal strength with handicaps, when strong states bear a small burden for handicapping themselves.

However, some scholars argue that pure sunk cost signals are rare and that most sunk cost signals take the form of arming, shifting the balance of power if a war occurs. I also show that arming does not function as a conventional signal in an ultimatum game when the sender has private information regarding their strength. First, if arming is at least moderately expensive then the Challenger will not arm at all. This occurs because the receiver can deter the sender from arming at any level by decreasing their acceptance probability whenever the sender arms. The only apparent exception is when arming is cheap enough that the sender can arm themselves to such a degree that the receiver would concede rather than risk war even when all types of the sender pool on the maximal demand. Note that in this case the results diverge from Slantchev (2005) in that weaker types arm themselves the most and the strongest types not at all.

The paper proceeds as follows. First, I show that sunk cost signaling is not viable in an ultimatum crisis bargaining game when strong types do not enjoy a signaling advantage. Second, I explain this result extend by using a “budgeting" approach, identifying the problem as weaker states increased willingness to spend on sunk costs. Third, I revert back to the ultimatum crisis bargaining model to show that sunk cost signaling is viable when the Challenger either enjoys a signaling advantage or when sunk cost signals take the form of an index signal. Finally, I demonstrate that sunk cost investments in arming are only viable in a crisis bargaining game when the weakest type of Challenger can arm itself enough that the Defender would prefer to concede.

The Ultimatum Model

Two countries play an ultimatum game for a divisible good of value 1. The Challenger makes a demand $x \in [0, 1]$ of the Defender. If the Defender accepts the offer, then the Challenger receives x, the Defender receives $1 - x$, and the game ends. If the Defender rejects the offer, then the two countries fight a war. War is modeled as a costly lottery where the winner receives the good (Fearon 1995). The Challenger's probability of winning is the war is $p(s)$ and is strictly increasing in their strength $s \in [\underline{s}, \bar{s}]$. The Defender has the reciprocal probability of winning $1 - p(s)$. I assume that the Challenger has private information regarding its strength and that the Defender's belief about this parameter are distributed according to the cumulative distribution function F with continuous support. Finally, if the countries go to war they each pay a cost for fighting c_1, c_2 respectively.

The Challenger can try to convey its strength by sending a sunk cost signal. When issuing a demand x, the Challenger can also choose to send a signal $m \in \mathbb{R}^+$ and incurs a linear cost m for doing so. Formally, a strategy for the Challenger is therefore a mapping from its private information strength type s to a choice of demand x and signal m and will be denoted by $\sigma : s \rightrightarrows [0,1] \times \mathbb{R}^+$. The Defender observes both the Challenger's demand and its signal choice m and must then choose with what probability it will concede. Formally, its strategy will be a mapping $\psi : x \times m \to [0,1]$. Together this implies that the Challenger's utility function will be given by

$$\psi(x,m)x + (1 - \psi(x,m))(p(s) - c_1) - m \tag{1}$$

I am looking for Perfect Bayesian Equilibrium as a solution concept. This requires that each type of the Challenger select an optimal strategy, that the Defender update its beliefs using Bayes' Rule whenever possible, and then select an optimal strategy conditional on those beliefs. I will use the function $G(s|x,m)$ to denote the Defender's posterior beliefs. A Perfect Bayesian Equilibrium will consist of a triple (σ, ψ, G) that satisfies the specified criteria.

Baseline Result

Without signaling, the Defender can only discourage weak types from bluffing via the risk-reward trade-off, rejecting larger demands with higher probability. Since weaker types have more to lose if the countries go to war, they will prefer to take on less risk and make strictly lower demands. It turns out that this results in separation, with each type of the Challenger making the unique demand that leaves the Defender indifferent. In turn, the Defender can identify the Challenger's type and select the unique probability of rejection that dissuades any other type from mimicking the demand. This separation can fail locally, if there exist types that the Defender would strictly prefer to concede to even if they demanded the entire good. In this case, weaker types take advantage of the Defender's reluctance to fight by bluffing and pooling on the maximal demand. The following Proposition summarizes the equilibrium strategies of both countries and is represented in Figure 1 (The proof of the claim can be found in Dal Bó and Powell (2009) and Reich (forthcoming) and is therefore omitted).

Proposition 1

In the ultimatum bargaining game with asymmetric information:

(i) There exists a Perfect Bayesian Equilibrium in which the Challenger plays

$$\sigma_x^*(s) \equiv \begin{cases} p_1(s) + c_2 \text{ if } s < \rho \\ 1 \text{ if } s \geq \rho \end{cases} \tag{2}$$

where ρ is given by

$$\rho \equiv \left\{ s' : c_2 = \frac{\int_{s'}^{\bar{s}} f(s) p_2(s) ds}{1 - F(s')} \right\} \tag{3}$$

The Defender responds by playing

$$\psi^*(x) \equiv \begin{cases} e^{-\frac{x - p(s) - c_2}{c_1 + c_2}} & \text{if } x < 1 \\ \frac{(c_1 + c_2) e^{-\frac{p_1(\rho) - p_1(s)}{c_1 + c_2}}}{1 - p_1(\rho) + c_1} & \text{if } x = 1 \end{cases} \tag{4}$$

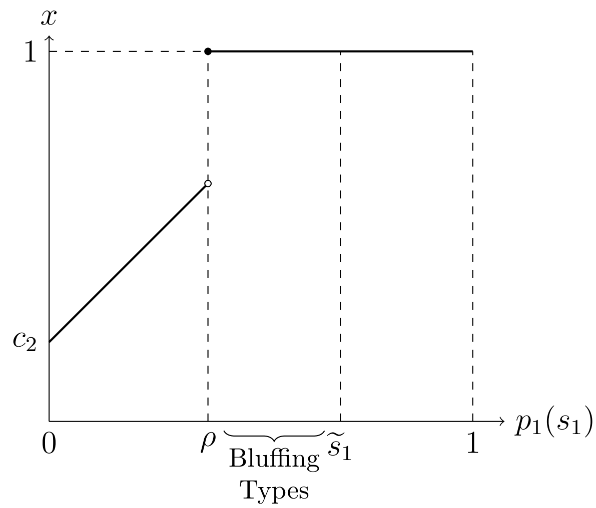

Figure 1: **Equilibrium Demands with a TIOLI Offer:** This figure plots how the Challenger's demand increases with their probability of winning. Challenger's weaker than type ρ demand $x = p(s_1) + c_2$ so that demands increase linearly up to that point. Type ρ is the weakest type to demand $x = 1$, and all types stronger than ρ do the same. Types in the range $[\rho, \widetilde{s}_1)$ are called bluffing types because they are the only one whose demands do not coincide with the demands they would make if the game were one of complete information.

and has beliefs

$$G^*(s|x) \equiv \begin{cases} 1 \ \textit{if}\ s = \sigma_x^{*-1}(x)\ \textit{and}\ x < 1 \\ \frac{f(s)}{1-F(\rho)}\ \textit{if}\ x = 1 \end{cases} \tag{5}$$

(ii) σ_x^* *is the unique signaling function subject to D1.*

Baseline Sunk Costs Result

The following Proposition states that sunk cost signals cannot change the equilibrium outcome and prevent bluffing.[1]

Proposition 2

No change occurs in the ultimatum game when the cost required to produce a signal m *is independent of type.*

[1]A proof of this Proposition is not provided, as it follows directly from the proofs in the following section.

The intuition for this result is straightforward. The D1 criterion continues to rule out any equilibrium that features pooling on any pooling on a demand except $x = 1$. The Challenger will only be willing to engage in sunk cost signaling if doing so increases the probability that the Defender concedes in response to a given demand. However, if the Defender ever increases the probability with which it concedes in response to larger demands, then the risk-reward trade-off breaks down. Any increase in utility obtained from costly signaling would encourage a weaker type to mimic the signal sent by a stronger counter-part.[2]

A Budget Approach to Signaling

In this section, I show that stronger or more resolved states will always be willing to invest less in sunk costs to achieve a peaceful resolution to a dispute. I will do so by using what I call a “budgeting" approach that asks how much a state is willing to pay to send a signal that will achieve an increased rate of concession. My results show that as a state becomes stronger or more resolved, the “budget" it is willing to invest in costly signals must strictly decrease; its better outside options implies that it will value a concession less. This in turn implies that weaker states will always be willing to “match" the investment of a stronger type in any concession and makes the signal uninformative.

A simple example illustrates the budgeting approach. Suppose that a type $s > \widetilde{s}$ were to send a signal m. So long as types weaker than s won't mimic the signal, then the Defender would strictly prefer to concede even when the Challenger demands the entire good. Let $U^*(s_1)$ denote the equilibrium expected utility for type s_1 in a Perfect Bayesian Equilibrium to the ultimatum game. Type s will be willing to send the sunk cost so long as the utility it

[2]Technically, there exist an infinite number of trivial equilibria in which the all types of the Challenger pool on some signal m and the Defender makes a corresponding decrease in the probability of war. However, the Challenger's demands are the same across equilibria and any of the Defender's strategies supportable in the equilibrium without signaling.

obtains from doing so is than its equilibrium expected utility

$$1 - m - U^*(s_1) \geq 0$$

The largest sunk-cost signal m that type s_1 will be willing to adopt is that for which this inequality holds at equality. Substituting for $U^*(s_1)$ using equation (1), we can rearrange to find that the largest signal s_1 that will be willing to signal whenever it is the case that

$$\overline{m}(s_1) = 1 - U^*(s_1) = [1 - \phi(1)][1 - p_1(s_1) + c_1] \tag{6}$$

It is immediately clear that the maximum amount that type s_1 is willing to invest in costly signaling is strictly decreasing in its wartime payoff. This implies that any type weaker than s_1 also issuing demand $x = 1$ in the Perfect Bayesian Equilibrium must also be willing to spend even more to send a signal that would induce the Defender to concede with certainty.

Signaling with Differential Costs

As the above section demonstrates, the issue with sunk cost signaling is that stronger or more resolved states are less willing to invest effort in achieving a peaceful outcome. One way to circumvent this issue is to have stronger or more resolved types receive a discount on signaling. When this occurs, higher-quality types will have their “budget" for signaling increased and so will be willing to expend greater amounts of effort than weaker types. When strong types enjoy such a signaling advantage, sunk cost signaling achieves the classic result in Fearon (1997), in which sunk costs do not eliminate bluffing but do induce the Defender to concede with probability 1.

Differential Costs in International Relations

Such signaling advantages might take many different forms in practice. For example, large military exercises or mobilizations designed to display strength might be easier for stronger or more organized militaries. Alternatively, consider military parades which can serve as periodic reminders of a state's resolve.[3] Insofar as a country's citizens are more likely to support such parades when they share in the nationalist sentiment such a parade represent, ease of staging such a parade and broader participation are more likely to come more easily to highly resolved states. For example, when US President Donald Trump attempted to stage a parade on July 4th, 2019, he struggled to achieve broad public support for the spectacle.

Introducing Differential Costs

As discussed there are two types of signaling advantage. The first, is one where stronger types require less effort to produce a signal. Formally, the cost function satisfies increasing differences so that for two types $s' > s''$ the larger types require less effort to produce a smaller signal. Letting m' and m'' be two signals with $m' > m''$ this requires that the cost or effort function to produce such a signal satisfy

$$e(m', s'') - e(m'', s'') > e(m', s') - e(m'', s') \tag{7}$$

Formally, equation (7) requires that the cost of producing a larger signal increase as a type becomes weaker.

The following Proposition demonstrates that if this discount is sufficiently large, then costly signaling will occur.

Proposition 3

If the following condition holds,

$$-\frac{\partial e(s,m)}{\partial s} > (1 - \psi(x))\frac{dp(s)}{ds} \tag{8}$$

then there exists an equilibrium where the Challenger issues demands according to equation

[3] This whole section could use some more references. If you know of any works in this vein, I would really appreciate it.

(2), and types $s \in [\rho, \bar{s}]$ also issue a sunk cost signal

$$m^* : e(m, \rho) = 1 - U^*(\rho) \tag{9}$$

where $U^*(\rho)$ is the utility obtained by ρ in Proposition 1. The Defender responds by conceding with probability 1 if the Challenger issues m^* and according to (4) otherwise and has beliefs given by G^* as defined in equation (12).

Again, the intuition for this result is straightforward. Type who pool on $x = 1$ produce a signal large enough such that the weakest type pooling on that demand is indifferent between signaling or not. So long as the condition in equation (8) is satisfied, then any type stronger than that type must strictly prefer to signal since they can do so for less effort, while any type weaker must strictly prefer not to signal. The receiver than knows that upon observing the signal that coincides with the demand $x = 1$ that the sender must be of the strong type and concedes with probability 1. By contrast the weaker separating types demanding $x < 1$ do not signal for the same reason as before, any gains from signaling would disrupt the risk-reward trade-off and entice weaker types to mimic that signal.

Index Signals

Whereas the signaling advantage described above allows for costly communication by ensuring that weaker types don't want to mimic costly signals, index signals are costly signals of strength that that weak types cannot mimic because they do not have the capacity to do so (Maynard-Smith Harper 2003). Such signals were originally identified in theoretical biology when researchers studying Scottish red deer observed that during the mating season large males protected their harems by roaring incessantly and loudly (Clutton-Brock and Algon 1979). Roaring in this way was costly and was found to correlate with a stag's ability to both deter and defeat challenges. However, index signals can also be quite cheap. For example, tigers have been observed to claw trees as high as they can possibly reach throughout their territories to advertise their size and strength to competitors. In both of these cases the signal is hypothesized to be informative because weak types cannot roar or mark trees in this way.

There is no direct parallel to index signals in the costly signaling literature in international relations. Slantchev (2011, 78) echoes the concept when he discusses how "military moves can, under certain circumstances reveal one's strength unambiguously." Additionally, Green and Long (2020) discuss the revelation of secret military technology as a source of strength and argue that whether such signals will be used depends on their impact on the military technology. Index signals of this form are probably incredibly common place. For example, the Doolittle Raid, the first instance in which the US bombed the Japanese mainland, is heralded as an important and symbolic victory that led Japan to pivot its strategy from expansion southwards to the east instead to head of the US (Jordan 2021). However, there is little reason to believe that index signals need to reveal new military technology. For example, much like the roaring of Scottish red deer, frequent missile testing by North Korea is likely designed with the appropriate frequency to continuously advertise capacity and the endurance capacity of missile stocks.[4]

Finally, it is important to note that it is possible for some types to bluff and issue index signals if the pool of types capable of doing so is sufficiently small. For example, in the 1930s Nazi Party's chief architect Albert Speer requested that the Luftwaffe provide him with a large number of search lights, then an important component of air-defenses, for use in party rallies for aesthetic purposes. Hermann Göring the head of the German Airforce refused arguing that the requested lights made up the bulk of Germany's strategic reserve. However, Hitler overruled Goring and recognized that the use of the search lights in this way provided a valuable signaling opportunity, arguing that it would suggest that Germany had a large reserve of searchlights (Speer 1970, 57-58).

[4]Seo, Yoongjun, Junko Ogura, and Jessie Yueng, "North Korea Launches 9th Missile Test of the Year" *CNN*, March 5th, 2022, URL: https://www.cnn.com/2022/03/04/asia/north-korea-missile-test-intl-hnk/index.html [last accessed: March 25th 2022].

Modeling Index Signals

To introduce index signals into the ultimatum game, we will now assume that there exist a set of types $[s^i, \bar{s}]$ with $\underline{s} < s^i$ that can choose whether or not they want to issue an index signal $m = \{1, 0\}$ at a cost e. The following Proposition demonstrates that there are two conditions required for an equilibrium in which the strongest types issue an index signal. First, the technology cannot be available to too many weak types. If $s^i < \rho$, then the signal cannot be informative. Second, the signal must be sufficiently cheap for the strongest types to be willing to issue it

Proposition 4

If $s^i \geq \widetilde{s}$ and $e < U^*(\bar{s})$ as achieved by $\bar{s}$ in proposition 1, then there exists an equilibrium in which all types capable of issuing an index signal do. In addition,

(i) If $s_i > \widetilde{s}$ then the Challenger issues demands according to the strategy in equation (2) where the only difference is that type ρ is now defined analogously as

$$\rho^i \equiv \left\{ s_1' : c_2 = \frac{\int_{s'}^{s^i} f(s) p_2(s) ds_1}{1 - F(s')} \right\} \tag{10}$$

(ii) If $s^i < \widetilde{s}$, then the Challenger plays

$$x^i(s) \equiv \begin{cases} p_1(s) + c_2 \ \mathit{if} \ s_1 < s_i \\ 1 \ \mathit{otherwise} \end{cases} \tag{11}$$

In both cases the Defender responds by playing ψ_x^* as defined in equation (4) if no index *signal is sent and $\psi = 1$ otherwise. Finally, the Defender has beliefs*

$$G^*(s_1|x,m) \equiv \begin{cases} 1 \text{ if } s = \sigma_x^{*-1}(x) \text{ and } x < 1 \\ \frac{f(s)}{F(s^i)-F(\rho^i)} \text{ if } x = 1 \text{ and } m = 0 \\ \frac{f(s)}{1-F(s^i)} \text{ if } x = 1 \text{ and } m = 1 \end{cases} \tag{12}$$

Moving forward, I plan to determine whether a conjecture that the sufficient conditions specified in the Proposition above for index signaling are also necessary for the use of index signals. However, the proof that these conditions are sufficient is straightforward and therefore omitted.

Sunk Costs as Investments in Arming

It is often argued that sunk cost signals reflect investments in military readiness, such as troop mobilizations or arming. Beyond burning money, such investments can shift the balance of power in favor of the signaler (Slantchev 2005, Post and Sechser 2022). Under these circumstances signaling generates two new benefits: (1) it allows the signaler to increase the size of its demand and (2) it increases the Challenger's payoff in case the Defender chooses to fight. Therefore, a Challenger could still invest in sunk costs signals to obtain these benefits even if the signal generates no shift in the Defender's posterior beliefs. This section studies how costly signaling operates when sunk costs increase the Challenger's strength. This can also be interpreted as states making observable arming decisions while making demands.

It turns out that arming is not always a viable strategy and that when arming does occur it is not the strong types who choose to arm. In the former case, the Defender can deter the Challenger from investing in arming when the costs of of arming are high enough to prevent the Challenger from making sufficiently large investments in arming. In the latter case, the cost of arming is low enough that the Challenger cannot be deterred from arming. In this case, the weakest types all separate, each arming themselves just enough to make the

Defender indifferent between fighting them and not when they demand the full value of the good $x = 1$. Strong types of the Challenger also demand $x = 1$, but do not arm. In turn the Defender concedes with greater probability in response to larger investments in arms. This occurs because stronger types are less willing to make the sunk cost investments necessary to reduce the probability of war.

Formally, let the Challenger's probability of winning a war now be a concave function of its level of signal $p_1(s_1, m)$. Moreover, assume that no type enjoys any advantage in arming so that $\frac{\partial^2 p_1(s_1,m)}{\partial s_1 \partial m} = 0$. The Challenger has the following expected utility

$$U(x, m|s_1) = x\psi(x, m) + (1 - \psi(x, m))[p_1(s_1, m) - c_1] - m \tag{13}$$

The following Proposition provides a complete characterization of the PBE.

Proposition 5

When sunk cost signals contribute to an increase strength that is independent of type, then there exists a Unique Perfect Bayesian Equilibrium subject to the D1 refinement that has the countries play as follows:

(i) If the following inequality holds

$$e(m^*) > 1 - p_1(\underline{s}) + c_2 \textit{ for } m = 1 - p(\underline{s}) - c_2 \tag{14}$$

then no arming occurs.

(ii) Otherwise there exists a PBE where the Challenger plays

$$\sigma^*(s) = \begin{cases} x = 1 & \\ m = 1 - p(s_1) - c_2 & \textit{if } s < \widetilde{s} \\ m = 0 & \textit{otherwise} \end{cases} \tag{15}$$

(iii) The Defender responds by playing

$$\frac{\psi(1,m)}{1-\psi(1,m)} = \frac{1}{c_1+c_2} - \frac{\frac{\partial p_1(s_1,m)}{\partial m}}{1-\psi(1,m)} \tag{16}$$

in response to any $m > 0$ *with the initial condition that* $\psi(1, 1-p(\underline{s})-c_2) = 1$ *and will play*

$$\psi(1,0) = \frac{\psi(1,m(\rho))[1-p(\rho)-m(\rho)+c_1] + p_1(\rho,m(\rho)) - p_1(\rho,0) - e(m)}{1-p(\rho)+c_1} \tag{17}$$

(iv) The Defender will have posterior beliefs

$$G(s|x,m) = \begin{cases} 1 & \textit{if } x = 1 \textit{ and } p(s) = 1-m-c_2 \\ \frac{f(s)}{1-F(\rho)} & \textit{if } x = 1 \textit{ and } m = 0 \\ 0 & \textit{otherwise} \end{cases} \tag{18}$$

The following is the intuition underlying the result. If condition stated in part (i) is satisfied, then the weakest type of the Challenger is unwilling to pay the costs of arming required for the Defender to strictly prefer to concede even when the weakest type demands $x = 1$. If the weakest type of the Challenger were to instead select any lower level of arms $m \in (0, 1-p(\underline{s})-c_2)$, it could then separate by issuing the demand $x = p(\underline{s})+c_2+m$. Though the Defender is indifferent between accepting and rejecting the demand, the Defender is strictly worse off now that the Challenger has armed. Therefore, the Defender has an incentive to discourage the Challenger from arming and can do so by increasing the probability that it will go to war if the Challenger arms itself. The Defender can increase the risk of war in this way because the risk-reward trade-off only requires that the Defender increase the probability it goes to war in response to higher demands, while leaving the Defender free to select the *baseline level of risk*. Being indifferent, the Defender can can increase its baseline level of rejection in response to the Challenger arming itself and increase the risk of war

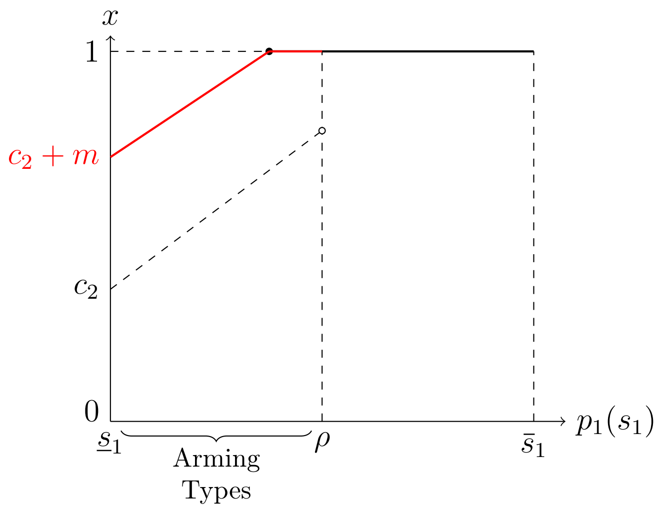

Figure 2: **Equilibrium Demands under Sub-optimal Arming:** This figure plots how the Challenger's demand would change under medium levels of arming. Challengers weaker than type ρ either separate by pooling on a level of arming m and then demand $x = p(s) + m + c_2$ or they pool on $x = 1$ and select a unique m so that $1 = p(s) + m + c_2$. Their new demands are colored in red while their demands in the bargaining baseline are represented by the dashed line. Types in the range $[\rho, \bar{s})$ do not arm.

enough to deter the Challenger from arming. Figures 2 and 3 illustrate the logic underlying the equilibrium.

On the other hand, if the inequality in equation (14) is not satisfied, then arming does occur. This is because the weakest type of the Challenger is willing to acquire a quantity of arms large enough to ensure that the Defender would strictly prefer to concede in response to a demand $x = 1$. In this case, the weakest type of the Challenger proceeds to arm enough to leave the Defender indifferent between fighting and accepting the demand $x = 1$. The Defender proceeds to accept that demand with probability 1, because if it did not, the weakest type of the Challenger could simply increase its level of arming by any arbitrary small $\epsilon > 0$ and the Defender would strictly prefer to concede. Similarly, types $s_1 \in (\underline{s}_1, \rho)$ separate, arming themselves to a level that leaves the Defender indifferent between accepting $x = 1$ and going to war. Since the Defender is indifferent in response to any such demand, the only requirement of its strategy is that ensures that the Challenger's strategy is incentive compatible. To do this, the Defender increases the probability that it goes to war

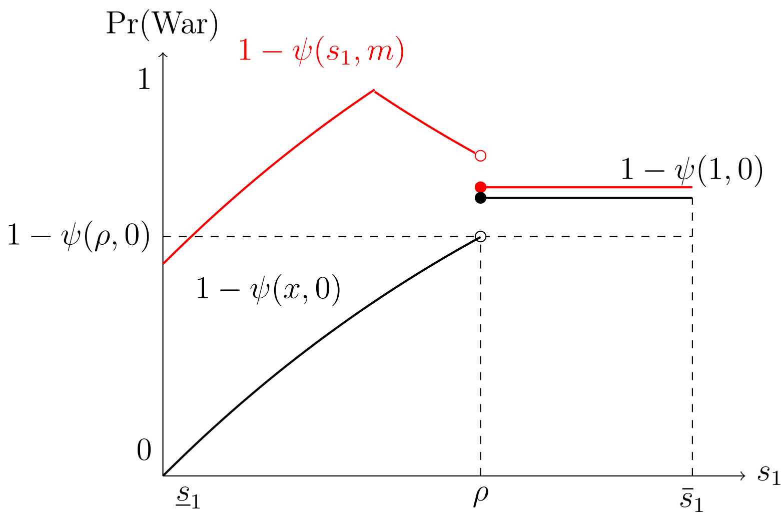

Figure 3: **The Probability of War under Sub-optimal Arming**: This figure plots the probability of war when the inequality in equation (14) is not satisfied, and weaker types of the Challenger arm anyway as depicted in Figure 2. The baseline probability of war in the equilibrium without arming is plotted in black, increasingly steadily as demands increase with a discontinuity between the separating demand type ρ would have made and $x = 1$. The probability of war when the Challenger arms is depicted in red and features a higher baseline level of risk of war for all demands. When weak types of the Challenger arm, there is a similar initial continuous increase in the probability of war as the Challenger's strength and demands increase, a manifestation of the risk-reward trade-off. The continuous decrease in the probability of war under arming occurs when as weak types demand $x = 1$, but steadily lower their level of arms. The discontinuity at ρ occurs because type ρ discontinuously reduces their level of arms to zero. The increase in the baseline rate of war is sufficient to deter the Defender from arming.

in response to smaller levels of arms. The risk of fighting discourages initially weaker types from arming to lower levels, while the cost of arms discourages initially stronger types from arming themselves in exchange for a higher probability of concession. However, sufficiently strong types $s_1 \in [\rho, \bar{s}_1]$ do not arm themselves as the Defender is already indifferent between accepting and not when these demand $x = 1$ without additional arming. The probability that the Defender concedes in response to no arming is designed to ensure that type ρ is indifferent between arming and separating as the weaker types do and pooling on no arming with the stronger unarmed types.

These result contradict much of the standard inherited wisdom from costly signaling theory which postulates that stronger states should invest more in signaling. Rather, it demonstrates that absent a signaling advantage and when sunk cost signals represent an investment in arming, it is possible to achieve an equilibrium in which weaker types invest more in sunk cost signals. However, still upheld from the conventional wisdom is that larger signals are more likely to elicit a concession (and that consequently ex-ante weaker states are less likely to fight). This should make empirical scholars wary of drawing conclusions about the pool of types that do or do not engage in signaling behavior.

Conclusion and Future Directions

This paper builds on a burgeoning literature in signaling theory and mechanism design to argue that sunk costs signals are only possible for revealing strength and resolve if the sender has a signaling advantage. I show that absent such a signaling advantage weaker types will be willing to mimic any signal and make it uninformative. I also show that sunk cost signaling may not be possible even when it takes the form of arming and that in any case weaker types invest more in arming than do stronger types. Finally, I show that when the sender has a signaling advantage, either in the form of lower differential costs or the ability to send index signals, then sunk cost signaling is possible.

Appendix

Proof of Proposition (??)

Following the arguments in the main text, $\overline{m}(s)$ will be given by

$$\overline{m}(s) = x - x^*\phi^*(x^*) - (1 - \phi^*(x^*))(p_1(s) - c)$$

which by the envelope theorem has the following derivative with respect to s

$$\frac{d\overline{m}(s)}{ds} = -(1 - \phi^*(x^*))\frac{dp(s)}{ds}$$

which is decreasing for all $\phi^*(x^*) < 1$. Fey and Ramsay (2016) show that $\phi^*(x^*)$ can only be equal to 1 in response to the lowest possible demand. Since Lemma 1 demonstrates that any equilibrium without signaling must satisfy the single-crossing property. It follows that any set of types issuing the lowest demand must form a (possibly degenerate) connected interval $[\underline{s}, s']$. ■

Proof of Proposition 5

The proof makes use of the following two lemmas. The first, demonstrates that for any fixed m, the Challenger's strategy must be weakly increasing in strength. The second demonstrates that when we impose the D1 Criterion, the only pair of demand and arming level that can be pooled upon is $(1, 0)$. Together these two lemmas imply that in the set $s_1 \in [\underline{s}_1, \rho)$ must separate and select a unique pair (x, m), while types $s_1[\rho, \bar{s}_1]$ must pool on $(1, 0)$.

Lemma 1

For the Challenger's demand x must be weakly increasing in type s_1.

Proof: Following Ashworth and Bueno de Mesquita (2006), to prove this claim it is only necessary to verify that the Challenger's utility function specified in equation (13) satisfies single crossing. The cross-partial of equation (13) with respect to x and s_1 is given by

$$-\frac{\partial \psi(x,m)}{\partial x}\frac{\partial p_1(s_1,m)}{\partial s_1} > 0$$

which is in fact positive given that $\frac{\partial \psi(x,m)}{\partial x}$ is decreasing. To check that is true it is only necessary to take the first-order condition of equation (13) with respect to x and rearrange to find that

$$\frac{-\frac{\partial \psi(x,m)}{\partial x}}{\psi(x,m)} = \frac{1}{x - p(s_1,m) + c_1}$$

This is sufficient to prove the claim.■

The following proposition follows identical steps to that in Reich (2022, Proposition 1) and is therefore omitted.

Lemma 2

No equilibrium to the ultimatum game can survive D1 that has countries pooling on any pair (x, m) *other than* $(1, 0)$.

We can now proceed to show that types $s_1 \in [\rho, \bar{s}_1]$ must pool on $(1, 0)$ and that the remaining types must separate. Let type $\widetilde{s}_1$ denote the type for which $p_1(\widetilde{s}_1) + c_2 = 1$, so that the Defender is indifferent between fighting and not when that type demands the whole value of the good and does not arm. Type $s_1 \in (\widetilde{s}_1, \bar{s}_1)$ cannot separate - for any pair $(x, 0)$, the Defender would strictly prefer to concede. Any type demanding a lower x, then the largest x demanded by any such type would then prefer to deviate to that largest x. It follows that the Defender must mix, which requires that types $[\widetilde{s}_1, \bar{s}_1]$ pool on $(1, 0)$ with enough weaker types for the Defender to be indifferent between fighting and not. Lemma (1), implies that it must be the set of types $[\rho, \widetilde{s}_1]$ who pool in this way. Thus types in the set $[\rho, \bar{s}_1]$ pool on $(1, 0)$ and types in the range $s_1, \in [\underline{s}_1, \rho)$ must separate, selecting a unique combination (x, m) for which $p_1(s_1, m) + c_2 = x$.

It turns out that the separating types will only separate on a single dimension - either they will issue a unique x and pool on a level of arms, or they will choose to pool on a demand x and each issue a unique level of arms. This is proven using two lemma.

Lemma 3

Let $[s_1', s_1'']$ be a (possibbly degenerate) set of types who pool on a level of arms m^p. Then,

- *(i) Each type issues a unique demand x satisfying $x = p_1(s_1, m^p) + c_2$*
- *(ii) The Defender responds by playing*

$$-\frac{\psi'(x, m^p)}{1 - \psi(x, m^p)} = \frac{1}{c_2 + c_1}$$

(iii) If $s_1' = \underline{s}_1$, then the Defender can make the baseline level of acceptance $\psi(p_1(\underline{s}_1 + +m^p + c_2, m^p)$ a function of m^p.

Proof: Parts (i) and (ii) follow from the arguments laid out above, and the following: taking the first order condition of equation (13) with respect to x produces

$$-\frac{\psi'(x, m^p)}{1 - \psi(x, m^p)} = \frac{1}{x - p_1(s_1, m^p) + c_1}$$

To see that this is indeed a maximum, it is only necessary to show that the second order condition is negative. Taking the derivative of equation (13) with respect to x, we are left with

$$\psi''(x, m^p)(1 - p_1(s_1, m^p) + c_1) + 2\psi'(x, m)$$

Recall, that the Defender's strategy satisfies

$$-\frac{\psi'(x, m^p)}{1 - \psi(x, m^p)} = \frac{1}{c_2 + c_1}$$

thereby implying that

$$\psi''(x, m^p) = -\frac{\psi'(x, m)}{c_1 + c_2}$$

Substituting this into the second order condition, we are left

$$\psi'(x) < 0$$

This suffices to prove parts (i) and (ii).

Part (iii) follows from the observation that

$$-\frac{\psi'(x, m^p)}{1 - \psi(x, m^p)} = \frac{1}{x - p_1(s_1, m^p) + c_1}$$

is a differential equation with solution

$$e^{\frac{x}{c_1+c_2}} + d$$

If type $\underline{s}_1$ pools on m^p, then Defender can condition d on m^p. Otherwise, d must be selected to ensure that weaker types find their demands incentive compatible and don't want to mimic the strategy select by s' and vice versa. ■

Lemma (1) implies that the Challenger's demand must be increasing in type s_1. Let type $\widetilde{s}_1(m) < \rho$ be the weakest type to demand $x = 1$ for $m > 0$. Types in the set $[\widetilde{s}_1(m), rho)$ must therefore all demand $x = 1$ and select the unique m for which $p_1(s_1, m) + c_2 = 1$. Let $d(m)$ denote the value of d that the Defender selects in response to the Challenger's level of arms. The following lemma demonstrates that all $s_1 \in [\underline{s}_1, \widetilde{s}_1(m)]$ must prefer to pool on the same level of arms m^p whenever the Defender conditions d on the Challenger's decision to acquire a positive amount of arms, but not the quantity of arms.

Lemma 4

Let $d(m) = k$ for all positive level of arms, where k is a constant. Types of the Challenger demanding $x < 1$ will all pool on the same level of arms m^p whenever the Defender otherwise sets $\frac{\partial \psi(x,m)}{\partial m} = 0$.

Proof: Because separating types must demand $x = p_1(s_1, m) + c_2$, x is a function of m. Substituting in for the value of x, we can rewrite the Challenger's expected utility function in equation (13) as

$$p_1(s_1, m) - c_1 + \psi(x, m)[c_1 + c_2] - m$$

Taking the first-order condition with respect to m, we find that we are left with

$$\frac{\partial p_1(s_1, m)}{\partial m} = 1$$

The left-hand side is the same for all types s_1 by assumption. The second-order condition is given by

$$\frac{\partial^2 p_1(s_1, m)}{\partial^2 m}$$

which is negative and also independent of type. ■

This previous lemma implies that the level of arming optimal for type $\underline{s}$ determines the strategy for all types. There remain two steps in the proof. The First is to show that types in the range of $[\widetilde{s}_1(m), \rho)$ do not have an incentive to deviate from their strategy. The second, is to show that so long as condition (14) is not met, then the Defender can select a value for k such that arming is not optimal and must select d such that $\psi(1, p^{-1}(\underline{s}_1, x - c_2)) = 1$ otherwise.

First, the expected utility function for a type $s_1 \in (\widetilde{s}_1(m), \rho)$ is given by

$$\psi(1, m) + (1 - \psi(1, m))[p(s_1, m) - c_1] - m$$

Taking the first-order condition with respect to m, we find that we are left with the expression in equation (16). To check that this indeed produces a maximum, we will check that the second-order condition is negative. Taking the second-derivative of the Challenger's expected utility function with respect to m, we find that it is negative whenever

$$\frac{\partial^2 \psi(1,m)}{\partial^2 m}[1 - p_1(s_1,m) + c_1] - 2\frac{\partial \psi(1,m)}{\partial m}\frac{\partial p_1(s_1,m)}{\partial m} + \frac{\partial^2 p_1(s_1,m)}{\partial^2 m}$$

Once again, we cannot proceed without additional information regarding the second-derivative of $\psi(1,m)$. To derive its value, we can rearrange the expression in equation (16) and take its derivative with respect to m to find that

$$\frac{\partial^2 p_1(s_1,m)}{\partial^2 m}(c_1 + c_2) = -\frac{\partial^2 p_1(s_1,m)}{\partial^2 m}(1 - \psi(1,m)) + \frac{\partial \psi(1,m)}{\partial m}\frac{\partial p_1(s_1,m)}{\partial m}$$

Substituting this back into the second-order condition, we find that we are left with

$$-\frac{\partial \psi(1,m)}{\partial m}\frac{\partial p_1(s_1,m)}{\partial m} < 0$$

which must hold.

References

Arena, Philip (Sept. 2013). *Costly Signaling, Resolve, and Martial Effectiveness*. Tech. rep.

Banks, Jeffrey S. (1990). "Equilibrium Behavior in Crisis Bargaining Games". In: *American Journal of Political Science* 34.3, pp. 599–614.

Carroll, Robert and Amy Pond (July 2021). "Costly signaling in autocracy". In: *International Interactions* 47.4, pp. 612–632.

Clutton-Brock, T. H. and S. D. Albon (Jan. 1979). "The Roaring of Red Deer and the Evolution of Honest Advertisement". In: *Behaviour* 69.3-4, pp. 145–170.

Dal Bó, Ernesto and Robert Powell (2009). “A Model of Spoils Politics”. In: *American Journal of Political Science* 53.1, pp. 207–222.

Fearon, James D. (1995). “Rationalist Explanations for War”. In: *International Organization* 49.3, pp. 379–414.

— (Feb. 1997). “Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs”. In: *Journal of Conflict Resolution* 41.1, pp. 68–90.

Fey, Mark and Brenton Kenkel (Jan. 2021). “Is an Ultimatum the Last Word on Crisis Bargaining?” In: *The Journal of Politics* 83.1, pp. 87–102.

Fey Mark and Ramsay Kristopher W. (Oct. 2010). “Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict”. In: *American Journal of Political Science* 55.1, pp. 149–169.

Green, Brendan Rittenhouse and Austin Long (Jan. 2020). “Conceal or Reveal? Managing Clandestine Military Capabilities in Peacetime Competition”. In: *International Security* 44.3, pp. 48–83.

Jordan, Richard (Sept. 2021). “Symbolic victories and strategic risk”. In: *Journal of Peace Research* 58.5, pp. 973–985.

Reich, Noam (n.d.). “Signaling Strength with Handicaps”. In: *Journal of Conflict Resolution* ().

Slantchev, Branislav L. (Nov. 2005). “Military Coercion in Interstate Crises”. In: *American Political Science Review* 99.4, pp. 533–547.

— (Feb. 2011). *Military Threats: The Costs of Coercion and the Price of Peace*. Cambridge University Press.

Smith, John Maynard and David Harper (Nov. 2003). *Animal Signals*. OUP Oxford.

Speer, Albert (1970). *Inside the Third Reich: Memoirs*. New York: Macmillan.

Wolton, Stephane (June 2019). *Signaling in the Shadow of Conflict*. SSRN Scholarly Paper ID 3100989. Rochester, NY: Social Science Research Network.