Online Appendix for Talking to the Enemy: Explaining the Emergence of Peace Talks in Interstate War

Oriana Skylar Mastro* David A Siegel†

In this appendix we prove the propositions in the text, using throughout the Perfect Bayesian Equilibrium solution concept. To that solution concept, we apply the Iterated Intuitive Criterion equilibrium refinement (see Fudenberg and Tirole 1991, 449). We choose this refinement because it matches the logic of the model: opening negotiations is a signal of state B's costs of war, and even if opening occurs off the equilibrium path, we expect state A to rationally respond to that signal. The intuitive criterion effectively disallows off-the-path beliefs that place positive probability on types of B that would do strictly worse if they were to deviate from equilibrium play. Iterating the criterion allows the standard game theoretic infinite regress: state A's belief updating affects its future offer, shrinking the set of types of state B that would consider a deviation beneficial, leading state A to further update its beliefs, and so on. Use of the Iterated Intuitive Criterion is not required, however. We could have equally well used the less involved but stricter D1 criterion (see Fudenberg and Tirole 1991, 452), but we felt that the process encoded in the Iterated Intuitive Criterion better matched the unraveling present in our model.

Proof of Proposition 1

Given the cutoff equilibrium described in the text, and the fact that $\int_{\hat{c_B}}^{\bar{c_B}} f(z)dz = 1 - F(\hat{c_B})$, A's posterior beliefs after B opens negotiations are

$$F(c_B|open) = F(c_B|c_B \geq \hat{c_B}) = \int_{\hat{c_B}}^{c_B} \frac{f(z)}{1 - F(\hat{c_B})} dz.$$

*Stanford University, omastro@stanford.edu
†Duke University, david.siegel@duke.edu

From that fact, it follows that $F(c_B|open) = \frac{F(c_B)-F(\hat{c_B})}{1-F(\hat{c_B})}$, so $1 - F(c_B|open) = \frac{1-F(c_B)}{1-F(\hat{c_B})}$ and $f(c_B|open) = \frac{\partial}{\partial c_B}\int_{\hat{c_B}}^{c_B} \frac{f(z)}{1-F(\hat{c_B})}dz = \frac{f(c_B)}{1-F(\hat{c_B})}$. Together, that implies that hazard rates before and after B opens are identical for any value of the cutoff $\hat{c_B}$, though of course the cutoff is still informative about B's costs. In other words, the hazard rate satisfies this important property:

$$\frac{f(c_B|open)}{1-F(c_B|open)} = \frac{f(c_B)/1-F(\hat{c_B})}{(1-F(c_B))/(1-F(\hat{c_B}))} = \frac{f(c_B)}{1-F(c_B)}.$$

With that result in hand, we turn to A's optimization problem. A maximizes the Lagrangian $F(x-p|open)(p-c_A)+(1-F(x-p|open))x-\mu_1(x-1)+\mu_2(x-p-\hat{c_B})$ to yield Karush–Kuhn–Tucker conditions; μ_1 and μ_2 are multipliers. Either the first condition binds, the second condition binds, or neither binds. We consider each in turn.

- If the first condition binds, then $x^* = 1$ and $\mu_2 = 0$. $\mu_1 = f(1-p|open)(p-c_A) + (1 - F(1-p|open)) - f(1-p|open)$, and the requirement that μ_1 not be negative leads to condition (i): $\frac{f(1-p)}{1-F(1-p)} \leq \frac{1}{1-p+c_A}$.
- If the second condition binds, then $x^* = p + \hat{c_B}$ and $\mu_1 = 0$. $\mu_2 = -f(\hat{c_B}|open)(p-c_A) - (1-F(\hat{c_B}|open)) + f(\hat{c_B}|open)(p+\hat{c_B})$, and the requirement that μ_2 not be negative leads to condition (ii): $\frac{f(\hat{c_B})}{1-F(\hat{c_B})} \geq \frac{1}{c_A+\hat{c_B}}$.
- If neither condition binds, then $\mu_1 = \mu_2 = 0$ and $x^* = \hat{x}$, where $\hat{x}$ solves $f(\hat{x}-p|open)(p-c_A) + (1-F(\hat{x}-p|open)) - f(\hat{x}-p|open)\hat{x} = 0$. This value is $\hat{x} = p - c_A + \frac{1-F(\hat{x}-p)}{f(\hat{x}-p)}$.

Putting these together, we get that

$$x^* = \begin{cases} 1 & \text{if } \frac{f(1-p)}{1-F(1-p)} \leq \frac{1}{1-p+c_A}, \\ p + \hat{c_B} & \text{if } \frac{f(\hat{c_B})}{1-F(\hat{c_B})} \geq \frac{1}{c_A+\hat{c_B}}, \\ \hat{x} = p - c_A + \frac{1-F(\hat{x}-p)}{f(\hat{x}-p)} & \text{otherwise.} \end{cases}$$

Having found A's optimal offer given B has opened, we now must determine if there are conditions under which our proposed cutoff strategy is an equilibrium for B. We need to check two things. One, that the cutoff $\hat{c_B}$ type is indifferent between closed and open, since all types with higher cost than $\hat{c_B}$ will strictly prefer an open stance with bargaining, and all types with lower cost will strictly prefer a closed stance and war if this is true. Two, that the $\hat{c_B}$ for which this is true is between $\underline{c_B}$ and $\bar{c_B}$. When such a cutoff strategy is an equilibrium, we will also need to find its value(s) in equilibrium.

The payoff to the $\hat{c_B}$ type for a closed stance is its war payoff, or $1 - p - \hat{c_B}$. Its payoff for an open stance is $1 - x^* - \kappa$, so our condition of indifference is $1 - p - \hat{c_B} = 1 - x^* - \kappa$. We work through the three cases of x^*.

- First, if condition (i) holds, then $x^* = 1$ and the indifference condition becomes $1-p-\hat{c_B} = 1-1-\kappa$ or $\hat{c_B} = (1-p)+\kappa$. This would be a viable equilibrium as long as $(1-p)+\kappa \geq \underline{c_B}$, which is true by assumption, and $(1-p)+\kappa \leq \bar{c_B}$, which is never true by assumption. In other words, the only equilibrium that exists under condition (i) requires that at least one type of B would be willing to pay a cost to engage in a bargain in which it would gain nothing, just to avoid the high costs of war. Since we have assumed no such types are in the support of $F(c_B)$, there is no such equilibrium.

- Second, if condition (ii) holds, then $x^* = p + \hat{c_B}$ and the indifference condition becomes $1 - p - \hat{c_B} = 1 - \kappa - p - \hat{c_B}$, or $\kappa = 0$. Thus, an equilibrium in this case can only exist if opening negotiations is costless. Such an equilibrium is possible if $\frac{f(1-p)}{1-F(1-p)} > \frac{1}{1-p+c_A}$ and $\frac{f(\hat{c_B})}{1-F(\hat{c_B})} \geq \frac{1}{c_A+\hat{c_B}}$. All cutoffs that satisfy that second condition are viable; however, in equilibrium, the cutoff chosen by B will be the one that yields it the highest payoff, which is the lowest value of $\hat{c_B}$ that A would accept. That is the $\hat{c_B}$ that solves $\frac{f(\hat{c_B})}{1-F(\hat{c_B})} = \frac{1}{c_A+\hat{c_B}}$. As $f(c_B)$ is finite, $\frac{f(\hat{c_B})}{1-F(\hat{c_B})} \geq \frac{1}{c_A+\hat{c_B}}$ for $\hat{c_B} = \bar{c_B}$, and so at least one such solution exists. Further, as $f(\underline{c_B})(\underline{c_B} + c_A) < 1$, $\hat{c_B} > \underline{c_B}$ and no pooling equilibrium in which all types open negotiations exists. In fact, because f is both continuous and finite, sets of positive measure will choose both open and closed in equilibrium. Finally, rearranging the equality to $\hat{c_B} = \frac{1-F(\hat{c_B})}{f(\hat{c_B})} - c_A$, we see that since the LHS is increasing in $\hat{c_B}$ and the RHS is non-increasing in the same (it is the inverse hazard rate, and the hazard rate is non-decreasing), the solution is unique.

- Finally, if neither holds, then $x^* = \hat{x}$ and the indifference condition becomes $1 - p - \hat{c_B} = 1 - \hat{x} - \kappa$. Plugging in $\hat{x}$ yields $p + \hat{c_B} - \kappa = p - c_A + \frac{1-F(\hat{x}-p)}{f(\hat{x}-p)}$, or $\hat{c_B} = \kappa - c_A + \frac{1-F(\hat{x}-p)}{f(\hat{x}-p)} = \kappa - c_A + \frac{1-F(\hat{c_B}-\kappa)}{f(\hat{c_B}-\kappa)}$, with the last equality arising from $\hat{x} - p = \hat{c_B} - \kappa$. Since the LHS is increasing in $\hat{c_B}$ and the RHS is non-increasing in the same (it is the inverse hazard rate, and the hazard rate is non-decreasing), if there is a solution to this equation it is unique.

  The equilibrium lies in the interior when (ii) fails to hold, which happens when $\frac{f(\hat{c_B})}{1-F(\hat{c_B})} < \frac{1}{c_A+\hat{c_B}}$, which can be written as $\frac{1-F(\hat{c_B})}{f(\hat{c_B})} > c_A + \hat{c_B}$. The RHS is equal to $\kappa + \frac{1-F(\hat{c_B}-\kappa)}{f(\hat{c_B}-\kappa)}$, so that the condition holds whenever $\kappa < \frac{1-F(\hat{c_B})}{f(\hat{c_B})} - \frac{1-F(\hat{c_B}-\kappa)}{f(\hat{c_B}-\kappa)}$. As the inverse hazard rate is non-increasing, the RHS is never positive for all $\kappa \geq 0$, so this condition is never satisfied. Thus, we have a contradiction, and there is no interior equilibrium.

Taken together, we see that when $\kappa = 0$, there are conditions under which a unique equilibrium exists in which states with $c_B \geq \hat{c_B}$ open negotiations and all types with lower c_B do not. As long as A is unwilling to make an offer to B so high that every type of B would prefer it to fighting ($f(\underline{c_B})(\underline{c_B} + c_A) < 1$), there is no pooling equilibrium with all types of B opening negotiations, and all types of B that open negotiations accept the adverse inference arising from that action.

However, when $\kappa > 0$ there is no equilibrium in which B opens negotiations, save when at least one type of B would be willing to pay a cost to engage in a negotiation in which it would gain nothing, just to avoid the high costs of war. To see that the pooling equilibrium in which all types stay closed to negotiation is, in fact, an equilibrium, we must invoke the Iterated Intuitive Criterion equilibrium refinement (Fudenberg and Tirole 1991, 449) to reduce the set of possible off-the-path beliefs to better match our model's substantive setting. To do this, start by defining $J(open)$ as the set of types of state B that would obtain less than their equilibrium payoff of $1 - p - c_B$ by deviating from equilibrium and opening negotiations. The maximum payoff B can obtain in equilibrium is that which makes the lowest cost type indifferent to war minus the cost of opening, or $1 - p - \underline{c_B} - \kappa$; thus, for $\kappa > 0$ there will be some types that would always do worse by opening negotiations, and so $J(open)$ is a set of strictly positive measure. Let $c_{B1} = \sup J(open)$

be a higher cost than any other in $J(open)$ and $\Omega_1 = [\underline{c_B}, \bar{c_B}]\backslash J(open)$ be the set of types left over after removing $J(open)$. We can now iteratively apply the intuitive criterion, defining $J(\Omega_1)$ as the set of types in the set Ω_1 that would obtain less than their equilibrium payoffs by deviating to an open stance. As the maximum payoff any type of B in set Ω_1 can obtain in equilibrium is now $1 - p - c_{B1} - \kappa$, for $\kappa > 0$ there will always be some types that would do worse by opening negotiations, and so $J(\Omega_1)$ is a set of strictly positive measure. Defining each additional c_{Bn} and Ω_n similarly, because $[\underline{c_B}, \bar{c_B}]$ and each $J(\Omega_n)$ have finite measure, iterating will eventually yield an empty Ω_n. At that point, no types would benefit from deviating to open, and the equilibrium we posited exists and is the unique cutoff equilibrium.

To show that our cutoff equilibrium is unique, we compare our model's results to a model in which all types of B are automatically open to negotiations. We start by noting that A's maximization problem remains identical, save that in condition (ii) we must replace $\hat{c_B}$ with $\underline{c_B}$, and the assumption that $f(\underline{c_B})(\underline{c_B} + c_A) < 1$ ensures that condition (ii) would never hold. Next, note that because we have assumed that there is no type of B in the support of $F(c_B)$ that would prefer getting nothing in a negotiation to war, an offer of $x^* = 1$ would be rejected by all types, yielding the war payoff, $p - c_A$ for A. Since that payoff is always less than the payoff to A from the interior solution, A will never make that offer in equilibrium. That leaves the interior solution as the only option. Note that this matches Fearon (1995, pg 411) under our additional assumption on the types of B in the support of $F(c_B)$, as it must.

Any B type with $1 - p - c_B \leq 1 - \hat{x}$ will accept A's offer in the model of Fearon (1995); that works out to types $c_B \geq -c_A + \frac{1-F(\hat{x}-p)}{f(\hat{x}-p)}$. But the RHS of that is the optimal cutoff chosen in our model; thus, any type that would accept an offer in the absence of the opportunity to open negotiations would choose to open negotiations in our model, and any type that would reject an offer would not open negotiations in our model. Further, since $1 - p - \hat{c_B} = 1 - \hat{x}$ for the optimal cutoff, equilibrium payoffs for both A and B are identical for each type of B and for A in expectation under both our model and that in Fearon (1995). B's lexicographic preferences then imply that all types of B that would reject A's offer in the interior equilibrium of Fearon (1995) strictly prefer the unique cutoff equilibrium to one in which all types start open to negotiations.

As such, those types would deviate to closed in our model, and having all types choose open is not an equilibrium. That leads to Proposition 1 in the text.

Robustness of Proposition 1 to Changes in Bargaining Framework

Rather than have A make an ultimatum, we could limit bargaining only by the need for negotiated settlements to be incentive compatible and individually rational on the part of both states. To make our point, however, it is simpler merely to consider the opposite bargaining framework: a take-it-or-leave-it offer by B. The value of the game to B must be between those two ultimatums, so if our general conclusions hold for an ultimatum by B, they hold for all other bargaining frameworks as well. Considering the ultimatum by B yields the following proposition, which we prove in the next subsection.

Proposition 1A *Consider the alternative game in which B can make an ultimatum. If there exist in that ultimatum game cutoffs $\hat{c_B} \leq \bar{c_B}$ and $\kappa^* > 0$ such that, for any cost of opening $\kappa \leq \kappa^*$, higher c_B-cost types take an open stance, then there exists an equilibrium for at least some bargaining frameworks in which at least some types of B would take an open stance. Such cutoffs are more likely to obtain when the probability A wins is low, its cost of continued war is high, B's maximum cost of war is high, and B has more power to set the terms of a negotiated settlement.*

The intuition behind Proposition 1A stems from the concept of a bargaining surplus. In the basic adverse inference model, the type of B at the cutoff receives a surplus of zero and so has no reason to take an open stance and pay an additional cost of adverse inference. In contrast, the bargaining surplus when B can make an ultimatum is maximal. If that surplus exceeds the cost of adverse inference, the cutoff type will take an open stance.

Moving away from the optimal framework for B reduces its bargaining surplus. However, if the surplus were large enough in the optimal framework in which B can make an ultimatum, alternative frameworks sufficiently close to the optimal would produce sufficiently large surpluses so as to still satisfy the condition for an open negotiating stance. That would occur, for example, under high discount factors and low costs of opening negotiations in a simple alternating offer framework.

One could more precisely characterize the set of bargaining frameworks to extend Proposition 1A further, but that is beyond the scope of our model. Rather, the intent of Proposition 1A is to indicate that an open stance may be taken in equilibrium, albeit only for certain parameter values and when more bargaining power accrues to the state taking the open stance. That naturally leads to the question: how often do we observe the required parameter values and levels of bargaining power? Though we cannot answer that question at the dyadic level at this time, the aggregate empirical record does not suggest we often observe the prerequisites for an open stance under Proposition 1A. In only forty percent of interstate wars since WWII did talks emerge at some point during the conflict, and in only 34 percent of the cases did talks emerge because one of the belligerents proposed them.

As a consequence of those empirical regularities, it remains important to consider how adverse inference costs accrue, and how they might be reduced or eliminated.

Proof of Proposition 1A

We solve for the equilibrium of the model in which B makes a take-it-or-leave-it offer to A. Note that this offer is only made in equilibrium after B has taken an open stance, and we again, for the same reasons, look for a cutoff equilibrium of the same nature. Thus, in the bargaining phase, $c_B \in [\hat{c_B}, \bar{c_B}]$.

We again solve the model via backward induction. A accepts any offer for which $x \geq p - c_A$. x must be non-negative, but since $p > c_A$ by assumption, A's acceptance condition guarantees a non-negative x in any equilibrium in which an offer is made. We also need B to be willing to make an offer, which means that the lowest cost type that takes an open stance, $\hat{c_B}$, is at worst indifferent between having the offer accepted and having the offer rejected. That implies that $1 - x \geq 1 - p - \hat{c_B}$ or $x \leq \hat{c_B} + p$. Finally, it must be the case that $x \leq 1$. B chooses its offer to maximize $1 - x$ under the three aforementioned constraints, so that B maximizes $1 - x + \mu_1(x - p + c_A) - \mu_2(x - p - \hat{c_B}) - \mu_3(x - 1)$.

If the first constraint binds, then $x = p - c_A$ and $\mu_1 = 1 \geq 0$, so this is a possible equilibrium.

If the second condition binds, then $x = \hat{c_B} + p$, but also $\mu_2 = -1 < 0$, so this cannot be an equilibrium. Similarly, if the third condition binds, then $x = 1$ and $\mu_3 = -1 < 0$, so this too cannot be an equilibrium. Finally, if no condition binds, then $\mu_1 = \mu_2 = \mu_3 = 0$ and the derivative is decreasing in x. The minimum value of x allowable occurs at $x^* = p - c_A$, so this is the unique possible equilibrium bargain.

That implies, in any equilibrium, $1 - x^* = 1 - p + c_A$. B's initial decision is thus to open negotiations and receive $1 - p + c_A - \kappa$, or opt for continued war and receive $p - \hat{c_B}$. The $\hat{c_B}$ that makes the cutoff type indifferent is therefore $\hat{c_B} = 2p - 1 - c_A + \kappa$. If $2p - 1 - c_A + \kappa \leq \underline{c_B}$, then all types open negotiations. If $2p - 1 - c_A + \kappa > \bar{c_B}$, no types open negotiations. Otherwise, $\hat{c_B} = 2p - 1 - c_A + \kappa \in (\underline{c_B}, \bar{c_B}]$.

There exist types that open negotiations as long as $2p - 1 - c_A + \kappa \leq \bar{c_B}$, or $\kappa \leq \bar{c_B} + c_A + 1 - 2p$. Let $\kappa^* = \bar{c_B} + c_A + 1 - 2p$, so that for all $\kappa \leq \kappa^*$, $\kappa \leq \bar{c_B} + c_A + 1 - 2p$ holds. If $\kappa^* > 0$, then there exists a cutoff $\hat{c_B} \leq \bar{c_B}$ such that at least some set of types of B choose to take an open stance in equilibrium.

That proves that there is at least one bargaining framework in which B would take an open stance should those cutoffs exist: the framework in which B can make an ultimatum. Now consider an alternative bargaining framework. Any such framework will result in A's receiving an amount in equilibrium between what it receives when it makes the take-it-or-leave-it offer, and what it receives when B does. The same is true for B. Thus, we can classify all bargaining frameworks by the difference between the payoff in equilibrium they would provide to B, and that provided in the equilibrium in which B can make a take-it-or-leave-it offer. Let this difference be Δ_B. By the logic above, whenever $\kappa \leq \bar{c_B} + c_A + 1 - 2p - \Delta_B$, there exists an equilibrium in which at least some types of B open negotiations. That condition is easier to satisfy when $\bar{c_B}$ and c_A are both large, and p and Δ_B are both small, which completes the proof of the proposition.

Note that Proposition 1A does not specify that the set of bargaining frameworks is non-degenerate. That decision was made in order to avoid having to formally specify what constitutes a valid bargaining framework, since the bargaining framework in which B makes a take-it-or-leave-it offer to A suffices to comprise a set. That said, it is not difficult to construct an array of alternative bargaining frameworks that would also suffice. For example, for a sufficiently high shared discount factor δ, a bargaining framework in which B made the first offer, followed by A, followed by B, would yield the payoff $1 - \delta(1 - \delta(1 - p + c_A))$ to B, which approaches the take-it-or-leave-it payoff of $1 - p + c_A$ as $\delta \to 1$. Thus, for sufficiently high discount factors a range of alternating-offer bargaining frameworks will also admit equilibria involving the taking of an open stance.

Proof of Proposition 2

Our resilience model uses a two-dimensional type for B comprising costs of war, c_B, and resilience, $\rho \in \{L, H\}$, with the prior on the latter $Pr(\rho = H) = \int Pr(\rho = H|c_B)f(c_B)dc_B = q$. When $\rho = L$, A can increase its probability of victory to p' by escalating, but it cannot when $\rho = H$. We assume that $Pr(\rho = H|c_B)$ is strictly decreasing in c_B. As an open stance in our cutoff equilibrium removes the lowest values of c_B from the support of A's beliefs, A's belief about B's resilience satisfy $Pr(\rho = H|open) = \int_{\hat{c_B}}^{c_B} Pr(\rho = H|c_B)f(c_B|open)dc_B = \hat{q} < q$, when starting from A's prior belief regarding B's resilience type. In other words, if the only signal of B's resilience were B's decision to open, opening would induce A to reduce its belief that B was the high-resilience type. However, the resilience model includes two potential signals of B's resilience: an initial signal, and the choice of open or closed. A can choose to escalate after the choice of open or closed: escalation replaces the exogenous cost κ from the adverse inference model. We focus in this proof on the possibility that a low-resilience type will open in equilibrium.

The decision of high-resilience types of B to open or stay closed is unaffected by A's later escalation decision, since A's escalation does not change B's payoffs in either case. As a result, B's decision to open when $\rho = H$ is identical to that analyzed in our adverse inference model under $\kappa = 0$, save only that the cutoff B employs may differ from that given in Proposition 1: the cutoff depends on the inverse hazard rate, which depends on A's belief about the distribution of c_B, which depends on A's belief about B's resilience, which depends on B's initial signaling strategy. However, the same logic as in Proposition 1 guarantees a unique cutoff for any initial signaling strategy. Further, since both signals are costless for high-resilience types, they are indifferent between the two initial signals, regardless of A's responses to them.

The decision of open or closed for low-resilience types of B does depend on A's later escalation decision, so we must now consider that. There are four possible escalation outcomes relating to the open or closed decision for low-resilience types, which lead to different possible decisions according to Proposition 1.

1. B expects A not to escalate under either open or closed. In this case, B's decision of open or closed does not affect A's escalation decision, and there is no cost of adverse inference. This is the best achievable outcome for low-resilience types, and Proposition 1 implies that a unique cutoff equilibrium holds here. Note that the cutoff may differ from that in Proposition 1 for the reason given earlier. This is the only type of equilibrium that could occur in a version of the resilience model that lacked the initial signal.

2. B expects A to escalate after both open and closed. In this case, B's decision of open or closed does not affect A's escalation decision, and there is no cost of adverse inference: B's probability of winning declines to $1 - p'$ regardless of what it chooses, and the cutoff type of B remains indifferent. Proposition 1 implies that, in the absence of other preferable alternatives to which to deviate, a unique cutoff equilibrium holds here, replacing p with p', though again the cutoff may differ from that in Proposition 1.

3. B expects A to escalate only after closed. In this case, no low-resilience type of B would choose closed as it could always instead choose open and decline the offer if need be, which would yield it a higher payoff according to the proof of Proposition 1. Equilibria in which all low-resilience types open but some decline the offer are possible in this case.

4. B expects A to escalate only after open. In this case, the decreased probability of winning for B serves as a cost of adverse inference, and by Proposition 1 there is no equilibrium in which a low-resilience type of B would open. In the absence of other preferable alternatives to which to deviate, Proposition 1 would have all low-resilience types staying closed in

equilibrium; however, that equilibrium may not exist for the resilience model if the presence of high-resilience types allows for low-resilience types to productively deviate to open.

From those four escalation outcomes, we see that there are three ways in which low-resilience types would open in equilibrium. One, following some initial signal, it would be too costly for A to escalate given its beliefs. Two, following some initial signal, A would escalate after both open and closed stances, and escalation cannot be avoided. Three, following some initial signal, A would escalate after a closed stance only, leading all low-resilience types that could not avoid escalation to open, and some to decline the resulting offer. Limitation to those three ways in which low-resilience types would open forms the backbone of Proposition 2, but to show existence of at least one equilibrium in which each way occurs, we must consider A's beliefs, as they drive A's escalation decision.

Specifically, we are concerned with A's belief that it faces a high-resilience type, given that A would receive no benefit from escalating against it and escalation is costly for A. Let q' be A's belief that it faces a high-resilience type after both B's initial signal and B's decision of open or closed. A will escalate if its expected payoff for doing so exceeds its payoff for not doing so. As $\hat{c_B}$ does not depend on p and therefore not on A's escalation decision, A's gain for escalating depends only linearly on its probability of winning, p or p', and its belief that it faces a high-resilience type, and A escalates if $(p'-p)(1-q') > \nu$. That's true regardless of the signaling strategy in place. For example, in any equilibrium in which all types pool on an initial signal, we must have $q' = \hat{q} < q$ after opening on the equilibrium path. A would not escalate after closed on the equilibrium path as $q' > q$ after closed, but would escalate after open if $q'(p+\hat{c_B})+(1-q')(p'+\hat{c_B}))-(p+\hat{c_B}) > \nu$, which reduces to $(p'-p)(1-q') > \nu$.

A's belief that B is a high-resilience type depends on both B's initial signaling strategy and its subsequent choice of open or closed. As there are only four different combinations of initial signal and opening decision but a two-dimensional type space with one continuous dimension, there is an infinite number of potential signaling strategies different resilience-types of B could play, depending on their values of c_B. As we are only concerned with proving existence, however, we can limit our attention to a subset of possible strategies. We start by simplifying our notation a bit, denoting any c_B-type that would open according to Proposition 1 when $\kappa = 0$ an O-type, and any c_B-type that would not a C-type. We'll denote the four relevant type combinations as OL-, CL-, OH-, and CH-types. Note that an OL-type need not necessarily open in equilibrium if it were to face escalation from doing so; instead it could stay closed, as in Proposition 1 when $\kappa > 0$. Nor would a CL-type necessarily need to stay closed. Equilibria do exist in which OL- and CL-types with different values of c_B choose different strategies; however, as we can still prove existence while sticking to equilibria in which each OL- or CL-type chooses the same strategy as other OL- or CL-types, respectively, we will do so.

We begin proving existence of the relevant equilibria by eliminating some possible classes of equilibrium. First, we note that OH-types will always play open, and CH-types will always play closed, as they are unaffected by escalation. Second, if OH- and OL-types choose different initial signals, then OL-types cannot be the ones to choose S. If they were to do so, A would know they were low-resilience types with certainty after opening and, since A would also know with certainty that types opening after ∼S were high-resilience, an OL-type could always deviate to ∼S to avoid escalation. By similar logic, in any equilibrium either the CH- and CL-types must choose the same initial signal, or the CH-types must choose S.

Third, if OH- and OL-types choose the same initial signal, then A cannot escalate in equilibrium after observing that signal followed by open. The reason is that, by Proposition 1, no OL-type would open if opening would induce an escalation. However, if OL-types were not to open in equilibrium, then A would know with certainty that the only types opening in equilibrium after that initial signal were OH-types, A would not escalate, and OL-types would open, a contradiction that implies the lack of such an equilibrium.

Now we complete our existence proof by working through, in reverse order, each of the three possible ways for a low-resilience type to open in equilibrium. From our earlier discussion, the only way a CL-type would open is if A were to escalate only after closed and there were no better options for the CL-type to which to deviate. By assumption, A would not escalate if both CH- and CL-types played the same initial signal, since A's posterior belief in that case would be $q' > q$. That implies that for CL-types to open, CH-types must play S and CL-types must play ∼S.

For there to be no beneficial deviation, CL-types must prefer to open after $\sim$S instead of either staying closed after $\sim$S or staying closed after S. The latter deviation is not beneficial when the cost of the signal is too high, or $k \geq p' - p$. The former requires that A not escalate after open. To see when A would not escalate after open, given $k \geq p' - p$, consider that there are three viable alternatives for signaling strategies for OH- and OL-types: both OH- and OL-types also play $\sim$S, only OL-types do, or neither type does. If both types do, then A does not escalate after open if $(p' - p)(1 - \hat{q}) \leq \nu$ and this is an equilibrium. If only OL-types do, A escalates after $\sim$S and open. If neither type does, then A does not escalate only if $(p' - p)(1 - \hat{q}) \leq \nu$; however, even in that case OL-types would not be willing to play S when $k \geq p' - p$: depending on off-the-path beliefs after $\sim$S and open they could either deviate to $\sim$S to avoid escalation, or deviate to suffer escalation along with CL-types, but not pay the cost of the signal. Thus, there exists an equilibrium in which CL-types play open when $k \geq p' - p$ and $(p' - p)(1 - \hat{q}) \leq \nu$, with all types but CH-types playing $\sim$S. Low-resilience types with sufficiently small costs of war will decline the offer.

Again from our earlier discussion, the only way for an OL-type to open after an escalation is if A would also escalate after closed and there were no better options for the OL-type to which to deviate. OL-types that open after escalation must play $\sim$S, and OH-types must play S. To prevent OL-types from deviating, it must be the case that the cost of the signal is too high, or $k \geq p' - p$. To ensure that A would escalate after closed, we need only the CL-type to play $\sim$S, which also happens when the cost of the signal is too high. Thus, there exists an equilibrium in which OL-types open after A escalates when $k \geq p' - p$, with low-resilience types playing $\sim$S and high-resilience types playing S. The third and last way for a low-resilience type to open is if A would not escalate under either open or closed after some signal. There are two possibilities: the OH- and OL-types both play $\sim$S, or they both play S, as if the OL-type plays $\sim$S and the OH-type plays S, A would escalate after $\sim$S and open. We know if both OH- and OL-types play the same signal, then there is no equilibrium in which A escalates, so we need $(p' - p)(1 - \hat{q}) \leq \nu$. All types pooling on $\sim$S is an equilibrium under that condition: A does not escalate in equilibrium and all types receive their highest payoffs so they have no incentives to deviate. Thus, there exists an equilibrium in which OL-types open and A does not escalate when $(p' - p)(1 - \hat{q}) \leq \nu$, with all types playing $\sim$S. (There are many other equilibria that are similar. There are two other potentially viable equilibria in which OH- and OL-types both play $\sim$S. The first has both CH- and CL-types playing S. That is an equilibrium if the CL-type is willing to play S, which requires that $k < p' - p$ and off-the-path beliefs after $\sim$S and closed induce A to escalate. The second has CL-types playing $\sim$S and CH-types playing S; this is the previously derived equilibrium in which CL-types open and exists under the conditions given there. All types pooling on S is also an equilibrium when A doesn't escalate after open on the equilibrium path, as long as off-the-path beliefs induce escalation after $\sim$S regardless of the decision to open or close and the cost of the signal is low enough to justify paying it ($k < p' - p$). There are two other potentially viable equilibria in which the OH- and OL-types both play S. The first has both CH- and CL-types playing $\sim$S. That is an equilibrium if OL-types are willing to play S, which requires that $k < p' - p$ and off-the-path beliefs after $\sim$S and open induce A to escalate. The second has the CL-type playing $\sim$S and the CH-type playing S; this is not an equilibrium as whenever OL-types would be willing to pay the signaling cost, so would CL-types to avoid the escalation that occurs in equilibrium.)

This completes the proof of Proposition 2.