Dynamic Screening in International Crises

Noam Reich*

December 13th, 2023

Revised and Resubmitted to the *Journal of Politics*

Abstract

In this paper, I challenge costly signaling theory and present a rival theoretical framework for understanding how states behave and assess resolve in in crises. Modeling crisis diplomacy as a war of attrition, I ask how long countries will be willing to participate in costly negotiations and invest in sunk costs and audience costs before going to war or conceding. I show that more resolved states display greater impatience with diplomacy, preferring to fight instead. In turn, the least resolved states prefer to concede quickly to avoid having to fight. Finally, moderately resolved states negotiate longer, pay more sunk costs, and accumulate more audience costs. Consequently, moderately resolved states are more likely to obtain concessions, not because belief in their resolve increases, but because they grant their rival more time to concede. The model also features stalemated negotiations, providing new microfoundations for a common crisis outcome.

*I would like to thank Muhammet Bas, Ekrem Baser, Brendan Cooley, Haonan Dong, Matias Iaryczower, Kathy Ingram, Michael Kistner, Adam Meirowitz, Kris Ramsay, Jacob Shapiro, Jessica Sun, and three anonymous reviewers for their helpful comments. Additional thanks to the attendees of Peace Science 2019, SPSA 2020, APSA 2020, APSA 2021, Online Peace Science Colloqium 2020, and the summer 2020 Virtual Formal Theory Workshop.

For decades, costly signaling theory has been the dominant theoretical framework for explaining how states communicate resolve in international crises. This theory maintains that states can demonstrate their willingness to fight by taking costly actions, namely paying sunk costs or accumulating audience costs (Fearon 1994, 1997; Kurizaki 2007; Slantchev 2005). A state that engages in such behavior can convince its rival that it is not bluffing about its intentions to go to war and intimidate it into making a concession. For these reasons, costly signaling is thought to be an important tool of coercive diplomacy that enables resolved states to avoid war.

In this paper, I challenge costly signaling theory and present a rival theoretical framework in which states that invest more in sunk costs and audience costs can appear less resolved. The logic of the argument is straightforward. The more attractive an option war is to a state, the lower its incentive to try to avoid fighting. It follows that the unresolved states most desperate to avoid war are those most willing to invest in sunk costs, audience costs, or any other form of diplomacy that can secure a peaceful concession. In an international crisis where states have private information about their resolve, a state that chooses to spend time and effort on producing costly signals reveals that it believes diplomacy to be more profitable than immediate war. The more a state pursues a non-violent solution to a conflict, the more intimidated by war it must be.

Indeed, states' investments in costly signals are often overshadowed by their efforts to avoid war. For example, during the lengthy prelude to the First Gulf War, US allies remained concerned that the US would waver despite its massive military mobilization and efforts in organizing an international coalition. Though costly signaling theory predicts that these actions should have dispelled any doubt regarding US resolve, US allies remained wary of the lengthy delay and outright panicked when the US organized a summit with Iraq (Baker 1995, 346-353; Freedman and Karsh 1993, 240-243). More recently, the Obama administration failed to convince Israel that it would use force to prevent Iran from acquiring a nuclear weapon. As I argue below, the administration's determined pursuit of peace via open-ended negotiations undermined US assurances and diminished the commitment displayed by its economic sanctions campaign.

However, if resolved states do not engage in costly signaling, how do they behave in crises? Why do states invest in coercive diplomacy if it conveys hesitation? How do states learn about their rivals? To answer these questions, I develop a theory of dynamic screening that builds on the above arguments and offers a framework for explaining how states behave and assess resolve in international crises.

To construct the dynamic screening theory, I model an international crisis between two countries as a war of attrition, an open-ended affair that continues until either one of the two parties involved takes one of two unilateral actions. First, either of the two countries can choose to terminate the crisis at any time by conceding to its rival's demands. Alternatively, they can choose to escalate the crisis by starting a war, modeled as a costly lottery, and in doing so force their rival to go to war as well. I assume that countries have private information about their war payoff to fighting and do not know whether their rival prefers to fight or concede or how long they will prolong the crisis before acting. As a result, crises resemble a "war of the nerves" (Fearon 1994) with both countries claiming to be resolved and perpetuating the crisis in the hope that their rival is bluffing and will concede.

Crucially, participation in a crisis is not cost-free. First, I assume that countries must pay sunk costs for every second they choose to delay conceding or escalating the crisis. These include the costs of troop deployments, the economic costs of heightened tensions, and the opportunity costs to leaders for having to manage the crisis. Second, I assume that countries continuously accumulate audience costs that must be paid if a country concedes. How long a country is willing to prolong the crisis depends on its willingness to pay these costs. Dynamic screening refers to how the costs of participating in a crisis compel countries with different levels of resolve to exit the crisis at different times. A state is "screened" whenever its resolve is too low or too high to let it continue negotiating.

The equilibrium to the model is characterized by two dynamic screening processes. First, resolved states, defined here as states who would prefer to go to war than concede, are screened by sunk costs. This occurs because resolved states face a trade-off: though they would rather their rival concede than start a war, delaying war in the hope of a concession requires that they continuously pay sunk costs. Because war is inefficient, resolved states can benefit from granting their rival a short opportunity to concede peacefully. However, there is no guarantee that postponing a war will lead to a concession as countries cannot be sure whether their rival will ultimately concede. Resolved states have limited patience for paying sunk costs and, if enough time passes without a concession, will choose to fight. The higher a state's wartime payoff, the less worthwhile it is to wait for a concession, and the earlier it will to go to war.

The second dynamic screening process occurs when unresolved states, defined here as those states that would rather concede than fight, are screened by the risk of war. One implication of resolved states being screened by sunk costs is that states cannot know whether their rival will abruptly escalate the crisis and start a war until they actually do so. As a result, states that remain in the crisis long enough can have war thrust upon them even if they had been willing to concede peacefully. This presents unresolved states with their own trade-off. On the one hand, unresolved states can benefit from misrepresenting their type – instead of conceding, an unresolved state can perpetuate a crisis and pretend to be resolved, threatening to go to war if their rival does not concede in the hope that their rival is also unresolved and will choose to concede first. However, doing so requires that unresolved states risk having to fight. Since unresolved states with worse payoffs to fighting have more to lose if their rival suddenly decides to start a war, they will opt to concede earlier.[1]

Together, these dynamic screening processes generate three key results. First is a set of novel comparative statics relating a state's resolve to crisis outcomes. In particular, the model predicts that resolved states spend less time and effort negotiating, are more likely to go to

[1] Sunk costs and audience costs also make waiting less worthwhile for unresolved states. However, willingness to delay and pay these costs is independent of a state's wartime payoff conditional on having chosen to concede.

war, and less likely to obtain concessions. Conversely, moderately resolved types perpetuate crises the longest. In doing so they pay more sunk costs, accumulate more audience costs, and are more likely to obtain concessions. However, moderately resolved states are more likely to obtain concessions because they grant their rival more time to concede, not because their rival's belief in their resolve increases. These results stand in direct contradiction with costly signaling theory, which predicts that more resolved states invest more in sunk costs and audience costs and are more likely to obtain concessions because their rivals will view their actions as an indication of resolve.

Second, under dynamic screening states gradually learn about their rival's resolve from the length of delay. When a state allows a crisis to drag on, it reveals that it prefers to pay sunk costs rather than go to war. Simultaneously, a state that does not concede even after a lengthy crisis reveals that it is at least willing to risk fighting even if it does not want to initiate a war itself. As a result, states will come to view their rival as being moderately resolved over time. This contrasts with costly signaling theory's emphasis on large and discontinuous shifts in beliefs following ostentatious actions undertaken with the purpose of demonstrating resolve.

Third, dynamic screening theory provides a micro-foundation for stalemate outcomes in international crises. I show that there can exist an endogenous date in a crisis after which both states recognize that if their rival has not yet conceded or escalated the dispute, then they will never do so. In this case the crisis continues in perpetuity and neither state receives the good or issue under dispute. Such a stalemate occurs because the remaining types are both (1) unwilling to pay the audience costs required to concede and (2) lack the resolve to start a war. Moreover, a stalemate can occur even when states are penalized for failing to settle the dispute. To the best of my knowledge, this is the first theory to allow for endogenous stalemates as a crisis outcome.[2]

The theory in this paper builds on existing war of attrition models that also feature dy-

[2]Previous work has shown that audience costs can lock states into fighting, but not perpetual negotiations (Fearon 1994, 1997; Kurizaki 2007; Leventoğlu and Tarar 2009).

namic screening in both international relations and economics. However, my war of attrition model is designed to apply to a diplomatic crisis and so differs from these works in two key respects. First, following Fearon (1994) states have two exit options: concession or war. When states can escalate the dispute, remaining in the war of attrition for a lengthy period of time is perceived as a sign of hesitancy and therefore as irresolution. By contrast, most existing war of attrition models in the international conflict literature are designed to study ongoing interstate wars or civil wars which participants can only end by conceding (Nalebuff and Riley 1985, Slantchev 2003, Langlois and Langlois 2012, Powell 2017). In this case, the costs of remaining in the war of attrition are interpreted as the costs of fighting which screen low-quality types so that remaining in the war of attrition is interpreted as a sign of determination or strength. Second, in my model states that choose to utilize their outside option and go to war impose this outside option on their rival. This differs from wars of attrition in the economics literature, which are typically used to study firm's decision whether to leave a crowded market and exercise an outside option of switching to a different sector (Fudenberg and Tirole 1986, Takahashi 2015). Since exiting firms do not impose this outside option on competitors, firms with poor outside options do not preemptively concede and the dual screening result I achieve here is not present.

This paper is the first to demonstrate that resolved states can be screened by sunk costs. Within the literature on crisis diplomacy, Fearon's (1994) seminal article on audience costs is closest to this one. Modeling a crisis as a war of attrition, Fearon sought to demonstrate that audiences could commit states to conflict even when he imposed severe assumptions against fighting. To this end, Fearon assumed that states paid no sunk costs for delay, that no state had a positive payoff from fighting, and that crises end in finite time. Though I relax all three of these assumptions, only the first needs to be relaxed for resolved states to be screened by sunk costs. Intuitively if delay is free, resolved will prefer to wait until all types who wish to concede have done so before fighting.[3] Other scholars studying the

[3] The assumption that resolved states have negative payoffs for fighting need not prevent screening. However, screening becomes likely - the lower a state's payoff to fighting the more tolerant of delay it will be

two-exit war of attrition, where countries can either concede or escalate, have focused on other dynamics, introducing behavioral types or power fluctuations which do not produce the screening results observed here (Özyurt 2014, 2016; Kim 2018).

In the next section, I present the model setup. I then solve for equilibrium, demonstrating that countries' behavior is governed by two different dynamic screening processes. This is followed by a discussion of the results that compares dynamic screening theory and existing theory. Finally, I present two illustrative case studies of international crises that can be explained by dynamic screening theory and in which costly signaling theory falls short.

The Model

I model an international crisis between two countries as a war of attrition in continuous time. The countries seek to attain an indivisible good of value 1. Both countries will remain locked in the crisis until either one of them exercises one of two exit options: conceding or going to war.[4] If a country chooses to concede, then it surrenders the good to its rival and receives a payoff of 0. If either of the two countries chooses to escalate, then the countries fight and the game ends in a costly lottery (Fearon 1995). Let p_i and $1 - p_i$ be the probability that country i and j ($j \neq i$) respectively win the fight and receive the good. Let c_i ($i = 1, 2$) denote a country's resolve, the cost that each country pays for fighting regardless of the lottery outcome. Each country's cost for fighting is private information and is selected by a random draw from a common knowledge distribution $c_i \sim C_i$ with continuous and strictly positive density over its support $[\underline{c}_i, \bar{c}_i]$. To simplify matters, let each country i's payoff to fighting be denoted with $w_i = p_i - c_i$ and the transformed cumulative distribution F_i have a support over $[\underline{w}_i, \overline{w}_i]$. I assume that $\overline{w}_i \in (0, 1)$ and that $\underline{w}_i \in (-1, 0)$ such that there always

and if payoffs for fighting are sufficiently low, then it is possible for all unresolved states to concede before either states becomes impatient with diplomacy. The online appendix demonstrates this point by numerically simulating the model. The assumption that wars of attrition occur end in finite time is unnecessary as sunk costs ensure that resolved states will exit endogenously. Moreover, relaxing this assumption allows for a richer analysis that can accomodate the occurence of stalemates.

[4] To avoid confusion between "war of attrition," the class of model, and "going to war," an exit strategy in the model, I will substitute the term "crisis" for "war of attrition" whenever possible.

exist both types with a positive and negative expected utility for fighting and no type prefers fighting to obtaining a concession.[5]

Following Fearon (1994), each country accumulates audience costs that must paid if that country concedes. Such costs are designed to capture punishments imposed by domestic audiences on leaders who fail to follow through on a threat that they have made. I assume that these costs accrue at a linear rate a_i so that if country i chooses to concede at time t, she pays $a_i t$ audience costs for doing so. This parameterization of audience costs implicitly assumes that countries pay no audience costs for conceding the conflict immediately at time $t = 0$. This reflects a belief that countries which concede "quietly, without a public contest" incur no penalties from their domestic audiences (Fearon 1994, 585). After this point, the model assumes that domestic audiences punish leaders more severely for conceding after lengthier crises.

In addition, each country must also pay a sunk cost k_i for every moment that they remain in the crisis. These sunk costs represent any and all expenses that might arise in a diplomatic dispute that must be paid regardless of the outcome of the conflict, such as the costs of mobilizing troops. I assume that countries pay no sunk costs at the start of the crisis so that a country who exits at $t = 0$ incurs no sunk costs. Together, these assumptions imply that a country that remains the crisis until time t incurs sunk costs $k_i t$.

Since I am interested in studying stalemate outcomes where both countries remain locked in the crisis forever, I must consider what happens if neither country goes to war or concedes. In such a case, I assume that countries cease to pay sunk costs and incur a one-time penalty $K_i > 0$. Substantively, this term represents the costs that arise when countries fail to settle a diplomatic dispute, such as market actors viewing investments in the country as being riskier or a permanent increase in troop deployments. Mathematically, the assumption that countries treat the costs of a stalemate as a one-time penalty, as opposed to paying sunk costs in perpetuity, is useful because it bounds payoffs. This is a necessary condition for

[5] This contrasts with Fearon (1994) who assumed that $\overline{w}_i = 0$.

stalemates. The online appendix explores an alternative solution in which states discount future payoffs and pay sunk costs in perpetuity and achieves similar results.

A strategy for country i is a function mapping country i's type to its choice of exit time and choice of exit strategy. The former will be denoted with a choice t_i in the set $\mathbb{R}_+ \cup \{\infty\}$ where a choice of $t_i = \infty$ represents a choice not to exit. Exit options will be denoted with $\theta_i \in \{0, 1\}$ where $\theta_i = 0$ represents a choice to concede and $\theta_i = 1$ represents a choice to go to war. Formally country i's strategy is denoted σ_i and defined as $\sigma_i : [\underline{w}_i, \overline{w}_i] \to \mathbb{R}_+ \cup \{\infty\} \times \{0, 1\}$. It will often be useful to work with the inverse image of the strategy function that maps exit times and exit choice to a country's type $\tau_i : \mathbb{R}_+ \cup \{\infty\} \times \{0, 1\} \to [\underline{w}_i, \overline{w}_i]$. Abusing notation, let $\sigma_i(w_i|\theta)$ denote the strategy function mapping type to exit time for types choosing exit option θ. Given these strategies, country i's expected utility function can be written as

$$\begin{aligned} U_i(t_i, \theta_i, \sigma_j | w_i) = & \int_{\{w_j | t_j < t_i, \theta_j = 0\}} [f_j(w)(1 - k_i \sigma(w|0))] dw + \int_{\{w_j | t_j < t_i, \theta_j = 1\}} [f_j(w)(w_i - k_i \sigma(w|1))] dw \\ & + \mathbb{1}_{\{t_i \neq \infty, \theta_i = 0\}} \left[\int_{\{w_j | t_j = t_i, \theta = 0\}} \left[f(w) \left(\frac{1}{2}(1 - a_i t_i) - k_i t_i \right) \right] dw + \right. \\ & \left. \int_{\{w_j | t_j = t_i, \theta = 1\}} \left[f(w) \left(\frac{1}{2}(w_i - a_i t_i) - k_i t_i \right) \right] dw - \int_{\{w_j | t_j > t_i)\}} [f_j(w)(a_i t_i - k_i t_i)] dw \right] \\ & + \mathbb{1}_{\{t_i \neq \infty, \theta_i = 1\}} \left[\int_{\{w_j | t_j = t_i, \theta = 0\}} \left[f(w) \left(\frac{1}{2}(1 + w_i) - k_i t_i \right) \right] dw \right. \\ & \left. + \int_{\{w_j | (t_j = t_i, \theta = 1) \vee (t_j > t_i)\}} [f(w)(w_i - k_i t_i)] dw \right] - \mathbb{1}_{\{t_i = \infty\}} \int_{\{w_j | t_j = \infty)\}} [f(w)(K_i + k_i \overline{T})] dw \end{aligned} \quad (1)$$

where the first line represents the payoff if country j exits before country i, the second and third lines represent the payoffs country i can expect from conceding at time t_i, the fourth line and first term on the fifth line are the payoffs country i can expect from going to war at time t_i, and the last term is the payoff for remaining locked in the crisis forever.

Throughout the paper I solve for Perfect Bayesian Equilibria. This requires that each country update its beliefs using Bayes' Rule whenever possible and maximize their expected utility in light of these beliefs. Throughout the crisis, each country will be able to update its beliefs at every instant after learning that its opponent has not yet chosen to exit. Let $g_i(w_j|t)$ denote country i's posterior beliefs that country j has wartime payoff w_j after observing country j remain in the crisis up until time t. Each country must continue to find its choice of exit time and strategy optimal as it updates these posterior beliefs. An equilibrium is therefore a pair (σ_i^*, g_i) for each country.[6]

Characterizing the Equilibrium

In this section I demonstrate that the game has an equilibrium featuring three distinct phases during which countries play different strategies. I characterize behavior in each phase and define the conditions which must be met for the countries to transition between them. These phases must occur in a strict sequence for virtually any possible set of parameters.[7] The game always begins with a peaceful phase during which no country goes to war. If countries have types that are sufficiently resolved, then sunk costs cause the game to transition to one of two different screening phases. First, a screening phase where only one country gradually goes to war and then a second screening phase where both countries gradually go to war. Unresolved types will concede throughout the three phases, though the strategy by which they do so will change from phase to phase. All equilibria must take this sequential form.[8]

The game may end during any one of the three phases. Specifically, the game ends when at least one country no longer has any types remaining who wish to concede. The assumption that states accumulate a strictly increasing quantity of audience costs ensures that such a date must exist. Intuitively, audience costs must eventually grow so large such that there is a date by which no type would ever choose to concede. Following Fearon (1994), I refer to this time as the horizon date and label it $\overline{T}$. Once a country no longer has any

[6] A formal definition of a Perfect Bayesian Equilibrium is provided in the online appendix.

[7] The online appendix contains a formal proof for the argument that the sequence of phases is unique.

[8] The online appendix contains a formal proof for the argument that the war of attrition must have delay occur with positive probability and that the sequence of phases must be unique.

types remaining who wish to concede, its rival can no longer justify paying the sunk costs required to delay exiting, thereby triggering an end to strategic behavior. This leads us to the following Lemma (all proofs provided in the online appendix):

Lemma 1

In any equilibrium there exists a finite time $\overline{T}$ after which no type exits the crisis.

Strategic behavior can end in one of three ways. First, both countries may run out of types willing to concede, and all remaining types can go to war. Second, under certain conditions, only one country may run out of types willing to concede. In response its rival immediately exits, either by going to war or concede at time $\overline{T}$. Due to space constraints, I restrict attention in the main text to the first of these two possibilities. A full characterization of the equilibria that accounts for this second possibility is provided in the online appendix.

Third, it is possible for the countries to remain in the crisis forever. This requires that $K_i < a_i(\overline{T})$ so that there exist types who would prefer to pay the sunk costs required to sustain the crisis forever rather than concede. I will refer to such an outcome as a stalemate because the dispute remains unsettled; neither party is sufficiently resolved to go to war and audience costs prevent both countries from conceding. Stalemates require that both countries have their unresolved types finish conceding by $\overline{T}$. In turn, any resolved types still participating in the crisis to go to war at $\overline{T}$ and the only types remaining past $\overline{T}$ are those who prefer to remain in the crisis forever. In the remainder of this section, I describe countries' behavior during each of the three phases in sequence and then at the horizon date.

Characterizing the Peaceful Phase

The game begins with a peaceful phase during which neither country goes to war. Since war is costly, resolved types prefer that their rival concede peacefully and will start by granting them the opportunity to do so. Knowing that their rival might abruptly choose to declare war at a later time, the least resolved types will choose to concede during the peaceful phase.

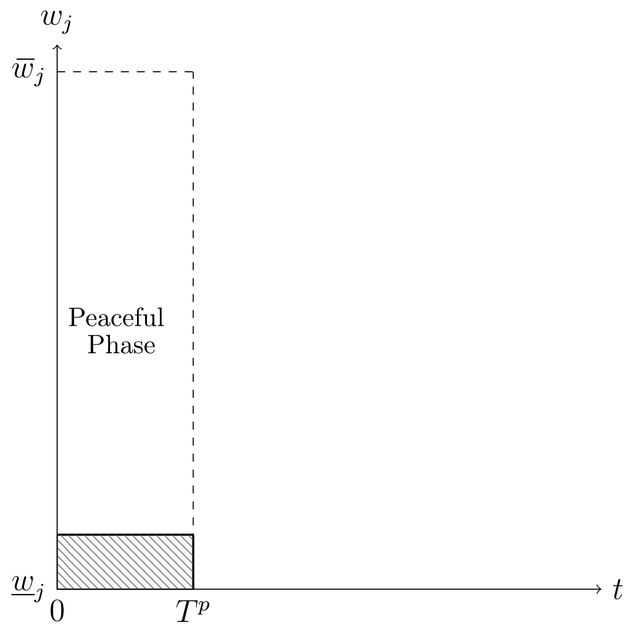

Figure 1: **The Peaceful Phase:** This figure maps the time at which different types exit the crisis, with the y-axis measuring the possible war payoffs and the x-axis measuring the time at which they exit. Crises will always begin with a peaceful phase during which no type goes to war. This is represented by the dashed line at the top of the figure indicating that type $\overline{w}_j$ does not exit. Knowing that war may occur in the future, the least resolved types of each country will concede during the peaceful phase by playing a mixed strategy. This is represented by the patterned box, indicating that the least resolved types of country j concede at a random time in that interval.

However, these least resolved types can benefit from delaying their concession if their rival is similarly unresolved and concedes first. As a result, types who concede during the peaceful phase will do so via a mixed strategy, delaying their concession for a random length of time in the hope of outlasting the rival state. Such delay is costly and both countries accumulate sunk costs and audience costs while waiting for their rival to concede. Though resolved types would prefer their rival concede peacefully, their willingness to pay sunk costs while waiting for their rival to concede is limited. When the most resolved type of both countries is no longer willing to incur sunk costs, they will go to war and cause the game to transition to the first screening phase. Figure 1 illustrates these strategies.

Concession Behavior During the Peaceful Phase

I begin by characterizing the strategies for types that concede during the peaceful phase. At the start of the dispute one of the two countries may concede with positive probability. If that country does not concede, then a crisis begins and the countries start to accrue sunk costs and audience costs. From that point on either one of the two countries may concede at any time. Though the probability of a concession at any *particular* time t is vanishingly small, the probability that either country has conceded *by* a particular time t is strictly increasing and does not cease to increase until the end of the peaceful phase.

Formally, let $Q_i(t)$ be the cumulative distribution function describing the probability that types of country i who have chosen to concede during the peaceful phase do so by time t. Moreover, let T^p denote the date at which the peaceful phase transitions into a screening phase. Lemma 2 establishes some useful properties of $Q_i(t)$.

Lemma 2

Let $T^1 = \min\{T^p, \overline{T}\}$. $Q_i(t)$ must satisfy the following properties in an equilibrium in which both countries finish conceding by the horizon date: (i) $Q_i(t)$ must be continuous and strictly increasing; (ii) $Q_i(t) < 1$ if and only if $t < \min\{T^p, \overline{T}\}$; and (iii) $Q_i(0)Q_j(0) = 0$.

The following Proposition characterizes $Q_i(t)$ explicitly.

Proposition 1

Let $T^1 = \min\{T^p, \overline{T}\}$. In any equilibrium in which both countries finish conceding by the horizon date, types $w_i \in [\underline{w}_i, \beta_i^p]$ $(i = 1, 2)$ concede on the interval $[0, T^1]$ according to the following strategy

$$\frac{q_j(t)F_j(\beta_j^p)}{1 - F_j(\beta_j^p)Q_j(t)} = \frac{a_i + k_i}{1 + a_i t} \tag{2}$$

Since no type goes to war during the peaceful phase, both countries beliefs are given by

$$g_i(w_j|t) = \begin{cases} \frac{f_j(w_j)[1-Q_j(t_i)]}{1-Q_j(t)F_j(\beta_j^p)} & \text{if } w_j \in [\underline{w}_j^t, \beta_j^p] \\ \frac{f_j(w_j)}{1-Q_j(t)F_j(\beta_j^p)} & \text{if } w_j \in [\beta_j^p, \overline{w}_j] \end{cases} \tag{3}$$

The intuition underlying Proposition 1 is straightforward. Since country i is mixing, it has to be indifferent as to when it concedes during the peaceful phase. This requires that the marginal benefits of delaying concession at any given moment are equal to the marginal costs of doing so. Proposition 1 characterizes these quantities. The left-hand side of equation (2) is country j's hazard rate, representing the probability that country j will concede if country i decides to wait a moment longer. The right-hand side of equation (2) represents the weighted marginal costs to waiting, i.e. the additional sunk and audience costs that would have to paid for conceding later divided by the difference in payoffs to having country j concede as opposed to country i conceding. Together Lemma 1 and equation (2) form a comprehensive strategy for types conceding during the peaceful stage. Finally, over the course of the peaceful phase each country continuously reduces their belief that their rival comes from a subset of the types with the lowest payoffs to fighting - because these types concede with positive probability over the course of the the peaceful, the longer a state holds out the less they are unresolved.

How Long will the Peaceful Phase Last?

The peaceful phase ends whenever there is a type that is no longer willing to delay going to war. Since types $\overline{w}_i$ $(i = 1, 2)$ have the highest payoffs for fighting, they will be the types of country i and country j that will prefer to go to war earliest. However, types $\overline{w}_i$ and $\overline{w}_j$ may prefer to go to war at different times. The peaceful phase will end whenever the first of these two types decides to go to war or upon arrival at the horizon date, whichever comes first.

Proposition 2 provides a formal characterization for T^p, the length of the peaceful phase that transitions into a first screening phase. Let T_i^p $(i = 1, 2)$ denote the amount of time that type $\overline{w}_i$ is willing to wait before going to war if its rival will play according to equation (2) up until that time. The following proposition provides a formal characterization for these quantities.

Proposition 2

During the peaceful phase, type $\overline{w}_i$ $(i = 1, 2)$ will choose to go to war at time

$$T_i^p = [1 - \overline{w}_i]\left[\frac{1}{k_i} + \frac{1}{a_i}\right] - \frac{1}{a_i} \tag{4}$$

Let $T^p = \min\{T_1^p, T_2^p\}$ $(i \neq j)$. If $T^p < \overline{T}$, then the game transitions to the first screening phase and country $i = \min\{T_1^p, T_2^p\}$ has type $\overline{w}_i$ go to war at time T^p. Otherwise, the game proceeds directly to the horizon date at $\overline{T}$.

The following is the intuition underlying Proposition 2. Country i will seek to wait before going to war until the marginal benefit of doing so is equal to the marginal cost. This will be achieved whenever

$$\frac{F_j(\beta_j^p)q_j(t_i)}{1 - Q_j(t_i)F_j(\beta_j^p)} = \frac{k_i}{1 - \overline{w}_i} \tag{5}$$

The left-hand side of this equation is the probability that country j concedes if country i $(i \neq j)$ at time t and represents the marginal benefit of delaying the choice to go to war at that time. The right-hand side of the equation is the marginal cost, representing the sunk costs that are paid for delaying another moment weighted by the difference in payoffs between having country j concede and having country i go to war. Equation (4) is found by substituting in for the hazard rate (2) into (5).

Equation (4) also reveals three additional facts about the peaceful phase. First, T_i^p is strictly decreasing in $\overline{w}_i$, implying that as the upper bound of a countries' resolve increases, the peaceful phase becomes shorter. Second, it is each country's most resolved type $\overline{w}_i$ that will satisfy the equation at the earliest time. Third, the equation demonstrates that it is the presence of audience costs and the fact that they are strictly increasing that generates the peaceful phase.[9] This is because audience costs make delay more costly for unresolved types thereby requiring that their rival concede at a faster rate to keep the unresolved types indifferent as to when they concede. It is this accelerated rate of concession which makes

[9]To see this multiply both sides of equation (4) by a_i and then set $a_i = 0$.

delaying war worthwhile for resolved types.

Characterizing the First Screening Phase

Once type $\overline{w}_i$ or $\overline{w}_j$ chooses to go to war, the peaceful phase ends and the game transitions into the first screening phase.[10] Without loss of generality, let country j be the country who's most resolved type preferred to go to war first ($T_j^p < T_i^p$). During this phase, sunk costs screen resolved types of country j - at any given time, country j's most resolved type still participating in crisis the crisis will go to war because it is no longer willing to pay sunk costs. In turn, the threat of war screens country i's unresolved types - at any given time, country i's least resolved type still participating in the crisis will concede because it wants to avoid the risk that its rival will abruptly declare war. However, resolved types of country i are still willing to give their rival an opportunity to concede peacefully and do not go to war in this phase. Without the threat of war from their opponent, unresolved types of country j continue to concede by playing a mixed strategy. When the most resolved type of country i is willing to incur sunk costs no longer, they will escalate and cause the game to transition to the second screening phase. Figure 2 illustrates these equilibrium strategies.

The Switch to Pure Strategies

I begin by characterizing the properties which must be satisfied by the various exit strategies. Once again, though the probability that country j goes to war, or that country i or j concedes, *at* any particular moment is vanishingly small, the probability that they do so *by* a particular moment is strictly and continuously increasing right up until the very end of the first screening phase. Consequently, resolved types of country j can choose precisely how much sunk costs they are willing to pay in exchange for a probability of a concession by delaying war. Resolved types with higher costs of fighting have more to lose from going to

[10]If the most resolved types of both countries both want to transition from the peaceful phase at at the same time ($T_i^p = T_j^p, i \neq j$), then the game transitions directly to the second screening phase. However, this is highly unlikely given the multidimensional and continuous parameter space.

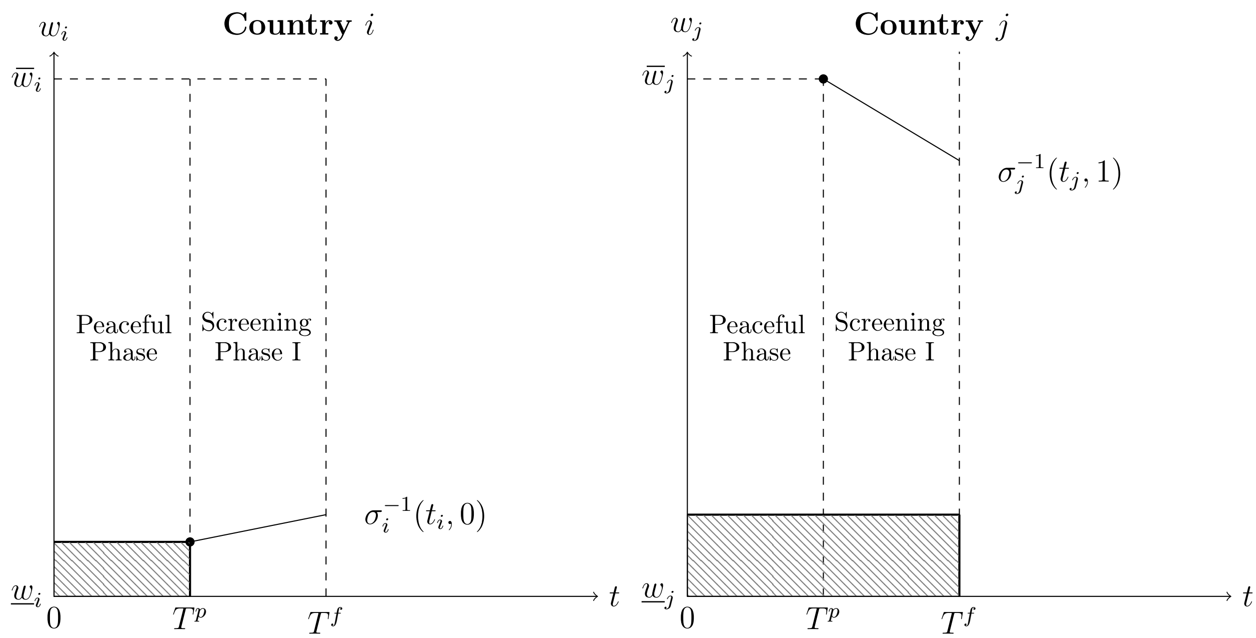

Figure 2: **The First Screening Phase:** The peaceful phase is followed by the first screening phase during which only one country has types that go to war. Beginning at time T^p, sunk costs begin to screen resolved types of country j. This is represented by the upper curve in the right sub-figure depicting the time at which resolved types of country j choose to go to war. In turn, the threat of war screens unresolved types of country i who switch from playing a mixed strategy to a pure strategy. This is represented by the lower curve in the left sub-figure depicting the time at which unresolved types of country i concede. Resolved types of country i remain peaceful during the first screening phase as depicted by the extension of the dashed line at the top of the left sub-figure into the first screening phase. Absent the threat of war, unresolved types of country j continue to concede via mixed strategy as depicted by a similar extension of the patterned box at the bottom of the right sub-figure.

war and will be willing to remain in the crisis longer. Similarly, unresolved types of country i can choose exactly how much risk they are willing to take by remaining in the war of attrition by choosing to delay until a particular time. Unresolved types of country i with higher payoffs to fighting will be willing to incur a little more risk and therefore delay longer.

Formally, let $S_j(t)$ denote the mixed strategy adopted by conceding types of country j. Let T^f, defined more precisely below, denote the date at which the first screening phase transitions into the second screening phase. Lemma 3 lays out the key characteristics that govern the dual screening process.

Lemma 3

If there exists a $T^p < \overline{T}$, then (i) both $S_j(t)$ and $\sigma_i(\cdot|0)$ must be continuous and strictly increasing (ii); $S_j(t) < 1$ if and only if $t < \min\{T^f, \overline{T}\}$; (iii) $S_j(T^p) = 0$; (iv) $\sigma_j(\cdot|1)$ must be continuous strictly decreasing on $[T^p, \min\{T^f, \overline{T}\}]$.

Proposition 3 characterizes the strategies of unresolved types of countries i and j and resolved types of country j that choose to exit during the first screening phase. Let β_i^f and β_j^f denote the lowest cost-of-fighting type of each country to concede during the first screening phase. Let $\underline{w}_i^t$ denote the least resolved type of country i yet to concede at time t. Proposition 3 provides a formal characterization of exit strategies for types exiting during the first screening phase.

Proposition 3

Let $T^2 = \min\{T^f, \overline{T}\}$. If there exists a $T^p < \overline{T}$, then during $[T^p, T^2]$, country i concedes by playing $\tau_i(\cdot, 0)$ as given by

$$\frac{f_i(\tau_i(t,0))\tau_i'(t,0)}{1 - F_i(\tau_i(t,0))} = \frac{a_j + k_j}{1 + a_j t} \tag{6}$$

Types $w_j \in [\beta_j^p, \beta_j^f]$ and resolved types of country j exit by playing

$$\frac{[F_j(\beta_j^f) - F_j(\beta_j^p)]s_j(t)}{F_j(\tau_j(t,1)) - [F_j(\beta_j^f) - F_j(\beta_j^p)]S_j(t) - F_j(\beta_j^p)} = \frac{a_i + k_i}{1 + a_i t}$$
$$+\frac{f_j(\tau_j(t,1))\tau_j'(t,1)}{F_j(\tau_j(t,1)) - [F_j(\beta_j^f) - F_j(\beta_j^p)]S_j(t) - F_j(\beta_j^p)} \times \frac{\underline{w}_i^t + a_i t}{1 + a_i t} \tag{7}$$

$$\sigma_j(w_j|1) = [1 - w_j]\left[\frac{1}{k_j} + \frac{1}{a_j}\right] - \frac{1}{a_j} \tag{8}$$

Each country's posterior beliefs posterior beliefs during this period are given by

$$g_i(w_j|t) = \begin{cases} \frac{f_j(w_j)[1-S_j(t)]}{F_j(\tau_j(t_i,1))-[F_j(\beta_j^f)-F_j(\beta_j^p)]S_j(t)-F_j(\beta_j^p)} & \textit{if } w_j \in [\beta_j^p, \beta_j^f] \\ \frac{f_j(w_j)}{F_j(\tau_j(t_i,1))-[F_j(\beta_j^f)-F_j(\beta_j^p)]S_j(t)-F_j(\beta_j^p)} & \textit{if } w_j \in [\beta_j^f, \overline{w}_j^t] \\ 0 & \textit{otherwise} \end{cases} \tag{9}$$

$$g_j(w_i|t) = \begin{cases} \frac{f_i(w_i)}{1-F_i(\tau_i(t,0))} & \textit{if } w_i \in [\underline{w}_i^t, \overline{w}_i] \\ 0 & \textit{otherwise} \end{cases} \tag{10}$$

The following is the intuition underlying the result. First, recall that without the threat of war, unresolved types of country j must be indifferent as to when they concede. From the discussion in the previous section, we know that this requires that unresolved types of country i concede at the rate given in equation (2). However, Lemma 3 requires that country i play a pure strategy with unresolved types of country i that have higher costs of fighting conceding earlier. Therefore, (2) is rewritten as (6). Surprisingly, this implies that the threat of war does not change the rate at which country i concedes, only that its unresolved types now do so via a pure strategy. Second, because the rate at which country i concedes does not change when the game transitions to the first screening phase, the trade-offs affecting how long a resolved type of country j should wait before going to war are identical to those faced by type $\overline{w}_i$ in the previous section. Therefore, equation (8) is analogous to (4).

Finally, because unresolved types of country j that concede during the first screening phase are playing a mixed strategy and are indifferent as to when they concede, their strategy only needs to ensure that unresolved types of country i find $\tau_i(t_i, 0)$ as given in equation (6) optimal. Unresolved types of country i will choose to delay their concession until the marginal costs of doing so are equal to the marginal benefits. The proof of the proposition shows that Country i's expected utility function is maximized by waiting until (7) is satisfied. Equation (7) is similar to equation (2) with the addition of a new term, an additional marginal cost to waiting that accounts for the possibility that delay will lead to war. Perhaps contrary to intuition, this new term is positive, implying that country j's rate of concession increases when compared to the peaceful phase. This increase is necessary to compensate country i for the increased risk of war now involved in delay.

These strategies determine how countries posterior beliefs change over the course of the first screening phase. First, country i continuously lowers its belief for country j's highest possible level of resolve. Because the most resolved types of country j are screened by sunk costs, i can eliminate the possibility of j being a type that should have already gone to war. Second, country j can similarly continuously increase its belief regarding the lowest possible level of country i's resolve - the fact that country i has not exited demonstrates that it has not yet been screened by the risk of war. Finally, as in the peaceful phase, country i continuously reduces its belief that country j is from the subset of the least resolved types as the crisis continues and country j does not concede.

How Long Will the First Screening Phase Last?

Eventually, resolved types of country i will grow tired of paying sunk costs while waiting for country j to concede. As before, type $\overline{w}_i$ will be the type who wants go to war at the earliest date because it has the lowest cost of fighting. The time at which type $\overline{w}_i$ goes to war is T^f and it marks the time at which the game transitions to the second screening phase.

The following Proposition characterizes the length of first screening phase.

Proposition 4

Type $\overline{w}_i$ will go to war whenever the following condition is satisfied

$$\overline{w}_i = 1 - \frac{k_i[1 + a_i t]}{k_i + a_i + [\underline{w}_i^t + a_i t]\frac{f_j(\tau_j(t_i,1))\tau_j'(t_i,1)}{F_j(\tau_j(t_i,1)) - (F_j(\beta_j^f) - F_j(\beta_j^p))S_j(t) - F_j(\beta_j^p)}} \tag{11}$$

at which point the game transitions to the second phase.

The intuition underlying Proposition 4 is similar to that underlying Proposition 2. Type $\overline{w}_i$ will delay going to war until the marginal costs of doing so are equal to the marginal benefits. This expected utility function will be maximized at time T^f when equation (11) holds. This equation is similar to (4) with the addition of a new term in the denominator. Whereas unresolved types want to avoid escalation, resolved types are willing to wait longer before going to war if there is some chance that their rival is going to initiate a war anyway. This additional benefit to waiting implies that type $\overline{w}_i$ delays their exit time relative to when they would have exited if it had been the case that $T_i^p < T_j^p$ and country i would have initiated the transition to the first screening phase. The date at which equation (11) holds is the date at which the game transitions from the first to second screening phases, providing us with a formal definition of T^f.

Characterizing the Second Screening Phase

Once the most resolved type of country i decides to go to war, the game transitions to the second screening phase.[11] In this phase, sunk costs screen both countries resolved types, causing the most resolved type of each country still participating in the crisis at any particular moment to go to war to avoid continuing to pay sunk costs. This generates a risk of war that similarly screens both countries' unresolved types, causing the least resolved type of each country still participating in the crisis at any particular moment to concede. Both of these screening processes continue until the horizon date, when all types who intended to concede

[11]Or alternatively, if $T_i^p = T_j^p$ the game proceeds directly to this phase instead of the first screening phase.

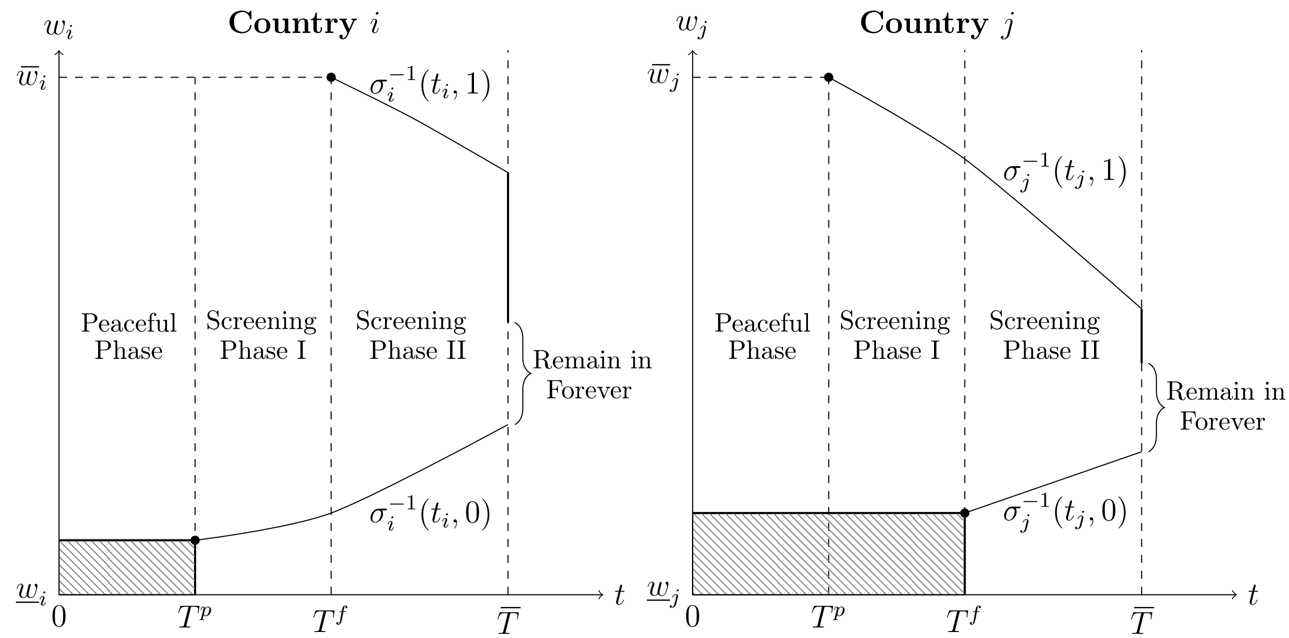

Figure 3: **The Second Screening Phase and the Horizon Date:** At time T^f sunk costs begin screening resolved types of country i causing them to gradually go to war. This is depicted by the upper curve in the left sub-figure. In turn, the risk of war that this generates begins to screen unresolved types of country j as depicted by the lower curve in the right-sub figure. Resolved types of country j and unresolved types of country i continue to be screened as they were in the first screening phase. Both countries continue being screened until the horizon date $\overline{T}$ when the last unresolved type of both countries concedes. At this point remaining types either go to war immediately or remain in crisis forever. Though the figure depicts the horizon date at the end of the second screening phase, it is possible for it to arrive earlier such that countries never arrive at either the first or second screening phases.

have done so. Figure 3 illustrates these results.

As in previous phases, screening involves strategies wherein countries gradually exit the crisis. Specifically, countries play strategies such that the timing of a war or concession is effectively zero at any particular moment, but that the probability of a war or a concession is strictly and continuously increasing right up until the horizon date. The following lemma establishes this result and dynamic screening's monotonicty properties:

Lemma 4

If there exists a $T^f < \overline{T}$*, then during* $[T^f, \overline{T}]$ $\sigma_i(\cdot|0)$ *is continuous and strictly increasing and* $\sigma_i(\cdot|1)$ *is continuous and strictly decreasing.*

The following Proposition characterizes the strategies for those types exiting during the second screening phase.

Proposition 5

If there exists a $T^f < \overline{T}$, then types of country i who exit during $[T^f, \overline{T}]$ play according to the strategy $\tau_i(\cdot, \theta)$ $(i = 1, 2)$ as defined by

$$\frac{f_j(\tau_j(t,0))\tau_j'(t,0)}{F_j(\tau_j(t,1)) - F_j(\tau_j(t,0))} = \frac{k_i}{1 - \overline{w}_i^t} \tag{12}$$

$$\frac{f_j(\tau_j(t,1))\tau_j'(t,1)}{F_j(\tau_j(t,1)) - F_j(\tau_j(t,0))} = \frac{k_i[\overline{w}_i^t + a_i t] - a_i[1 - \overline{w}_i^t]}{[1 - \overline{w}_i^t][\underline{w}_i^t + a_i t]} \tag{13}$$

and country i's $(i = 1, 2)$ posterior beliefs during $[T^f, \overline{T}]$ are given by

$$g_i(w_j|t) = \begin{cases} \frac{f_j(w_j)}{F_j(\tau_j(t,1)) - F_j(\tau_j(t,0))} & \text{if } w_j \in [\underline{w}_j^t, \overline{w}_j^t] \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

The intuition for this result is similar to that of Proposition 3. Resolved and unresolved types of both countries will remain in the war of attrition until the marginal costs of doing exceed the marginal costs. For resolved types, this means delaying until equation (12) is satisfied which is analogous in interpretation to equation (5). For unresolved types, this means delaying concession until equation (13) is satisfied.

The strategies stated in the proposition also determine how countries will update their posterior beliefs during the second screening phase. Now that unresolved and resolved types of each country are being screened, both countries will continuously increase their belief regarding their rival's lowest possible level of resolve and continuously decrease their belief regarding their rival's highest possible resolve. As before, this is because a country that remains in the crisis demonstrates that it is sufficiently resolved to have incurred the risk of war so far, but also lacks the resolve to have yet initiated a war.

Characterizing the Horizon Date

The game ends once all unresolved types have conceded. At this point, any type of either country that intends to go to war can no longer benefit from delay. As a result, a mass of types, possibly all those who remain, will go to war at the horizon date. However, it is also possible that some types still participating in the crisis may opt for a stalemate outcome and sustain the crisis in perpetuity. This latter outcome is illustrated in Figure 3.

Let β_i $(i = 1, 2)$ denote the lowest cost-of-fighting type to concede of country i. The following proposition summarizes countries' behaviour throughout the crisis and at the horizon date.

Proposition 6

Strategic behaviour ends at $\overline{T}$, which can arrive during any phase. In an equilibrium where both countries finish conceding by the horizon date, countries' choice of exit strategy and their behavior at the horizon date is determined by the following:

(i) All types exit: There exists an equilibrium where types $w_i \in [\underline{w}_i, \beta_i]$ concede and types $w_i \in (\beta_i, \overline{w}_i]$ go to war where $\beta_i = -a_i\overline{T}$. Any type still participating in the crisis at $\overline{T}$ goes to war at that time.

(ii) Some types remain in forever: If $K_i < a_i\overline{T}$ for both $i = 1, 2$, then there exists an equilibrium where types $w_i \in [\underline{w}_i, \beta_i]$ concede for β_i as given by

$$\frac{F_j(\overline{w}_j^{\overline{T}}) - F_j(-\overline{K}_j)}{F_j(\overline{w}_j^{\overline{T}}) - F_j(\beta_j)}\beta_i - \frac{F_j(-\overline{K}_j) - F_j(\beta_j)}{F_j(\overline{w}_j^{\overline{T}}) - F_j(\beta_j)}K_i = -a_i\overline{T} \tag{15}$$

Types $w_i \in (\beta_i, -K_i]$ remain in the crisis forever and types $w_i \in (-K_i, \overline{w}_i]$ go to war. Any type from the latter set still participating in the crisis at $\overline{T}$ go to war at that time.

The explanation for the result is as follows. If all types remaining at $\overline{T}$ choose to go to war, then type β_i must be indifferent between going to war or conceding and paying the audience costs required to concede at the horizon date. It follows that any type with a higher cost of fighting than type β_i must have already conceded by the horizon date and that any type with a lower cost of fighting than β_i will go to war at $\overline{T}$ if it has not already done so. Alternatively, the game can end in a stalemate whenever sunk costs are sufficiently low $K_i < a_i\overline{T}$ for both $i = 1, 2$. In this case type $w_i = -K_i$ is indifferent between fighting and sustaining the crisis forever. This implies that any type with a higher payoff to fighting than $-K_i$ strictly prefers to fight at $\overline{T}$ if they have not already exited the war of attrition. Type β_i is then defined by equation (15) as the type that is indifferent between paying the audience costs accumulated by the horizon date and opting for a stalemate while risking war with all the types that choose to fight at the horizon date. It follows that any type less resolved than β_i must have already conceded by the horizon date and that any type in $(\beta_i, -K_i]$ will choose to remain in the crisis forever.

Note that Proposition 6 implies that the costs of a stalemate being sufficiently low ($K_i < a_i\overline{T}$ for both $i = 1, 2$) are a necessary but insufficient condition for a stalemate outcome. Stalemates require that both countries choose not to fight at $\overline{T}$. If both countries' strategies require them to fight at the horizon date, then no single country has the ability to prevent a war by unilateral deviation from $\sigma(w_i) = \{\overline{T}, 1\}$ to $\sigma(w_i) = \{\infty, \theta\}$. [12]

Discussion

The war of attrition model offers a parsimonious and widely applicable framework with which to study the dynamics of diplomacy. True to the anarchic nature of the international system, the model imposes little structure on the countries' interactions. The crisis has no exogenously imposed end date. Nor can states commit to taking any future action. Instead, states are free to go to war or concede at any time. Moreover, the model is capable of incorporating standard aspects from the costly signaling literature, including sunk costs and

[12] This is why countries went to war at the horizon date in Fearon's (1994) model even though no type had a positive expected utility for fighting ($\overline{w}_i = 0$). However, without sunk costs ($K_i = k_i = 0$), Fearon's model could also support an outcome where both countries remain in the war of attrition forever. The potential for a stalemate is not recognized or discussed in the article.

audience costs, into a dynamic setting.

By contrast, costly signaling *models* are defined by a sender who can take a costly action to attempt to convey information to a receiver. For analytical convenience, signaling models have a discrete order of moves imposed on the players and typically only afford the sender one opportunity to take costly action. This is a weakness of costly signaling models as it requires that senders be able to instantaneously take large costly actions.[13] Costly signaling *theory* refers to a set of results arising from analyses of such models which demonstrate that higher quality types can communicate their willingness to fight by taking costly actions (e.g. Fearon 1997, Slantchev 2005).

Costly signaling models need not support the results of costly signaling theory. The ability of higher quality types to engage in costly signaling is sensitive to the source of uncertainty and type of signal considered in a signaling model (Arena 2013, Caroll and Pond 2021, Reich 2023). The online appendix demonstrates this claim and shows that even if one of the two countries were offered the opportunity to engage in sunk cost signaling prior to the war of attrition beginning, then resolved types would fail to distinguish themselves and a war of attrition with positive probability for delay must still occur. This is because resolved types preference to invest less in sunk costs is not driven by the game form, but by their weaker incentive to spend to avoid war. However, this is not to say that costly signaling signaling is never possible or that costly signaling theory cannot work under certain assumptions. Therefore, it is worth highlighting four general results from the model and how they compare to existing theory.[14]

The first general result is that more resolved states prefer to invest less in diplomacy and go to war earlier. Specifically, the model shows that more resolved states spend less time negotiating, pay less sunk costs, accumulate less audience costs, and are more likely to go to war. These findings represent a a major departure from costly signaling theory, which

[13]Alternatively, senders could produce the signals over a longer period during which any of the receiver's interim actions are unimpactful such that they may as well have been instantaneously produced.

[14]For brevity's sake, the introduction treats the first two as a single result.

maintains that more resolved states should invest more in sunk costs and audience costs and be more likely to achieve peaceful outcomes (Fearon 1994, Fearon 1997, Slantchev 2005, Reich 2022).

The second general result is that unresolved states with higher costs of fighting will choose to concede earlier to avoid the risk of war. Specifically, the model shows that unresolved types with higher costs of fighting spend less time negotiationg, pay less sunk costs, and accumulate less audience costs, are less likely to obtain concessions themselves, and are less likely to go to war.[15] At first glance, this result would seem to support findings in the costly signaling literature, wherein less resolved or weaker types invest less in costly signaling. However, the mechanism that drives this behavior differs across the two theories. In costly signaling theory unresolved states don't signal because they are deterred by the price of sunk costs and audience costs, even when these lower the risk of war. By contrast, in dynamic screening theory unresolved states concede early to avoid the risk of war, not because they are deterred by paying sunk costs or audience costs.[16]

Taken together, these two general results imply that moderately resolved states spend the most time negotiating. This comes across clearly in Figure 3. Accordingly, moderately resolved states pay more sunk costs, accumulate more audience costs, and are more likely to obtain concessions. This increased probability of obtaining a concession occurs because they grant their rival more time to concede not because states come to believe their rival is more resolved as they invest more in sunk costs and audience costs.

Indeed, the third general result concerns the gradual convergence in each countries posterior beliefs towards their being a moderately resolved type. Once sunk costs begin to screen

[15]Note that this relationship is not completely monotonic because Lemmas 2 and 3 require that countries play a mixed strategy absent the threat of war. Conditional on conceding during the peaceful phase or the first screening phase, there need not be a relationship between a country's cost of fighting and its exit time.

[16]Dynamic screening theory more closely resembles the literature on brinkmanship, which argues that states can demonstrate resolve by generating an *exogenous* risk of war (Schelling 1960). Formal models of brinkmanship showed that when remaining in a crisis required states to withstand the probabilistic risk of war, states with higher costs of fighting concede earlier (Powell 1988). This is similar to the method by which unresolved states are screened in my model. However, brinkmanship views remaining in a crisis and incurring risk as the primary method by which states demonstrate resolve. By contrast, in my model the risk of war is generated endogenously by resolved states who are tired of diplomacy and prefer to fight.

resolved types, countries conclude that their rival must lack sufficient resolve to have yet started a war and respond by decreasing their belief in the highest level of their rival's possible resolve. Similarly, each country concludes that their rival is at least sufficiently resolved to have not yet conceded, and responds with a continuous reduction in its belief that its rival is unresolved.[17]

Additionally, this slow and steady rate of learning also implies that countries use the length of delay – how long their rival has chosen to prolong the crisis - as their primary metric for assessing resolve. Formally, the propositions demonstrate that the length of delay is a sufficient statistic for a state's posterior distribution of their rival's possible resolve. This means that, conditional on knowing a rival's strategy, the length of delay contains all the information contained within the model necessary for a state to form beliefs about its rival - once a state knows how long its rival has negotiated, it can infer the amount of sunk costs it has paid, the amount of audience costs it has accumulated, and the risk of war it has incurred such that these quantities provide no additional information.[18] The gradual nature of learning under dynamic screening contrasts sharply with costly signaling theory in which large discontinuous shifts in beliefs are thought to occur following countries undertaking dramatic actions, e.g. following a sudden deployment of troops.

A fourth general result is the model's ability to incorporate stalemates as an endogenous outcome of crises. After the horizon date, countries recongize that they will be locked in crisis in perpetuity as it becomes common knowledge that neither country is sufficiently resolved to start a war and that the crisis has gone on long enough for audience costs to have grown too large for either state to concede. When a stalemate occurs neither state receives the good under dispute and both are assumed to continue to pay some penalty for failing to settle the dispute. Proposition 6 demonstrates that so long as this penalty is not too large,

[17]The sole exception is at time $t = 0$ when one country can have a mass of types concede. This implies that if a country meets the challenge of a crisis head on and does not concede immediately, there can be a discontinuous change in beliefs.

[18]Note that this does not mean that there cannot be additional sources of information used by states to form beliefs that are not included in the model.

then a stalemate is possible.

An examination of the data on militarized interstate disputes reveals that such stalemates are incredibly common (Palmer et al 2015). Fully 1,479 out of 2,143, or 69 percent, of MIDs that are settled peacefully are coded as ending in a stalemate outcome, i.e. as not having "any decisive changes in the pre-dispute status quo and [are] identified when the outcome does not favor either side in the dispute" (MIDs Dispute Coding Manual). Moreover, 1,413 of these disputes are coded as not having any negotiated outcome, so that "none of the pre-conditions that fueled the conflict are resolved nor is there any agreement between the parties that the dispute should be terminated." This suggests that stalemates as defined by the model are the modal crisis outcome.

Dynamic Screening in Practice

Both dynamic screening and costly signaling theories describe how states manage uncertainty in international crises. Because both theories incorporate uncertainty, sunk costs, and audience costs as essential components of crisis behavior they allow for direct comparisons of their predictions. This section examines the performance of the two theories in two high-profile cases: the First Gulf War and the Iranian Nuclear Crisis. I demonstrate that in both cases the US's decision to delay war and pursue diplomacy detracted from its large investments in sunk costs or audience costs and caused it to be perceived as unresolved. Together these cases illustrate the shortcomings of costly signaling theory and the explanatory power of dynamic screening theory.

Negotiating the JCPOA

When President Obama acceded to the presidency he decided to reach out to Iran and attempt to negotiate a peaceful resolution to the ongoing crisis over its nuclear program. In the fall of 2009, the P5+1 and Iran held a number of summits until negotiations collapsed when Iran rejected the P5+1's offer for a fuel swap.[19] Iran's rejection of this "confidence-building measure" coupled with the secrecy and nature of its nuclear program convinced the P5+1 that Iran was negotiating in bad faith and in June 2010 the UN Security Council sanctioned Iran. In the 15 months that followed, the US and Europe organized and imposed additional sanctions while Iran continued to develop its nuclear program without further negotiation. Starting in April 2012, negotiations resumed and the P5+1 and Iran held several meetings that produced little progress. Ultimately, a surprise change in Iranian leadership in 2013 jump-started diplomacy. Five months after the election of Iranian President Rouhani, the P5+1 and Iran reached the "Joint Action Plan" agreement and managed to avoid war.

Underlying these negotiations was the US threat to use force to prevent Iran from acquiring nuclear weapons should negotiations fail. Per costly signalling theory, a rational observer of the US should have believed the US threat to be credible. President Obama invested a significant amount of sunk costs in the crisis. His administration spent a great deal of time and political capital on coordinating the passage of sanctions in the UN Security Council and advocating for sanctions more broadly. According to Deputy National Security Advisor Ben Rhodes, sanctions against Iran were a top priority for every meeting Obama held with a foreign leader in 2011 (Parsi 2017, 120). Furthermore, Obama would have likely faced a great deal of audience costs if he would have conceded to unrestricted or unmonitored Iranian nuclear enrichment. In both public and private Obama promised that he would use military means to prevent Iran from acquiring a nuclear weapon, stating that "as President of the United States, I don't bluff" (Goldberg 2012; Parsi 2012, 77). Congress also repeatedly exerted pressure on the White House by threatening or passing sanctions bills at the height of sensitive negotiations (Parsi 2012, 73, 108-111, 132-133, 157-161; Parsi 2017 148).

However, US allies did not perceive it to be resolved. Specifically, Israel pressured the US to consider military options to terminate Iran's nuclear program and threatened to attack

[19]By this point Iran had collected a sufficient quantity of low-enriched uranium to produce one nuclear weapon. This deal would have required Iran to ship the lion's share of its stockpile out of the country in exchange for fuel pads for the Tehran Research Reactor which produced medical isotopes. This would have, in principle, provided more time for negotiations and reduced the threat of war (Parsi 2012, 114-116).

Iran on its own. Though Israel preferred that the US be the one to attack Iran and the US sought to avoid an Israeli strike, US attempts at reassurance failed to convince Israel that the US would attack should negotiations fail.[20] Though Israel did not ultimately strike Iran, this was not because of its belief in US resolve.[21]

Dynamic screening theory, and in particular the length of US delay can help explain Israel's lack of confidence in the US. Though the US managed to orchestrate a tough multilateral sanctions regime it refused to commit to a timeline for military action, turning down repeated Israeli requests for a deadline for sanctions or negotiations.[22] According to Gary Samore, the White House Coordinator for Arms Control and Weapons of Mass Destruction, the US recognized that if it were ever to admit that negotiations had failed, then they would be forced either to attack Iran or to concede to an unrestricted and unmonitored nuclear program (Parsi 2017, 117). As a result, US policy was to sustain diplomacy and sanctions, even when these offered no sensible path towards a resolution of the dispute.[23] Dynamic screening theory maintains that this delay conveyed hesitation thereby undermining Obama's reassurances that he would be willing to resort to military means to reign in the Iranian nuclear program.

Gulf War

Within days of Iraq's invasion of Kuwait (August 2nd, 1990), the United States mounted a tough response, beginning the process of deploying tens, and eventually hundreds, of thou-

[20]Israel preferred that the US be the one to attack Iran, if necessary, because of its superior capacity to damage Iranian's nuclear program and set back Iran's ability to acquire the bomb for longer (Parsi 2012, 28-30; Parsi 2017, 151; Barak 2018, 433). The US believed that an Israeli strike would undermine diplomacy, weaken the sanctions regime, and could compel the US to go to war anyway (Parsi 2017, 152-154).

[21]In November 2010, Israeli Prime Minister Netanyahu, the Minister of Defense Ehud Barak, and the Minister of Foreign Affairs Avigdor Liberman supported a strike and held a meeting with the heads of Israel's security organizations to order immediate preparations for one (Barak 2018, 426-427; TOI Staff 2012). However, the plan was opposed by the heads of Israel's security organizations who insisted on a vote in the security cabinet, a group of ministers authorized to approve acts of war. These organization heads opposed a strike and in doing so undermined Netanyahu's ability to secure a majority for a strike, thereby thwarting it (Netanyahu 2011, 477-488; TOI Staff 2015).

[22]See for example Parsi (2012, 50-51, 74-78, 165-169), Parsi (2017, 154-156)

[23]For example, in the summer of 2012, when the US agreed to schedule additional diplomatic meetings solely to keep negotiations alive and deny Israel the political cover for a strike (Parsi 2017, 148).

sands of troops to Saudi Arabia to deter further Iraqi aggression. Simultaneously, the US spearheaded an international coalition that passed numerous UN Security Council Resolutions condemning the invasion, imposing severe economic sanctions, and authorizing the use of force to enforce these sanctions. In an address to the nation, President Bush made clear that the goal was the "immediate, unconditional, and complete withdrawal" of Iraq from Kuwait (Freedman and Karsh 1993, 93). When it became evident that sanctions would not compel Iraq to leave Kuwait, the US successfully advocated for a UN Security Council Resolution that authorized the use of force if Iraq would not leave of its own volition.

Per costly signalling theory, these actions should have communicated strong American resolve on this issue. Bush put the US's reputation on the line and accrued audience costs by repeatedly stating that Iraq would have to leave Kuwait without any preconditions and that he would use force to compel it to do so if necessary.[24] Moreover, the Bush administration incurred a great deal of sunk costs in addressing the crisis. Beyond the time and effort invested in organizing an international coalition, the US deployed a massive force to the Gulf. According to Chairman of the Joint Chiefs of Staff Colin Powell, the US Deployment would be large enough to allow the US to "win decisively" and ensure that it would never be "operating in the margins" (Freedman and Karsh 1993, 207-208). Costly signalling theory would therefore predict that Iraq and US allies should have adjusted their beliefs and taken these actions as evidence that the US was willing to use force if its demands were not met.

However, despite these actions, key actors in the conflict continued to doubt US resolve. On November 30th, the day after the UN Security Council authorized the use of force to expel Iraq from Kuwait, Bush announced that, though he was "not hopeful," he would reach out to Iraq and attempt to go the "extra mile for peace" (Freedman and Karsh 1993, 235). The explicit purpose of this policy was to shore up US domestic support for the war by showing that every effort had been made to attain peace. However, upon learning of Bush's diplomatic initiative, US allies, influential pundits, and even members of the administration

[24]In addition to the UN Security Council resolutions and many private assurances given to world leaders, Bush made many public statements (Bush and Scowcroft 1999, 340-341, 345, 350, 368, 370-371, 388).

began to suspect that the US "didn't really want to use force" and was "desperately searching for an escape route" (Baker 1995, 346-353). Saudi Ambassador Bandar told National Security Advisor Scowcroft that sending Secretary of State Baker to meet with Iraqi officials would "suggest [to Saddam] you're chicken" (Freedman and Karsh 1993, 241). Though Bush made clear that Iraq's unconditional withdrawal from Kuwait was not on the table, onlookers grew concerned that the US would seek a compromise or allow negotiations to extend beyond the deadlines imposed by the UN Security Council.

Dynamic screening theory can explain why why US allies were so concerned by Bush's attempt to negotiate with Iraq, highlighting diplomacy and delay as a sign of irresolution. Moreover, it can also explain why the allies disregarded the US's sunk cost investment in the conflict and the audience costs it accumulated. Neither was this episode the first instance wherein state's behavior accorded with dynamic screening and was inconsistent with costly signaling theory. For example, as early as October, Prime Minister Thatcher was already pressuring Bush to go to war as soon as US forces arrived in sufficient number so as to avoid looking unresolved (Bush and Scowcroft 1999, 385). When informed about the US decision to double the number of troops in the theatre, Thatcher was not impressed by the US commitment as costly signalling theory would predict. Instead she grew concerned about the delay required for the troops to arrive and the potential that the Americans would "wobble" during that time (Freedman and Karsh 1993, 209, 228). Finally, it should be noted that US allies remained wary of US resolve, correctly assessing the Bush's reluctance to go to war. For example, while Saudi Arabia thought it important that Saddam not be allowed to withdraw from Kuwait unpunished and with his army intact, they recognized that Bush would have gladly allowed him to do so (Baker 1995, 352). [25]

[25] Though Scowcroft shared in the opinion that an Iraqi withdrawal at this late stage could be a disaster (Bush and Scowcroft 1999, 437-438), according to Baker, "war was the last thing" Bush wanted. Instead "all [Bush] really wanted was to get Iraq out of Kuwait" and would have refrained from going to war if Iraq withdrew as demanded (Baker 1995, 349).

Conclusion

In this article, I developed a theory of dynamic screening in international crises and argued that more resolved states should invest less in diplomacy. To do so, I modeled a crisis as a war of attrition in which states decide how long to pursue diplomatic options while incurring sunk costs, audience costs, and potentially risking war. The model is characterized by two different dynamic screening processes. First, more resolved states will go to war earlier, preferring their assuredly high payoff to fighting over paying sunk costs to prolong a crisis in the hope that their rival concedes. Second, the threat of war posed by this potential breakdown in diplomacy causes the least resolved states to concede earlier to avoid the risk of having war thrust upon them. These dynamic screening processes make the length of delay an important source of information about state's resolve.

This article is but a first step in the study of dynamic screening in international crises. First, this paper focused on the effects of audience costs and sunk costs as these are the two most prevalent signaling costs in the literature. Other potential processes might be capable of reshaping the screening dynamics. For example, it is possible that democratic institutions might constrain leaders and prohibit them from going to war before a threshold of sufficient negotiations has been reached. I leave it to future research to explore such dynamics. Second, there is room to incorporate bargaining into the war of attrition, as in Langlois and Langlois (2012), so as to study its impact on dynamic screening. Recent work in mechanism design and conflict has studied the properties of different bargaining protocols, demonstrating, for example, that ultimatums deliver proposers their best distribution of outcomes (Fey and Ramsay 2011, Fey and Kenkel 2021). In light of these dynamic screening results, the properties of the war of attrition as a bargaining protocol merit more attention. Finally, dynamic screening is the first theory capable of explaining the variation in the length of crises, generating novel comparative statics and predictions for which states are likely to go to war or concede and when. As such it deserves future empirical study.

References

Arena, Philip (2013). "Costly Signaling, Resolve, and Martial Effectiveness". *Unpublished Manuscript.*

Baker, James Addison (1995). *The Politics of Diplomacy : Revolution, War, and Peace, 1989-1992*. Ed. by Thomas M. DeFrank. New York: G.P. Putnam's Sons.

Barak, Ehud (2018). *My Country, My Life : Fighting for Israel, Searching for Peace*. New York: St. Martin's Press.

Bush, George and Brent Scowcroft (1999). *A World Transformed*. New York: Vintage Books.

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

Fearon, James D. (1994). "Domestic Political Audiences and the Escalation of International Disputes". *American Political Science Review* 88.3, pp. 577–592.

— (1995). "Rationalist explanations for war". *International Organization* 49.3, pp. 379–414.

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

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

Fey, Mark and Kristopher W. Ramsay (2011). "Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict". *American Journal of Political Science* 55.1, pp. 149–169.

Freedman, Lawrence and Efraim Karsh (1993). *The Gulf Conflict, 1990-1991 : Diplomacy and War in the New World Order*. Princeton, N.J.: Princeton University Press.

Fudenberg, Drew and Jean Tirole (1986). "A Theory of Exit in Duopoly". *Econometrica* 54.4, pp. 943–960.

Goldberg, Jeffrey (Mar. 2012). "Obama to Iran and Israel: 'As President of the United States, I Don't Bluff'". *The Atlantic.*

Kim, Jin Yeub (2018). "Counterthreat of Attack to Deter Aggression". *Economics Letters* 167, pp. 112–114.

Kurizaki, Shuhei (2007). "Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy". *American Political Science Review* 101.3, pp. 543–558.

Langlois, Jean-Pierre P. and Catherine C. Langlois (2012). "Does the Principle of Convergence Really Hold? War, Uncertainty and the Failure of Bargaining". *British Journal of Political Science* 42.3, pp. 511–536.

Nalebuff, Barry and John Riley (1985). "Asymmetric equilibria in the war of attrition". *Journal of Theoretical Biology* 113.3, pp. 517–527.

Netanyahu, Benjamin (2022). *Bibi: My Story.* Threshold Editions.

Özyurt, Selçuk (2014). "Audience Costs and Reputation in Crisis Bargaining". *Games and Economic Behavior* 88, pp. 250–259.

— (2016). "Building Reputation in a War of Attrition Game: Hawkish or Dovish Stance?" *The B.E. Journal of Theoretical Economics* 16.2, pp. 797–816.

Palmer, Glenn et al. (2015). "The Mid4 Dataset, 2002–2010: Procedures, Coding Rules and Description". *Conflict Management and Peace Science* 32.2, pp. 222–242.

Parsi, Trita (2012). *A Single Roll of the Dice : Obama's Diplomacy with Iran.* New Haven: Yale University Press.

— (2017). *Losing an Enemy : Obama, Iran, and the Triumph of Diplomacy.* New Haven: Yale University Press.

Powell, Robert (1988). "Nuclear Brinkmanship with Two-Sided Incomplete Information". *American Political Science Review* 82.1, pp. 155–178.

— (2017). "Taking Sides in Wars of Attrition". *The American Political Science Review* 111.2, pp. 219–236.

Reich, Noam (2022). "Signaling Strength with Handicaps". *Journal of Conflict Resolution* 66.7-8, pp. 1481–1513.

— (Mar. 2023). "When Can States Signal with Sunk Costs?" *Unpublished Manuscript.*

Schelling, Thomas C. (1960). *The Strategy of Conflict*. Cambridge: Harvard University Press.

Slantchev, Branislav L. (2003). "The Principle of Convergence in Wartime Negotiations". *American Political Science Review* 97.4, pp. 621–632.

— (2005). "Military Coercion in Interstate Crises". *American Political Science Review* 99.4, pp. 533–547.

Takahashi, Yuya (2015). "Estimating a War of Attrition: The Case of the US Movie Theater Industry". *American Economic Review* 105.7, pp. 2204–2241.

Tarar, Ahmer and Bahar Leventoğlu (2009). "Public Commitment in Crisis Bargaining". *International Studies Quarterly* 53.3, pp. 817–839.

TOI, Staff (Nov. 2012). "Security Chiefs Refused Order from PM in 2010 to Prepare Military to Strike Iran within Hours, TV Report Says". *Times of Israel*.

— (Aug. 2015). "Barak: Netanyahu Wanted to Strike Iran in 2010 and 2011, but Colleagues Blocked Him". *Times of Israel*.

Online Supplemental Appendix

Dynamic Screening in International Crises

Contents

A Proofs of Results in the Paper

A.1 Preliminaries

Before proceeding to the proofs from the main paper, it is necessary to provide a more formal definition of a Perfect Bayesian Equilibrium (PBE) to the game and the restrictions it places on an equilibrium. The following definition of a PBE is adapted from Takahashi (2015).

Definition A. 1 *(Perfect Bayesian Equilibrium)*
A PBE in the war of attrition consists of a pair strategies σ_i^* and σ_j^* and beliefs $g_i(w_i|t)$ that satisfy the following properties

(i) For any country i and every type w_i, σ_i^* must satisfy

$$U_i(t_i^*, \theta_i^*, \sigma_j^*|w_i) \geq U_i(\hat{t}_i, \hat{\theta}_i, \sigma_j^*|w_i)$$

for any possible combination of $\hat{t}_i \in [0, \infty)$ and $\hat{\theta}_i \in \{0, 1\}$ for the function $U_i(\cdot)$ as defined in equation (1).

(ii) $g_i(w_i|t)$ must be computed using Bayes' Rule whenever possible.

A.2 Proofs of the Lemmas

This appendix begins with a proof of the lemmas before providing a proof for each of the propositions. Though these lemmas make claims that are standard in war of attrition models, it is necessary to show that these properties still apply when states have and exercise two exit options. Lemma 2 can be applied to any interval during which neither country goes to war. Similarly, Lemma 3 can be applied to any interval of time that has only one country go to war and not the other. Lemma 4 can be applied to any interval of time during which both states go to war.

As stated in the main text, there are three possible outcomes at the horizon date: one in which both countries have finish conceding by the horizon date, one in which only one country has its unresolved types finish conceding by the horizon date and its rival has a mass of types concede on the horizon date, and a stalemate. Due to space constraints, the main text only described the first and third possibilities. However, the proofs in this supplemental appendix will address all three possibilities. To do so, I will restate Lemmas 2 and 3 and Propositions 1,3, and 6 to account for the possibility that one of the two countries can have a mass of types concede at $\overline{T}$. These will demonstrate that it is possible to have one country have a mass of types concede at the horizon date provided that its rival has not yet started to be screened by sunk costs and the game does not end in a stalemate.

A.2.1 Proof of Lemma 1

The proof of this lemma resembles the proof of lemma 1 in Fearon (1994), with minor adjustments resulting from the introduction of sunk costs. The goal is to show that countries will cease to exit in finite time. This must include two parts. The first showing that all types must concede in finite time. The second, showing that all resolved states must go to war in finite time.

First, all unresolved types must concede in finite time. Suppose not, that is suppose that for any time t', there exists a type w_i' who concedes at a time later than time t'. Recall that the cumulative audience costs that must be paid for a concession are strictly increasing. Let w_i^s denote the supremum of conceding types and $\hat{t}$ denote the time at which $a_i\hat{t'} = w_i^s$. Any type of country i conceding after time $\hat{t}$ must be strictly better off going to war instead. A contradiction. It follows that there must exist a date $\overline{T}$ after which no type of country i concedes.

Having demonstrated that all unresolved types concede in finite time, it is straightforward to demonstrate that no resolved type goes to war at a time past $\overline{T}$. Suppose not, that is suppose that type w_i went to to war at time $t > \overline{T}$. Then w_i has a strictly profitable deviation to going to war at time $t^{interim}$ for $\overline{T} < t^{interim} < t$ instead of time t and avoiding the payment of additional sunk costs.[1] It follows that there can be no strategy that has a country go to war after $\overline{T}$ that is part of an equilibrium. ■

A.2.2 Restatement of Lemma 2

Lemma A. 2 *Let $T^1 = min\{T^p, \overline{T}\}$. In any equilibrium $Q_i(t)$ must satisfy the following properties: (i) $Q_i(t)$ must be continuous and strictly increasing on the interval $[0, T^1)$ for both countries and on the interval $[0, T^1]$ for at least one country; (ii) if $T^1 = \overline{T}$, then one country can have a discontinuity in $Q_i(\cdot)$ at $\overline{T}$; (iii) $Q_i(t) < 1$ if and only if $t < \min\{T^p, \overline{T}\}$; $Q_i(0)Q_j(0) = 0$.*

A.2.3 Proof of Lemma 2

The proof of this lemma closely follows the proofs of similar lemmas and propositions in Fearon (1994) and Hendricks, Weiss and Wilson (1988). My proof departs from these papers in characterizing the properties of types who concede at the end of the peaceful phase $T^1 = \min\{T^p, \overline{T}\}$. In particular, the main difference is that in my lemma one country can have a mass of types concede at time $\overline{T}$.

Step 1: Types who concede during the peaceful phase must be playing a mixed strategy. To prove this we will show that the utility of any type on the interval $[0, \min\{T^p, \overline{T}\}]$ must be given by a constant that is independent of both its type and the time at which conedes.

[1] This is true regardless of whether it is assumed that type w_i is paying $k_i(t - \overline{T})$ sunk costs or $\overline{K}_i$ sunk costs.

This suffices to show that the countries are indifferent and can play mixed strategies represented by the cumulative distribution function $Q_i(t)$.

Suppose not, that is, suppose that there were two different types w_i^1 and w_i^2 who conceded on the interval $[0, \min\{T^p, \overline{T}\}]$ and had different expected utilities. Without loss of generality, assume that $U_i(\cdot|w_i^1) > U_i(\cdot|w_i^2)$. First, observe that this cannot be possible if the two types concede at the same time as the types face no risk of war by assumption. However, if the two types concede at different times, then type w_i^2 has a profitable deviation by switching to the strategy played by type w_i^1. But the two types would then have the same expected utility. A contradiction.

Step 2: $Q_i(0)Q_j(0) = 0$: Suppose not, that is suppose that $Q_i(0), Q_j(0) > 0$. This implies that the utility of country i if it concedes at time 0 is given by $U_i(0,0) = \frac{Q_j(0)}{2}$. However, if country i delayed concession by some arbitarily small ϵ, then it could strictly increase its utility to $U_i(\epsilon, 0) = Q_j(\epsilon) - \epsilon a_i$ which for ϵ approaching zero is equal to $Q_j(0)$.[2] A contradiction.

Step 3: $Q_i(t) < 1$ **if and only if** $t < \min\{T^p, \overline{T}\}$**.** Let $T^1 = \min\{T^p, \overline{T}\}$. The proof of this claim has two steps. First, it is necessary to show that it cannot be the case that both $Q_i(t) = 1$ and $Q_j(t) = 1$ for some $t < T^1$. To prove this, suppose not such that no type concedes on the interval $[t, T^1]$. Recall that, by definition, there must be at least one type of country j who goes to war at time T^1. This type of country j could however, strictly increase its utility by instead going to war at time t and paying less sunk costs, thereby implying that the peaceful phase would end at time t and not time T^1.

Second, it is necessary to show that that it cannot be the case that $Q_i(t) = 1$ for some $t < T^1$ while $Q_j(t) < 1$ for all $t < T^1$ and $Q_j(T^1) = 1$. To prove this, suppose not; i.e. suppose that the statement were true. In this case, any type of country i conceding at some time $t' \in [t, T^1]$ has a strictly profitable deviation to instead conceding at time $t' - \epsilon$ for some arbitrarily small $\epsilon > 0$. This is because country i is not conceding during the interval $[t' - \epsilon, t']$ and country j can pay less sunk costs and audience costs by conceding earlier. A contradiction.

Step 4: $\lim_{t \to \hat{t}} Q_i(\hat{t}) = Q_i(t)$ **for** $t \in (0, T^1)$**.** In other words, $Q_i(t)$ can have no mass points on the interval $(0, T^1)$. Suppose not and assume that country i has a mass of types conceding at time $t \in (0, T^1)$. Then any type of country j conceding on the interval $[t - \epsilon, t]$ can strictly increase their utility by conceding slightly after time t at time $t + \epsilon$.

$$U_j(t+\epsilon) - U_j(t-\epsilon) = \int_{t-\epsilon}^{t+\epsilon} [q_i(t)(1 - k_j(t))]dt$$
$$-[1 - Q_i(t+\epsilon)][a_j(t+\epsilon) + k_j(t+\epsilon)] + [1 - Q_i(t-\epsilon)][a_j(t-\epsilon) + k_j(t-\epsilon)]$$

which when taking the limit of ϵ to zero leaves

$$q_i(t) > 0$$

[2] Because $Q_i(t)$ is an increasing function on a compact interval it must be continuous almost everywhere. It follows that there must exist an ϵ small enough such that there is no mass point of types j conceding on the interval $(0, \epsilon]$ thereby implying that the probability of a concession by j on that interval is effectively zero for sufficiently small ϵ.

This implies that there exists some $\epsilon > 0$ such that no type of country j will concede on the interval $[t-\epsilon, t+\epsilon]$. However, the mass of types of country i conceding at time t have a strictly dominant deviation to conceding at time $t-\epsilon$ and paying less sunk costs and audience costs. A contradiction.

Step 5: $Q_i(t') \neq Q_j(t'')$ **for any** $t' \neq t''$**.** In other words there can be no interval during which unresolved types that concede during the peaceful phase do not exit with positive probability. Suppose not. That is suppose, that there existed an interval $[t', t'']$ during which Country j did not exit. Note that it cannot be the case that $t'' = T^1 = \min\{T^p, \overline{T}\}$ as this would violate Step 3. Therefore $t'' < T^1$. Note that Country j would never concede on the interval $(t', t'']$, since they could instead concede at time t' and avoid paying additional sunk costs and audience costs. In addition, there exists an arbitrarily small ϵ such that any type of Country i that concedes on some interval $[t'', t''+\epsilon]$ could strictly increase its utility by conceding at time t' instead. To see this observe that

$$U_i(t') - U_i(t''+\epsilon) = \int_{t''}^{t''+\epsilon} q_j(t)[1-k_i t]dt - k_i[t''+\epsilon-t']$$

which equals $k_i(t''-t')$ when taking the limit of $\epsilon \to 0$. This implies that Country i does not concede on the interval $[t', t''+\epsilon]$. However, in this case, any type of Country j conceding on the interval $[t'', t''+\epsilon]$ could strictly increase its utility by instead conceding at time t' and avoid paying the additional sunk costs and audience costs. This contradicts our original premise that the interval of time on which Country i did not concede was $[t', t'']$.

Step 6: $\lim_{t \to \hat{T^p}} Q_i(\hat{t}) = Q_i(T^p)$ **for** $T^1 = T^p < \overline{T}$**.** In other words, $Q_i(t)$ can have no mass points at T^p when the game transitions to the first screening phase (or in the case in which $T_1^p = T_2^p$, the second screening phase).[3] Suppose not. That is, suppose that there were a mass of types of country i who conceded at time T^p. Following Lemma 3, we know that an unresolved type of country j conceding at some point $T^p - \epsilon$ during the peaceful phase could increase their expected utility by instead conceding at some point $T^p + \epsilon$ instead

$$U_j(T^p+\epsilon, 0; w_j) - U_j(T^p-\epsilon, 0; w_j) = F_i(\beta_i^p)\int_{T^p-\epsilon}^{T^p} q_i(t)[1-k_j]dt$$
$$+\int_{T^p}^{t_j} f_i(\tau_i(t,0))\tau_i'(t,0)[1-k_j t]dt - (a_j t_j - k_j t_j)[1-F_i(\tau(t_j,0))]$$

Taking the limit of ϵ to 0, we find that this equation equals $q_i(T^p) > 0$. Since step 1 of this proof implies that all types conceding during the peaceful phase must have the same expected utility, any type of country j conceding during the first screening phase must strictly prefer to instead concede at some time $T^p + \epsilon$. But then, if no type of country j concedes during the first screening phase, then the mass of types of country i that concede at time T^p could strictly increase their expected utility by conceding instead at time 0 and not paying any sunk costs. A contradiction. The proof to demonstrate that Country j does not have a mass of types conceding at time T^1 follows identical

[3] Note that this proof relies on the properties stated in Lemma 3 (or Lemma 4 if the game proceeds from the peaceful phase to the second screening phase). However, the proof of the properties of Lemma 3 (and Lemma 4) do not rely on this proof.

steps.

Step 7: If $T^1 = \overline{T}$, then $\lim_{t \to \widehat{T}} Q_i(\widehat{T})$ may be greater than $Q_i(\hat{t})$ for one country. First, if Country i has a mass of types concede at time t, then no type of Country j will be willing to fight at $\overline{T}$ since they could strictly increase their utility from $\frac{F_i(Beta_i)q(\overline{T})}{2} + (1 - F_i(Beta_i))w_j$ to $F_i(Beta_i)q(\overline{T}) + (1 - F_i(Beta_i))w_j$ by delaying their decision to exit and seeing whether their rival will go to war or concede. Note that Country j's strategy would not violate Lemma 1 if and only if all types of Country i exit by $\overline{T}$. Therefore, Country i can only have a mass point at time $\overline{T}$ if all types exit at $\overline{T}$.

Second, only one country can have a measurable mass of types concede at time $\overline{T}$. If both countries had a mass of types concede at time $\overline{T}$, then following a similar logic to the above argument, neither country would go to war at that time. However, if neither country conceded at time $\overline{T}$ then both countries would seek to go to war after that time, a violation of Lemma 1. ■

A.2.4 Restatement of Lemma 3

Lemma A. 3 *Let $T^2 = \min\{T^f, \overline{T}\}$. If there exists a $T^p < \overline{T}$, then in any equilibrium $S_j(t)$ must satisfy the following properties: (i) $\sigma_i(\cdot|0)$ must be continuous and strictly increasing on $[T^p, T^2]$ (ii) $S_j(t)$ and $\sigma_i(\cdot|0)$ must be continuous and strictly increasing on the interval $[0, T^2)$; (iii) if $T^2 = \overline{T}$, then $S_j(\cdot)$ may have a discontinuity at $\overline{T}$; (iv) $S_j(t) < 1$ if and only if $t < \min\{T^f, \overline{T}\}$; (v) $\sigma_j(\cdot|1)$ must be continuous and strictly decreasing on $[T^p, T^2]$.*

A.2.5 Proof of Lemma 3

Step 1: Types of country j that concede during the peaceful phase must be playing a mixed strategy. Follows identical steps to the proof of Lemma 2, step 1.

Step 1: $S_i < 1$ if and only if $t < \min\{T^f, \overline{T}\}$. The proof of this claim follows analogous arguments to Lemma 2, Step 3.

Step 2: $S_j(t) < 1$ if and only if $t < T^2$. Suppose not. That is, suppose that $S_j(t') = 1$ for some $t' < \overline{T}$. This implies that country j does not concede on the interval $[t', T^2]$. Recall that there must be at least one type of country i that fights at time T^2 by definition of T^f and $\overline{T}$. This type has a strictly profitable to fight instead at time t' and avoid paying sunk costs. A contradiction.

Step 3: Country i cannot have a mass of types concede at any time. Suppose not. That is, suppose that there is a mass of types of country i conceding at time t where $T^p \leq t < \min\{T^f, \overline{T}\}$. It follows that there exists some interval $[t - \epsilon, t]$ in which country j never exits.

$$U_j(t+\epsilon, \theta; w_j) - U_j(t-\epsilon, \theta; w_j) = \int_{t-\epsilon}^{t+\epsilon} \int_{\{w_i \in \tau_i(t,0)\}} f_i(w)[1 - k_j t] dw dt$$
$$+ \int_{\{w_i \in \tau_i(t_i, \theta | t_i > t+\epsilon)\}} f_i(w)[\mathbb{1}_{\{\theta=1, t_i \neq \infty\}} w_i - \mathbb{1}_{\{\theta=0, t_i \neq \infty\}} a_i(t+\epsilon) - k_j(t+\epsilon)] dw$$
$$- \int_{\{w_i \in \tau_i(t_i, \theta | t_i > t-\epsilon)\}} f_i(w)[\mathbb{1}_{\{\theta=1, t_i \neq \infty\}} w_i - \mathbb{1}_{\{\theta=0, t_i \neq \infty\}} a_i(t-\epsilon) - k_j(t-\epsilon)] dw$$

Or taking the limit of ϵ to zero

$$\int_{\{w_i \in \tau_i(t,0)\}} f_i(w)[1 - \mathbb{1}_{\{\theta=1,t_i\neq\infty\}} w_i - \mathbb{1}_{\{\theta=0,t_i\neq\infty\}} a_i t] dw > 0$$

This implies that no type of country j concedes in the interval $[t-\epsilon, t]$. It then follows that country i only incurs sunk costs by delaying its concession time from $t-\epsilon$ to time t and could strictly increase its payoff by conceding at time $t-\epsilon$ instead. A contradiction.

Step 4: There can be no interval $[t', t'']$ during which no type of country i does not concede. Suppose not. That is, suppose that there is an interval $[t', t'']$ during which no type of country i concedes and where $T^p \leq t' < t'' \leq \min\{T^f, \overline{T}\}$. This implies that there exists an ϵ such that no type of country j exits during the interval $(t', t''+\epsilon]$ as they could instead exit at time t' and avoid paying sunk costs. The fact that there is no point in time at which a mass of types of country i concedes, ensures that there can be no benefit to country j from delaying their exit until $t''+\epsilon$ as the following equation shows

$$\begin{aligned} U_j(t', \theta; w_j) - U_i(t''+\epsilon, \theta; w_j) = -\int_{t''}^{t''+\epsilon} \int_{\{w_i \in \tau_i(t,0)\}} f_i(w)[1 - k_j t] dw dt \\ + \int_{\{w_i \in \tau_i(t_i,\theta | t_i > t')\}} f_i(w)[\mathbb{1}_{\{\theta=1,t_i\neq\infty\}} w_i - \mathbb{1}_{\{\theta=0,t_i\neq\infty\}} a_j t' - k_j t'] dw \\ - \int_{\{w_i \in \tau_i(t_i,\theta | t_i > t'')\}} f(w)[\mathbb{1}_{\{\theta=1,t_i\neq\infty\}} w_i - \mathbb{1}_{\{\theta=0,t_i\neq\infty\}} a_j (t''+\epsilon) - k_j (t''+\epsilon)] dw \end{aligned}$$

which taking the limit of ϵ to zero leaves

$$k_i t'' - k_i t' + \mathbb{1}_{\{\theta=0,t_i\neq\infty\}} a_i t'' - \mathbb{1}_{\{\theta=0,t_i\neq\infty\}} a_i t' > 0$$

This implies that no type of country j ever exits in the interval $[t'', t''+\epsilon]$. However, if no type of country j exits in the interval country $[t'', t''+\epsilon]$, then the types of country i exiting in the interval $[t'', t''+\epsilon]$ can strictly increase their payoff by conceding at time t' instead. This contradicts the premise that the interval of time during which country i does not exit is $[t', t'']$.

Step 5: $\sigma_j(\cdot|1)$ is continuous and strictly decreasing. Suppose that there is a mass of types of country j exiting at time t where $T^p \leq t < \min\{T^f, \overline{T}\}$. Given steps 1 and 2, there must exist a function $Z_i(t)$ that describes the probability that country i has conceded during the first screening phase that is both continuous and strictly increasing. Without loss of generality, assume for the purposes of this step that $\sigma_i(\cdot|0)$ is continuous and strictly increasing. This simplifies the notation and ensures that country j's expected utility can be represented by equation (A. 17).

Country j's utility function is continuous in both t and w_j. This implies that any type of country j choosing to go to war at time t must have its utility function satisfy

$$\frac{\partial U_j(t, 1; w_j)}{\partial t} = f_i(\tau_i(t, 0))\tau_i'(t, 0)[1 - w_j] - k_j[1 - F_j(\tau(t, 0)] = 0$$

It is straightforward to see that any type less resolved than w_j also exiting at time t would have a strictly positive benefit to waiting and exiting later. Moreover, any type more resolved than w_j would have a strictly positive benefit to exiting earlier. This contradicts the premise that there is a mass exiting at time t.

Alternatively suppose that there were an interval $[t', t'']$ during which no type of country j chose to fight. It is still without loss of generality to assume that country j's utility function would be given by equation (A. 17) which is continuous in both w_j and t. A type w'_j would only exit at time t' if $\frac{\partial U_j(t',1;w_j)}{\partial t'} = 0$. Similarly a type w''_j would only exit at time t'' whom $\frac{\partial U_j(t'',1;w''_j)}{\partial t''} = 0$. From the continuity of country j's utility in t and w_j, there must be a type $w_j \in (w''_j, w'_j)$ who strictly prefers to exit in the interval (t', t''), a contradiction.

To complete the proof it is sufficient to show that the equation (A. 17) satisfies single crossing. Taking the derivative with respect to t and w_j we are left with $-f_i(\tau_i(t,0))\tau'_i(t,0)$ which implies that it country j's strategy must be strictly decreasing (Ashworth and Bueno de Mesquita 2006).

Step 6: $\lim_{t \to \hat{t}} S_j(t) = S_j(\hat{t})$ **for** $\hat{t} \in [T^p, T^2)$**.** In other words, country j can have no mass of types conceding on the interval $[T^p, T^2)$. Suppose not. That is, suppose that country j had a mass of types conceding at time $t \in (T^p, T^2)$. Then there exists an $\epsilon > 0$ such that any type of country i conceding on the interval $[t-\epsilon, t]$ would strictly benefit by delaying their concession to time $(t, t+\epsilon]$

$$U_i(t+\epsilon, 0|w) - U_i(t-\epsilon, 0|w) = \int_{t-\epsilon}^{t+\epsilon} [s_j(t)(1 - k_i t - f_j(\tau_j(t,1))\tau'_j(t,1)(w_i - k_i t)dt$$
$$-[F_j(\tau_j(t+\epsilon)) - [F_j(\beta_j^f) - F_j(\beta_j^p)(1 - S_j(t+\epsilon)) + F_j(\beta_j^p)][a_i(t+\epsilon) + k_i(t+\epsilon)]$$
$$+[F_j(\tau_j(t-\epsilon)) - [F_j(\beta_j^f) - F_j(\beta_j^p)(1 - S_j(t-\epsilon)) + F_j(\beta_j^p)][a_i(t-\epsilon) + k_i(t-\epsilon)]$$

which, when taking the limit of ϵ to zero, must equal

$$s_j(t)[1 - k_i t]$$

which is strictly greater than zero. However, such a deviation by country i would violate Step 4 of this proof. Therefore, country j cannot have a mass of types conceding on the interval $[T^p, T^2)$.

Step 7: $S_j(t') \neq S_j(t'')$ **for any** $t' \neq t''$**.** In other words there can be no interval during which country j does not exit. Suppose not. That is suppose that there existed a non-degenerate interval $[t', t'']$ during which country j did not concede so that $S_j(t') = S_j(t'')$. Then there exists an $\epsilon > 0$ such that any type conceding on the interval $[t', t'' + \epsilon]$ would strictly prefer to concede at time t' instead. To see this note that

$$U_i(t', 0|w_i) - U_i(t'' + \epsilon|w_i) = -\int_{t'}^{t''+\epsilon} [s_j(t)(1 - k_i t) - f_j(\tau_j(t,1))\tau'_j(t,1)(w_i - k_i t)]dt$$
$$-[F_j(\tau_j(t')) - (F_j(\beta_j^f) - F_j(\beta_j^p))(1 - S_j(t')) - F_j(\beta_j^p)][1 - S_j(t)][a_i(t') + k_i(t')]$$
$$+[F_j(\tau_j(t''+\epsilon)) - (F_j(\beta_j^f) - F_j(\beta_j^p))(1 - S_j(t''+\epsilon)) - F_j(\beta_j^p)][1 - S_j(t)][a_i(t''+\epsilon) + k_i(t''+\epsilon)]$$

which when taking the limit of ϵ to 0, leaves

$$\int_{t'}^{t''}[f_j(\tau_j(t,1))\tau_j'(t,1)(w_i-k_it)]dt+[(F_j(\beta_j^f)-F_j(\beta_j^p))(1-S_j(t'))-F_j(\beta_j^p)](a_i+k_i)(t''-t')$$
$$+F_j(\tau_j(t'',1))(a_it''+k_it'')-F_j(\tau_j(t',1))(a_it'+k_it')$$

all these terms are positive, thereby implying that no type of country i concedes on the interval $[t'',t''+\epsilon]$ for some small positive $\epsilon>0$. However, in this instance, any type of country j conceding on the interval $[t'',t+\epsilon'']$ could instead concede at time t' and avoid paying additional sunk costs. But we assumed at the start the interval during which country j did not concede was $[t',t'']$. A contradiction.

Step 8: If $T^2=\overline{T}$, then $\lim_{t\to\hat{T}}S_j(t)$ may be smaller than $Q_i(\overline{T})$. That is, country j may have a mass of types concede at time $\overline{T}$ so that $S_j(t)$ may have a discontinuity at time $\overline{T}$. The proof of this step follows an identical logic to that in Lemma 2, Step 7.[4]

Step 9: $\sigma_i(\cdot|0)$ is strictly increasing. To prove this claim, it is sufficient to show that country i's expected utility function satisfies the single crossing property. First, note that in light of steps 3, 4, 5, and 6, country i's expected utility function can be represented with

$$U_i(t_i,0;w_i|T^p\leq t_i\leq T^f)=F_j(\beta_j^p)\int_0^{T^p}q_j(t)[1-k_it]dt \quad \text{(A. 1)}$$
$$+[F_j(\beta_j^f)-F_j(\beta_j^p)]\int_{T^p}^{t_i}s_j(t)[1-k_it]dt-\int_{T^p}^{t_i}f_j(\tau_j(t,1))\tau_j'(t,1)[w_i-k_it]dt \quad \text{(A. 2)}$$
$$-[F_j(\tau_j(t_i,1))-(F_j(\beta_j^f)-F_j(\beta_j^p))S_j(t_i)-F_j(\beta_j^p)][a_it_i+k_it_i] \quad \text{(A. 3)}$$

and is continuous in both t_i and w_i. The cross-partial of country i's utility with respect to t_i and w_i is $-f_j(\tau_j(t_i,1)\tau_j(t_i,1)$ which is positive. This is sufficient to show that $\sigma_i(\cdot|0)$ is strictly increasing (Ashworth and Bueno de Mesquita 2006). ■

A.2.6 Proof of Lemma 4

Step 1: Country i ($i=1,2$) can have a mass of types conceding at time t if it also has a mass of types escalating at time t. Suppose not. That, is, suppose that there were a time t where a mass of types of country j conceded but there were no mass of types of country j escalating. In that case, there exists an $\epsilon>0$ such that any types of country i exiting during the

[4] The sole difference is that we have already ruled out Country i having a mass of types concede at time $\overline{T}$ with Step 3 of this proof.

interval $[t-\epsilon, t+\epsilon)$ could strictly increase their payoff by conceding at time $t+\epsilon$

$$U_i(t+\epsilon,\theta;w_i) - U_i(t-\epsilon,\theta;w_i) = \int_{t-\epsilon}^{t+\epsilon} \int_{\{w_j \in \tau_j(t,0)\}} f_j(w)[1-k_i t)]dw + \int_{\{w_j \in \tau_j(t,1)\}} f_j(w)[w_i - k_i t]dwdt$$
$$+ \int_{\{w_j \in \tau(t_j,\theta|t_j > t+\epsilon)\}} f_j(w)[\mathbb{1}_{\{\theta=0,t_i\neq\infty\}} w_i - \mathbb{1}_{\{\theta=0,t_i\neq\infty\}} a_i(t+\epsilon) - k_j(t+\epsilon)]dw$$
$$- \int_{\{w_j \in \tau(t_j,\theta|t_j > t-\epsilon)\}} f_j(w)[\mathbb{1}_{\{\theta=0,t_i\neq\infty\}} w_i - \mathbb{1}_{\{\theta=0,t_i\neq\infty\}} a_i(t-\epsilon) - k_j(t-\epsilon)]dw$$

Taking the limit of ϵ to 0, we find that we are left with

$$\int_{\{w_j \in \tau_j(t,0)\}} f_j(w)[1-k_i t]dw > 0$$

This implies that no type of country i concedes during the interval $[t-\epsilon, t+\epsilon]$. But then any type of country j conceding at time t could strictly strictly increase their payoff by conceding at time $t-\epsilon$ and avoid avoid paying sunk costs. A contradiction.

Step 2: If country j has a mass of types concede at time t, then there exists an $\epsilon > 0$ such that country i does not go to war during the interval $[t-\epsilon, t]$. Suppose not. That is, suppose that country j had a mass of types conceding at time t and that for every $\epsilon > 0$ there were always some type of country i who went to war in the interval $[t-\epsilon, t]$. That type of country i could strictly increase its payoff by conceding at time $t+\epsilon$ as

$$U_i(t+\epsilon,1;w_i) - U_i(t-\epsilon,1;w_i) = \int_{t-\epsilon}^{t+\epsilon} \int_{\{w_j \in \tau_j(t,0)\}} f_j(w)[1-k_i t]dw + \int_{\{w_j \in \tau_j(t,1)\}} f_j(w)[w_i - k_i t]dwdt$$
$$+ \int_{\{w_j \in \tau(t_j,\theta|t_j > t+\epsilon)\}} f_j(w)[w_i - k_j(t+\epsilon)]dw - \int_{\{w_j \in \tau(t_j\theta|t_j > t-\epsilon)\}} f_j(w)[w_i - k_i(t-\epsilon)]dw$$

which when taking the limit of ϵ to zero leaves

$$\int_{\{w_j \in \tau_j(t,0)\}} f_j(w)[1-w_i]dw > 0$$

This shows that country i can strictly increase its payoff by going to war at time $t+\epsilon$ instead of time $t-\epsilon$. A contradiction.

Note that Step 2 also implies that both countries cannot have a mass of types concede at the same time t, as step 1 implies that this would also require that both countries have a mass of types go to war at time t, thereby contradicting the previous result.

Step 3: Suppose country j has a mass of types conceding at time t. Then there can be no interval $[t-\epsilon, t)$ during which no type of either country concedes. Suppose not.

That is, suppose that there were some $\epsilon > 0$ such that no type of country i conceded in the interval $[t-\epsilon, t)$. Step 2 establishes that there must be some interval $[t-\epsilon', t]$ during which country i does not go to war. Following the arguments in the main text, types of country j must be indifferent as to when they concede during the interval $[t-\epsilon, t)$. If there were an interval $[t-\epsilon, t)$ such that country i did not concede during that interval, then types of country j conceding at time t would have a profitable deviation to conceding at time $t-\epsilon$ and avoid paying sunk costs. A contradiction.

Alternatively, there can be no interval $[t-\epsilon, t)$ during which no type of country j concedes. If there were such an interval, then any type of country i conceding in that interval could simply concede at time $t-\epsilon$ and avoid paying sunk costs and the threat of war. This contradicts the previous result.

Step 4: Suppose country j has a mass of types conceding at time t'. Then there can be no interval $[t', t'']$ during which no type of either country i or country j does not concede. Suppose not. That is, suppose that there were an interval $[t', t'']$ during which no type of either country i or country j did not concede. Without loss of generality assume that it is country i that has no types concede during $[t', t'']$. Step 3 implies that there can be no mass point of conceding types at time t'' (by either country). It follows that any type of country j exiting during the interval $(t', t''+\epsilon]$ for some $\epsilon > 0$ could strictly increase their payoff by exiting at time t' and avoid paying sunk costs. But then any type of country i conceding at time t'' could strictly increase their payoff by conceding at time t'. A contradiction.

Step 5: There can be no mass of countries conceding. Suppose not. That is, suppose that country j had a mass of types conceding at time t. It follows from the previous steps that there must exist some cumulative distribution function $Z_i(t)$ that is continuous and strictly increasing describing the probability that country i concedes in the neighborhood of time t. This implies that the utility for a type of country j that escalates at time t is continuous on some interval $[t-\epsilon, t+\epsilon]$. Therefore a type of country j going to war at time j must satisfy $\frac{\partial U_j(t,1;w_j)}{\partial t} = 0$. However, this will only be true for country j when

$$\frac{z_i(t)}{1 - Z_i(t)} = \frac{k_j}{1 - w_j}$$

If this is satisfied for some type w_j escalating at time t, then any type less resolved and also escalating at time t that could strictly increase its expected utility by escalating later in that interval. Similarly, any type more resolved than w_j could strictly increase its expected utility by escalating prior to time t. This contradicts the requirement that there is a mass of types of country j going to war at time t established in Step 1. Therefore, there can therefore be no mass of types of country j conceding at time t.

Step 6: There can be no interval $[t', t'']$ during which no type of country i ($i = 1, 2$) does not concede. The proof of this claim follows identical steps to that in Lemma 3, Step 7.

Step 7: $\sigma_i(\cdot|1)$ is strictly decreasing (for $i = 1, 2$) Steps 5 and 6 imply that there must exist some continuous and strictly increasing cumulative distribution function $C_j(t)$ representing the probability that country j concedes during the second screening phase. The arguments presented in Step 5 of this Lemma can therefore be extended to rule out the possibility of a mass point of types of country i going to war at any time (as opposed to in conjunction with a mass of types of country i conceding). It is then possible to replicate the arguments made in Lemma 3, Step 5 to show that there there can be no interval $[t', t'']$ where no type of country i $(i = 1, 2)$ goes to war. To complete the step, it is only necessary to show that country i's utility satisfies single crossing. As in Lemma 3, Step 5, the cross-partial of country i's utility function with respect to its exit time t_i and its type w_i is $-z_i(t)$ which is negative as desired.

Step 8: $\sigma_i(\cdot|0)$ is strictly increasing As in Lemma 3, step 9, it is sufficient to show that country i's utility satisfies single crossing. Using the previous steps we can rewrite the expected utility of a type of country i conceding during the first screening phase as

$$U(t_i, \theta; w_i | T^f \leq t_i < \overline{T}) = F_j(\beta_j^p) \int_0^{T^P} q_j(t)[1 - k_i t]dt + [F_j(\beta_j^f) - F_j(\beta_j^p)] \int_{T^p}^{T^f} s_j(t)[1 - k_i(t)]dt$$
$$+ \int_{T^f}^{t_i} f_j(\tau_j(t, 0))\tau_j'(t, 0)[1 - k_i(t)]dt - \int_{T^p}^{t_i} (\tau_j(t, 1))\tau_j'(t, 1)[w_i - k_i t]dt$$
$$+ [F_j(\tau_j(t_i, 1)) - F_j(\tau_j(t_i, 0))] \max\{-a_i t - k_i t, w_i - k_i t\}$$

Taking the cross partial of this expected utility function, we are left with $-f_j(\tau_i(t_i, 1))\tau_j'(t_i, 1)$ which is positive as desired. ■

A.3 Proofs of the Propositions

To prove the propositions it is necessary to show that no type has a profitable deviation from its prescribed strategy. This requires checking that no type can increase its expected utility by changing its exit time and its exit option. In the proofs of Propositions 1-5, I hold countries exit option constant and only rule out deviations from prescribed exit times for their assigned phases. Deviating from prescribed exit times to those in different phases is easily ruled out because of the monotonicity of the pure strategies during the screening phases. I do not restate this argument for each individual proposition. Proposition 6 verifies the assignment of types to their exit strategies.

It is easiest to prove the propositions 3 and 4 that characterize strategies in the first screening phase after proving propositions 1,2 and 5, which characterize behavior in the other phases. Proposition 6 is proven last.

A.3.1 Restatement of Proposition 1

Proposition A. 1 *Let $T^1 = \min\{T^p, \overline{T}\}$. If $T^1 = T^p$, then types $w_i \in [\underline{w}_i, \beta_i^p]$ $(i = 1, 2)$ concede on the interval $[0, T^1]$ according to the following strategy*

$$\frac{q_j(t)F_j(\beta_j^p)}{1 - F_j(\beta_j^p)Q_j(t)} = \frac{a_i + k_i}{1 + a_i t} \tag{A. 4}$$

If $T^1 = \overline{T}$*, then one country may instead play according to equation (A. 4) on the interval* $[0, \overline{T})$ *and play*

$$q(\overline{T}) = \lim_{t \to \overline{T}^-} 1 - Q(t) \qquad \text{(A. 5)}$$

Since no type goes to war during the peaceful phase, both countries beliefs are given by

$$g_i(w_j|t) = \begin{cases} \frac{f_j(w_j)[1-Q_j(t_i)]}{1-Q_j(t)F_j(\beta_j^p)} & \text{if } w_j \in [\underline{w}_j^t, \beta_j^p] \\ \frac{f_j(w_j)}{1-Q_j(t)F_j(\beta_j^p)} & \text{if } w_j \in [\beta_j^p, \overline{w}_j] \end{cases} \qquad \text{(A. 6)}$$

A.3.2 Proof of Proposition 1

Types who concede during the peaceful phase play a mixed strategy and are indifferent as to when they concede. Therefore, they cannot profitably deviate to playing any other strategy that would have them concede during the peaceful phase. ■

A.3.3 Proof of Proposition 2

Proposition 1 implies that I can restate the utility function in equation (1) for a country going to war during the peaceful phase as

$$U_i(t_i, 1; w_i|t \leq T^p) = F_j(\beta_j^p)\int_0^{t_i} q_j(t)[1 - k_i(t)]dt + [1 - F_j(\beta_j^p)Q_j(t_i)][w_i - k_i t_i] \qquad \text{(A. 7)}$$

Time T_i^p is derived by taking the first-order condition of this expected utility function. Therefore, to prove that no type of any country will go to war before T^p, it is necessary to show that equation (A. 7) is concave in t_i. This can be verified by showing that the second order condition is negative. Taking the derivative of equation (A. 7) with respect to t_i twice, we find that it is is concave if

$$F_j(\beta_j^p)\frac{q_j(t)}{dt}[1 - \overline{w}_i] + k_i F_j(\beta_j^p)q_j(t_i) < 0$$

To find the value of $F_j(\beta_j^p)\frac{q_j(t)}{dt}$, we can rearrange the expression in equation (A. 4) to find that

$$F_j(\beta_j^p)q_j(t) = \frac{a_i + k_i}{1 + a_i t}[1 - F_j(\beta_j^p)Q_j(t)]$$

and then take the derivative with respect to t

$$F_j(\beta_j^p)\frac{dq_j(t)}{dt} = -\frac{a_i[a_i + k_i]}{[1 + a_i t]^2}[1 - F_j(\beta_j^p)Q_j(t)] - F_j(\beta_j^p)q_j(t)\frac{a_i + k_i}{1 + a_i t}$$

Substituting this back in to the second order condition, we are left with

$$\left[-\frac{a_i[a_i + k_i]}{[1 + a_i t_i]^2}[1 - F_j(\beta_j^p)Q_j(t_i)] - F_j(\beta_j^p)q_j(t_i)\frac{a_i + k_i}{1 + a_i t_i}\right][1 - \overline{w}_i] + k_i F_j(\beta_j^p)q_j(t_i) < 0$$

I begin by dividing by $[1 - F_j(\beta_j^p)Q_j(t_i)]$ and multiplying by $[1 + a_i t_i]^2$ to get

$$\left[-a_i[a_i + k_i] - \frac{F_j(\beta_j^p)q_j(t)}{1 - F_j(\beta_j^p)Q_j(t)}[a_i + k_i)][1 + a_i t_i]\right][1 - \overline{w}_i] + k_i \frac{F_j(\beta_j^p)q_j(t_i)}{1 - F_j(\beta_j^p)Q_j(t_i)}[1 + a_i t_i]^2 < 0$$

Substituting in for the hazard rates, we have

$$\left[-a_i[a_i + k_i] - [a_i + k_i]^2\right][1 - \overline{w}_i] + k_i[a_i + k_i][1 + a_i t_i] < 0$$

We divide by $[a_i + k_i]$ to get

$$-[2a_i + k_i][1 - \overline{w}_i] + k_i[1 + a_i t_i] < 0$$

Isolating $\overline{w}_i$ we have

$$-[2a_i + k_i] + k_i[1 + a_i t_i] < -\overline{w}_i[2a_i + k_i]$$

Or dividing by $-[2a_i + k_i]$ we have

$$\overline{w}_i < 1 - \frac{k_i}{2a_i + k_i}[1 + a_i(t_i)]$$

From equation (4), we know that

$$\overline{w}_i = 1 - \frac{k_i}{a_i + k_i}[1 + a_i t_i]$$

which is indeed less than the term on the right-hand side. ■

A.3.4 Proof of Proposition 5

To prove the proposition, we need to verify that $\sigma_i(\cdot|1)$ is strictly decreasing in the second screening phase and to show that each country's expected utility function is concave in t_i. If this is the case, then the strategies in equations (12) and (13) that were derived from first order conditions must be utility maximizing.

To start, we will verify that the expression in (13) is negative. The denominator in that expression

$$[1 - \overline{w}_i][\underline{w}_i + a_i t]$$

is negative. Therefore for the expression to negative as required, the numerator must be positive. The numerator will be positive if

$$k_i[\overline{w}_i + a_i t] - a_i[1 - \overline{w}_i^t] > 0$$

which can be rearranged to

$$\overline{w}_i > 1 - \frac{k_i}{a_i + k_i}[1 + a_i t]$$

To see that this holds, observe that we can rearrange the expression in (13) to show that $\overline{w}_i^t$ is given by

$$\overline{w}_i^t = 1 - \frac{k_i t[1 + a_i t]}{a_i + k_i \frac{f_j(\tau_j(t,1))}{F_j(\tau_j(t,1)) - F_j(\tau_j(t,0))}[\underline{w}_i^t + a_i t]}$$

Substituting in this value for $\overline{w}_i^t$ back into the inequality, it is straightforward to see that the numerator is positive.

Next, we examine the second order condition for a type that is going to war. Using Lemma 4, we can rewrite the utility function for types of country i $(i = 1, 2)$ who exit during the second screening phase as follows

$$\begin{aligned} U(t_i, \theta; w_i | T^f \leq t_i < \overline{T}) = F_j(\beta_j^p) \int_0^{T^P} q_j(t)[1 - k_i t]dt + [F_j(\beta_j^f) - F_j(\beta_j^p)] \int_{T^p}^{T^f} s_j(t)[1 - k_i(t)]dt \\ + \int_{T^f}^{t_i} f_j(\tau_j(t,0))\tau_j'(t,0)[1 - k_i(t)]dt - \int_{T^p}^{t_i} (\tau_j(t,1))\tau_j'(t,1)[w_i - k_i t]dt \\ + [F_j(\tau_j(t_i,1)) - F_j(\tau_j(t_i,0))] \max\{-a_i t - k_i t, w_i - k_i t\} \end{aligned} \quad \text{(A. 8)}$$

Next, Taking the derivative of (A. 8) twice with respect to t_i, we find that a resolved type's expected utility function will be concave if

$$\frac{df_j(\tau_j(t_i,0))\tau_i'(t_i,0)}{dt}[1 - \overline{w}_i] - k_i[f_j(\tau_j(t_i,1))\tau_j'(t_i,1) - f_j(\tau_j(t_i,0))\tau_i'(t_i,0)] < 0$$

To find the value of (13), we can rearrange country j's strategy in equation (12) to find that

$$f_j(\tau_j(t,0))\tau_i'(t,0)[1 - \overline{w}_i^t] = k_i[F_j(\tau_j(t,1)) - F_j(\tau_j(t,0))]$$

Taking the derivative with respect to t have

$$\frac{df_j(\tau_j(t,0))\tau_i'(t,0)}{dt}[1 - \overline{w}_i^t] - \frac{\overline{w}_i^t}{dt} f_j(\tau_j(t,0))\tau_i'(t,0) = k_i[f_j(\tau_j(t_i,1))\tau_j'(t_i,1) - f_j(\tau_j(t_i,0))\tau_j'(t_i,0)]$$

Substituting this back into the second derivative of country i's utility function we are left with

$$\frac{d\overline{w}_i^t}{dt} f_j(\tau_j(t_i,0))\tau_j'(t_i,0) < 0$$

which is negative as desired given that $\tau_j'(t_i,0)$ is increasing and that $\frac{d\overline{w}_i^t}{dt}$ is decreasing in t per

Lemma 4.

Finally, we examine the second order condition for a type that concedes. Taking the derivative of (A. 8) twice with respect to t_i, we find that an unresolved type's expected utility function will be concave if

$$\frac{f_j(\tau_j(t_i,0))\tau_j'(t_i,0)}{dt}[1+a_it_i] + f_j(\tau_j(t_i,0))\tau_j'(t_i,0)[k_i+2a_i]$$
$$-\frac{f_j(\tau_j(t_i,1))\tau_j'(t_i,1)}{dt}[\underline{w}_i+a_it_i] - f_j(\tau_j(t_i,1))\tau_j'(t_i,1)[k_i+2a_i]$$

Once again, we cannot proceed without additional information relating to the derivative of the τ terms. We know from the first order condition of a conceding type of country i that it is possible to express the hazard rate for concession as

$$\frac{f_j(\tau_j(t,0))\tau_j'(t,0)}{F_j(\tau_j(t,1))-F_j(\tau_j(t,0))} = \frac{f_j(\tau_j(t,1))\tau_j'(t,1)}{F_j(\tau_j(t,1))-F_j(\tau_j(t,0))}\frac{[\underline{w}_i^t+a_it_i]}{1+a_it}+\frac{a_i+k_i}{1+a_it}$$

The expression in (13) is derived after rearranging and substituting in (12) into this expression. However, to derive more information regarding the τ terms, we instead take the derive of this expression with respect to t to find that

$$\frac{f_j(\tau_j(t,0))\tau_j'(t,0)}{dt}[1+a_it] = -f_j(\tau_j(t,0))\tau_j'(t,0)[k_i+2a_i] + \frac{f_j(\tau_j(t,1))\tau_j'(t,1)}{dt}[\underline{w}_i^t+a_it)]$$
$$+f_j(\tau_j(t,1))\tau_i'(t,1)\left[\frac{d\underline{w}_i^t}{dt}+k_i+2a_i\right]$$

Substituting this back into the second order condition we are left with

$$f_j(\tau_i(t_i,1))\tau_i'(t_i,1)\frac{d\underline{w}_i^t}{dt}<0$$

which is negative as desired given that $\tau_j'(t_i,1)$ is decreasing and that $\underline{w}_i$ is increasing in t per Lemma 4. ■

A.3.5 Restatement of Proposition 3

Proposition A. 2 *Let $T^2 = \min\{T^f, \overline{T}\}$. If there exists a $T^p < \overline{T}$. During $[T^p, T^2]$, country i concedes by playing $\tau_i(\cdot, 0)$ as given by*

$$\frac{f_i(\tau_i(t,0))\tau_i'(t,0)}{1-F_i(\tau_i(t,0))} = \frac{a_j+k_j}{1+a_jt} \tag{A. 9}$$

If $T^2 = T^f$, then types $w_j \in [\beta_j^p, \beta_j^f]$ concede by playing

$$\frac{[F_j(\beta_j^f) - F_j(\beta_j^p)]s_j(t)}{F_j(\tau_j(t,1)) - [F_j(\beta_j^f) - F_j(\beta_j^p)]S_j(t) - F_j(\beta_j^p)} = \frac{a_i + k_i}{1 + a_i t} \quad \text{(A. 10)}$$

$$+\frac{f_j(\tau_j(t,1))\tau_j'(t,1)}{F_j(\tau_j(t,1)) - [F_j(\beta_j^f) - F_j(\beta_j^p)]S_j(t) - F_j(\beta_j^p)} \times \frac{\underline{w}_i^t + a_i t}{1 + a_i t} \quad \text{(A. 11)}$$

on the interval $[T^p, T^2]$. If $T^2 = T^f$, then conceding types of country j may choose to play (A. 11) on the interval $[T^p, T^2)$ and play

$$s_j(\overline{T}) = \lim_{t \to \overline{T}^-} 1 - S_j(t) \quad \text{(A. 12)}$$

Resolved types of country j play

$$\sigma_j(w_j|1) = [1 - w_j]\left[\frac{1}{k_j} + \frac{1}{a_j}\right] - \frac{1}{a_j} \quad \text{(A. 13)}$$

Each country's posterior beliefs posterior beliefs during this period are given by

$$g_i(w_j|t) = \begin{cases} \frac{f_j(w_j)[1-S_j(t)]}{F_j(\tau_j(t_i,1)) - [F_j(\beta_j^f) - F_j(\beta_j^p)]S_j(t) - F_j(\beta_j^p)} & \text{if } w_j \in [\beta_j^p, \beta_j^f] \\ \frac{f_j(w_j)}{F_j(\tau_j(t_i,1)) - [F_j(\beta_j^f) - F_j(\beta_j^p)]S_j(t) - F_j(\beta_j^p)} & \text{if } w_j \in [\beta_j^f, \overline{w}_j^t] \\ 0 & \text{otherwise} \end{cases} \quad \text{(A. 14)}$$

1

$$g_j(w_i|t) = \begin{cases} \frac{f_i(w_i)}{1 - F_i(\tau_i(t,0))} & \text{if } w_i \in [\underline{w}_i^t, \overline{w}_i] \\ 0 & \text{otherwise} \end{cases} \quad \text{(A. 15)}$$

A.3.6 Proof of Proposition 3

Using Lemma 3, we can rewrite the expected utility function for an unresolved type of country i who concedes during the first screening phase as

$$\begin{aligned} U_i(t_i, 0; w_i | T^p \leq t_i \leq T^f) = F_j(\beta_j^p) \int_0^{T^p} q_j(t)[1 - k_i t]dt \\ +[F_j(\beta_j^f) - F_j(\beta_j^p)] \int_{T^p}^{t_i} s_j(t)[1 - k_i t]dt - \int_{T^p}^{t_i} f_j(\tau_j(t,1))\tau_j'(t,1)[w_i - k_i t]dt \\ -[F_j(\tau_j(t_i,1)) - (F_j(\beta_j^f) - F_j(\beta_j^p))S_j(t_i) - F_j(\beta_j^p)][a_i t_i + k_i t_i] \end{aligned} \quad \text{(A. 16)}$$

Similarly the utility for a type of country j who exits during the first screening phase is given by

$$U_j(t_j,\theta_j;w_j|T^p \leq t_i \leq T^f) = F_i(\beta_i^p)\int_0^{T^p} q_i(t)[1-k_j]dt$$
$$+\int_{T^p}^{t_j} f_i(\tau_i(t,0))\tau_i'(t,0)[1-k_jt]dt + [1-F_i(\tau(t_j,0))]\max\{-a_jt_j-k_jt_j, w_j-k_jt_j\} \tag{A. 17}$$

Once again, conceding types of country j have no profitable deviation on account of their being indifferent and playing a mixed strategy. Similarly types of country j going to war have no incentive to deviate - their strategy is derived from a first-order condition that was already shown to produces a maximum in the proof of Proposition 2. The proof that that the first-order condition for a conceding type of country i produces a maximum requires showing that the utility function in (A. 16) is concave in t. The proof that this property is satisfied is identical to the proof showing that the utility for a conceding type is concave in the second screening phase and follows immediately from the proof of proposition 5.

A.3.7 Proof of Proposition 4

Using Lemma 3, we can restate type $\overline{w}_i$'s expected utility for going to war during the first screening phase as follows

$$U_i(t_i,1;\overline{w}_i|T_p \leq t_i \leq T^f) = F_j(\beta_j^p)\int_0^{T^p} q_j(t)[1-k_it]dt$$
$$+[F_j(\beta_j^f)-F_j(\beta_j^p)]\int_{T^p}^{t_i} s_j(t)[1-k_it]dt - \int_{T^p}^{t_i} f_j(\tau_j(t,1))\tau_j'(t,1)[\overline{w}_i-k_it]dt \tag{A. 18}$$
$$+[F_j(\tau_j(t_i,1))-(F_j(\beta_j^f)-F_j(\beta_j^p))S_j(t_i)-F_j(\beta_j^p)][\overline{w}_i-k_it_i]$$

To show that the most resolved type of country i has a maximum when going to war at time T^f, it is necessary to show that its utility is concave. The derivative of country i's utility function as given by equation (A. 18) with respect to t_i shows that country i will go to war when

$$\frac{F_j(\beta_j^f)-F_j(\beta_j^p)s_j(t)}{F_j(\tau_j(t_i,1))-(F_j(\beta_j^f)-F_j(\beta_j^p))S_j(t_i)-F_j(\beta_j^p)} = \frac{k_i}{1-\overline{w}_i}$$

Note that this first order condition is identical to that resulting from equation (A. 8). However, the hazard rate characterizing Country j's concession rate is given by (A. 11) instead of simply being (12). However, note that by taking the derivative of (A. 8) for a conceding country, it is possible to express the hazard rate in (12) exactly as (A. 11). It therefore follows that the escalating type of country i's utility must be concave since it is so in the the proof in Proposition 3.■

A.3.8 Restatement of Proposition 6

Proposition A. 3 *Strategic behaviour ends at $\overline{T}$, which can arrive during any phase. Countries' choice of exit strategy and their behavior at the horizon date is determined by the following:*

(i) All types exit by the horizon date: There exists an equilibrium where types $w_i \in [\underline{w}_i, \beta_i]$ concede and types $w_i \in (\beta_i, \overline{w}_i]$ go to war where $\beta_i = -a_i\overline{T}$. Any type still participating in the crisis at $\overline{T}$ goes to war at that time.

(ii) One country exits by the horizon date: If $\overline{T} < T^f$, then there exists an equilibrium where types $w_i \in [\underline{w}_i, \beta_i]$ concede and types $w_i \in (\beta_i, \overline{w}_i]$ go to war where

$$\beta_i = \frac{-a_i(\overline{T}) + \mu F_j(\beta)}{1 - F_j(\beta_j)} \tag{A. 19}$$

and $\mu = q_j(\overline{T})$ if $\overline{T} < T^p$ and $\mu = s_j(\overline{T})$ otherwise. In turn, types $w_j \in [\underline{w}_j, \beta_j]$ concede and types $w_j \in (\beta_j, \overline{w}_j]$ go to war where $\beta_j = -a_j\overline{T}$.

(iii) Some types remain in forever: If $K_i < a_i\overline{T}$ for both $i = 1, 2$, then there exists an equilibrium where types $w_i \in [\underline{w}_i, \beta_i]$ concede for β_i as given by

$$\frac{F_j(\overline{w}_j^{\overline{T}}) - F_j(-\overline{K}_j)}{F_j(\overline{w}_j^{\overline{T}}) - F_j(\beta_j)}\beta_i - \frac{F_j(-\overline{K}_j) - F_j(\beta_j)}{F_j(\overline{w}_j^{\overline{T}}) - F_j(\beta_j)}K_i = -a_i\overline{T} \tag{A. 20}$$

Types $w_i \in (\beta_i, -K_i]$ remain in the crisis forever and types $w_i \in (-K_i, \overline{w}_i]$ go to war. Any type from the latter set still participating in the crisis at $\overline{T}$ go to war at that time.

A.3.9 Proof of Proposition 6

Parts (i) and (iii) of the proof follow directly from the arguments in the main text. For part (ii), note that type β_i is the type that is indifferent between conceding and paying $a_i\overline{T}$ audience costs and remaining in the war of attrition and obtaining a concession with probability $q_j(\overline{T})F_j(\beta_j)$ and going to war with probability $1 - F_j(\beta_j)$. It follows that any type with a lower wartime payoff than β_i must strictly prefer to concede and any type stronger than β_i will prefer to remain in the crisis and risk going to war. Type β_j is the type indifferent between paying audience costs and going to war. It follows that any type with a lower wartime payoff prefers to concede and any type with a larger wartime payoff prefers to go to war.

Note that this requires that remaining resolved types of Country i play a strategy that has them exit at a time $t_i > \overline{T}$. In this case, country j has no incentive to deviate since the only types of country i remaining at $\overline{T}$ are resolved types who wish to fight, implying that delay can no longer lead to a concession and will increase the amount of sunk costs paid. Similarly, resolved types of Country i have no interest in going to war at an earlier time as per Propositions 2 and 4 since $\overline{T} < T^f$.

Moreover, observe that a stalemate is not possible when $q_j(\overline{T}) > \epsilon$ or $s_j(\overline{T})\epsilon$ for an arbitrarily small $\epsilon > 0$. This is because resolved types of Country i would not go to war at time $\overline{T}$, since they would prefer to wait until country j finished conceding. However, this would then require that any type $w_i \in [-\overline{K}_i, \overline{w}_i]$ go to war at a time $t > \overline{T}$, a violation of Lemma 1. ■ ■

Note that a stalemate is not possible if one type has ma

B Ruling out Alternative Equilibria

In this section I rule out (i) an equilibrium in which the war of attrition ends with probability 1 at $t = 0$ and (ii) any equilibrium which has types exit after time $t = 0$ and which does not feature the three phases described in the main text in the order that they are described. That is, I show that the equilibrium must begin with a peaceful phase, before transitioning into the first screening phase, and only then to the second screening phase.[5] The first proposition demonstrates that the war of attrition must proceed past $t = 0$ with positive probability

Proposition B. 1 *The war of attrition must feature a positive probability of concession and war after* $t = 0$*.*

The proof of this result follows directly from the way simultaneous exit is determined in the expected utility function in (1). Since the outcome of the model (i.e. war or concession) is adjudicated by a simple coin toss, there are strong pressures against going to war or conceding when one's rival is likely to concede with positive probability. This prohibits all countries from exiting immediately at $t = 0$, since there is an incentive to delay if one's rival is going to concede at $t = 0$ with positive probability.

The following proposition demonstrates the second result.

Proposition B. 2 *Any equilibrium that does not have the war of attrition end at* $t = 0$*, must have countries play strategies according to Propositions 1-5.*

The proof of this proposition relies on the fact that Lemmas 2, 3, and 4 still apply to any interval of time in which neither country has types go to war, one country has types go to war, or both countries have types go to war. It then shows that continuity and monotonicity and properties perscribed by these lemma are incongruent with countries going to war prior to when is stipulated by Propositions 1 through 5.

Proof of Proposition B. 1

It cannot be the case that resolved types of both countries cannot go to war at time $t = 0$. This is because unresolved types $w_i < 0$ for $(i = 1, 2)$ would respond by trying to concede at $t = 0$ implying that resolved types could increase their expected utility from $F_j(0)\frac{1+w_i}{2} + (1 - F_j(0))w_i$ to

[5] As mentioned in the main text there is an exception in that the first screening phase can be skipped when resoved types of both countries want to go to war at the same time at the end of the peaceful phase.

$F_j(0) + (1 - F_j(0))w_i$ by delaying concession by an arbitrarily small ϵ. Similarly, it cannot be the case that both countries concede at $t = 0$. As one country could strictly increase its utility from $\frac{1}{2}F_j(0)$ to $F_j(0)$ by delaying concession by an arbitrarily small ϵ. ■

Proof of Proposition B. 2

Any alternative equilibria must contain intervals where neither country has any types that fight, intervals where only one country has types that fight, or intervals where both countries have types that fight. Lemmas 2, 3, and 4 respectively would still apply to such intervals in any alternative equilibrium.

The requirement in Lemma 4 that $\sigma_i(\cdot|1)$ be continuous and strictly decreasing rules out equilibria that begin with an interval in which both countries go to war too early. To see why simply note that this requirement implies that (13) must be the hazard rate that determines when countries go to war in any such alternative equilibrium. Observing $\sigma_i(\cdot|1)$, we note that the denominator is always negative and that the numerator will only be negative as required if

$$k_i[\overline{w}_i^t + a_i t] > a_i[1 - w_i]$$

Rearranging, this requires that

$$\frac{k_i}{1 - \overline{w}_i^t} > \frac{a_i + k_i}{1 + a_i t}$$

We know from the discussion of the peaceful phase, that this will not hold for a single type until T^p. It follows that the game cannot begin with an interval in which both sides concede as both countries would be unable to play a strategy $\sigma_i(\cdot|1)$ that is continuous and strictly decreasing.

Therefore any alternative equilibrium must begin with an interval where at least one country does not have any escalating types and the other does. Without loss of generality, let country j be the country to have the type(s) escalating in such an interval. From the discussion of the first screening phase, we know that any such interval must have conceding types of country i play a strategy that keeps conceding types of country j indifferent and must therefore be given by (A. 9). However, from the proof of Proposition 2, we know that when country i concedes using this hazard rate that no type of country j will choose to escalate prior to T^p. It follows that the equilibrium where countries play according to the strategies described in Propositions 1-5 is the only type of equilibrium.■

C Comparison to Fearon (1994)

In the main text, I describe three differences between my model and that in Fearon (1994): (i) the addition of sunk costs for delay, (ii) the fact that states can have a positive payoff for fighting, and (iii) relaxing the assumption that war occurs in finite time. In this section, I provide a concise discussion of the differences and how their impact.

C.1 Difference 1: The Introduction of Sunk Costs

As discussed in the introduction in the main text, my model includes sunk costs whereas Fearon (1994) does not. This is a necessary condition for resolved types to be screened so that higher resolve types exit the crisis and go to war earlier. Intuitively, by making delay costly, the sunk costs cause those states with more appealing outside options to abandon crisis negotiations earlier in favor of going to war. By contrast, if there are no sunk costs then delay is cost free for any type who wants to go to war. This would cause resolved types to wait until they were certain that their rival would not concede before choosing to fight as in Fearon (1994).

C.2 Difference 2: Allowing for Types with Positive Payoffs to Fighting:

Fearon (1994) assumed that all types of both countries had negative payoffs for fighting . Formally, $\overline{w}_i \leq 0$. Fearon made this assumption to show that audience costs could lock states into conflict even when severe assumptions against war were imposed. Though I relax this assumption in the main text, and allow for $\overline{w}_i > 0$, screening of resolved types by sunk costs is still possible even if $\overline{w}_i = 0$. However, in general, the lower is $\overline{w}_i$, the higher sunk costs will need to be for screening of resolved types to occur.

To demonstrate this point consider the following series of numerical simulations of the model in the main text. To simplify matters, I will assume that both countries are symmetric such that if sunk costs do screen resolved types, the peaceful phase will proceed directly into the second screening phase. In all the examples that follow, I assume that $\underline{w}_i = -0.8$, $a_i = 0.3$ and that $F_i(\cdot)$ is a uniform distribution. Across example, I will vary $\overline{w}_i$ and k_i to illustrate the degree of sunk costs required to induce screening behavior.

First, consider this first example in the left panel of the Figure (C. 1) designed to replicate Fearon (1994) such that $\overline{w}_i = 0$ and $k_i = 0$. Following the logic described above, without sunk costs resolved types are never screened and the game ends in a peaceful phase regardless of of the value of $\overline{w}_i$. This is illustrated by the figure in the right-panel illustrating the equilibrium that occurs with $\overline{w}_i = 0.8$ and $k_i = 0$.

Second, figure (C. 2) screening is possible once we introduce sunk costs, even if $\overline{w}_i = 0$. In the left panel, k_i has been set to $k_i = 1.2$ and has led to a short screening phase. This value of k_i, four times the value of a_i, is the minimum k_i that can cause screening for these parameter values. when k_i is increased in increments of 0.1. The right panel simulates the model when k_i is increased to a value of 3, ten times the value of audience costs. However, in general screening of resolved states is more easily as the upper bound of each country's possible resolve increases. Figure (C. 3) demonstrates this. In it's right panel, the figure demonstrates that screening is possible when audience costs and sunk costs both have a value of 0.3 for $\overline{w}_i = 0.4$, the minimum value of $\overline{w}_i$ for which this is possible (in increments of 0.1). The left panel demonstrates that a short screening phase is possible when $k_i = 0.5$ for the interim value of $\overline{w}_i = 0.2$. In the latter case this is the minimum value of sunk costs (in incremenets of 0.1) for which screening of resolved types can be achieved for this particular valye of $\overline{w}_i$.

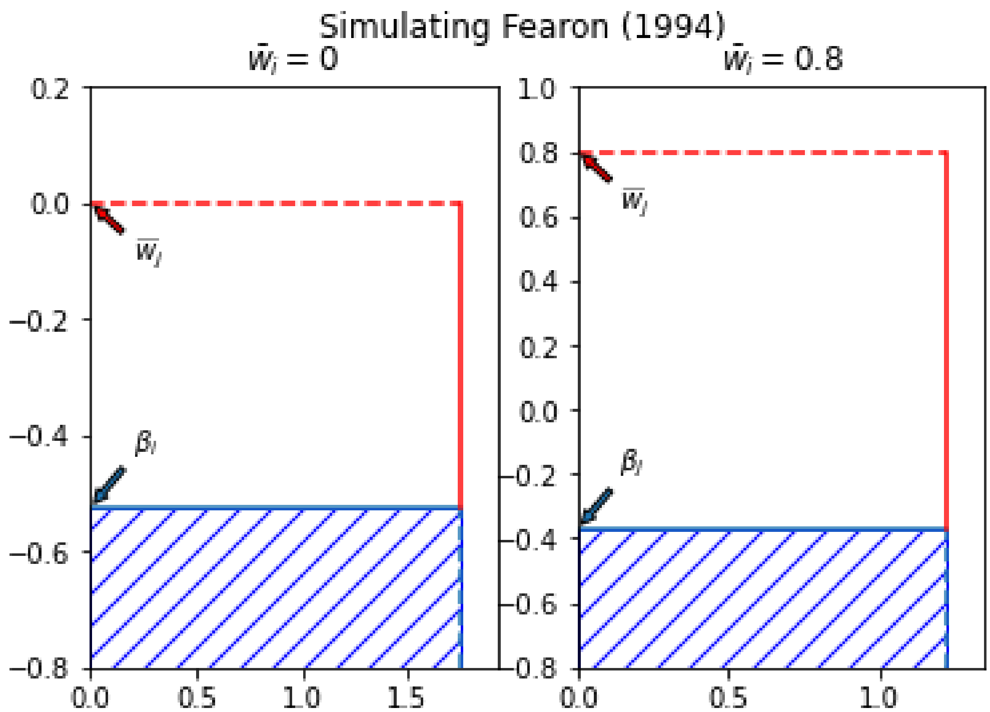

Figure C. 1: **Simulating Fearon (1994):** Both panels present strategies adopted by the two countries when there are no sunk costs for delay. In the left panel $\overline{w}_i = 0$, as in Fearon (1994). In the right panel $\overline{w}_i = 0.8$. So long as delay imposes no sunk costs, the equilibrium will only ever consist of a peaceful phase. In both panels $\underline{w}_i = -0.8$, $a_i = 0.3$ and $F_i(\cdot)$ is a uniform distribution.

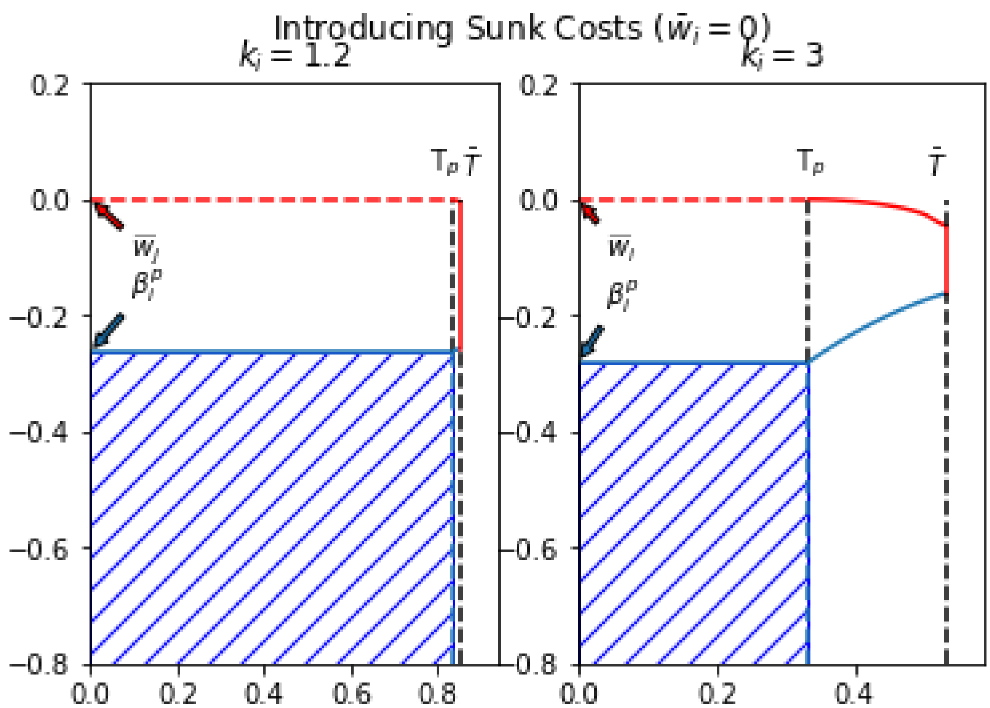

Figure C. 2: **Introducing Sunk Costs ($\overline{w}_i = 0$):** Both panels present strategies adopted by the two countries when $\overline{w}_i = 0$. In the left panel k_i has been set to 1.2, four times the value of audience costs, and a short screening phase can be achieved. The right panel has had k_i set to 3, ten times the value of audience costs and has a relatively longer screening phase. In both panels $\underline{w}_i = -0.8$, $a_i = 0.3$ and $F_i(\cdot)$ is a uniform distribution.

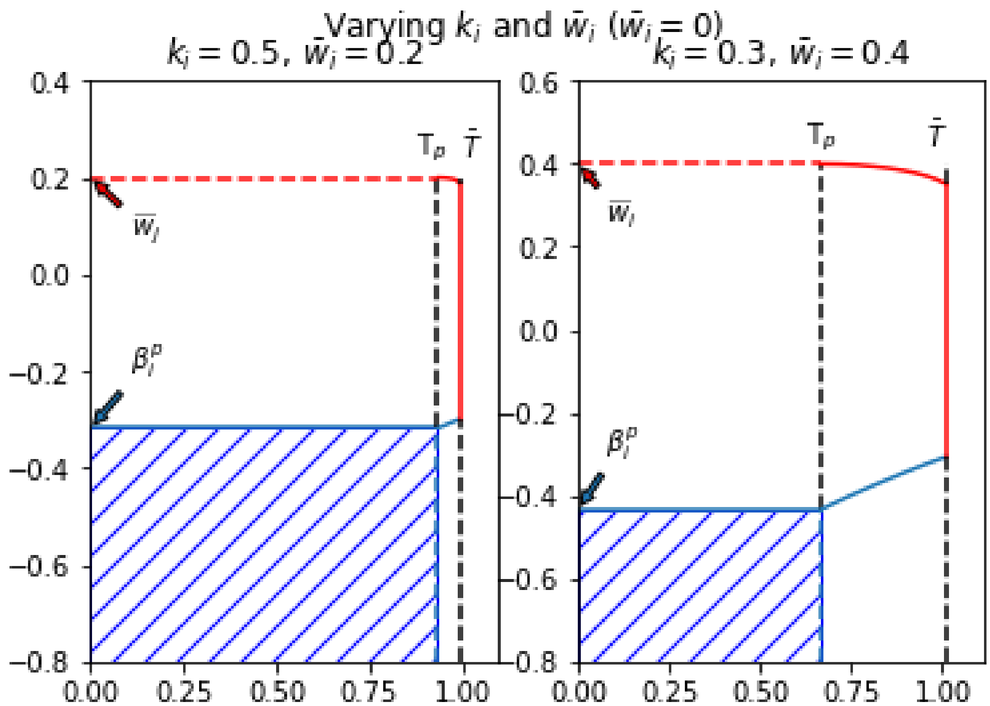

Figure C. 3: **Varying $\overline{w}_i$:** Thios figure demonstrates that when w_i is allowed to increase, screening of resolved types by sunk costs can occur with lower values of k_i. The right-panel has $\overline{w}_i = 0.4$, the minimum value of $\overline{w}_i$ (in incremenets of 0.1) for which audience costs and sunk costs can both have a value of 0.3 and have screening occur. The left panel demonstrates that for the interim value of $\overline{w}_i = 0.2$, a short screening phase is possible with sunk costs having a value of 0.5. In both panels $\underline{w}_i = -0.8$, $a_i = 0.3$ and $F_i(\cdot)$ is a uniform distribution.

C.3 Difference 3: Crises Must End in Finite Time:

In Fearon (1994), the requirement that crises end in finite time was necessary to make sense of the results. As in the model in the main text, in Fearon (1994) there is an endogenous horizon date after which no type concedes. Because Fearon's model did not include sunk costs, the remaining resolved types of the two countries, who are all presumed to have negative payoffs to fighting, have no incentive to go to war and would prefer to delay in perpetuity. To circumvent this problem, Fearon restricts attention to cases where the remaining types past the horizon date play a strategy where they both go to war *at some point* past the horizon date - absent any sunk costs penalizing delay, war could occur at any time $t \in [\overline{T}, \infty)$. Even though the remaining resolved types would prefer not to fight, an equilibrium exists if at any time $t \in [\overline{T}, \infty)$ all remaining types of both countries played (fight, fight) such that neither could change the outcome by unilateral deviation. This is discussed in the section characterizing the horizon date and is the reason why a stalemate need not occur, even the necessary conditions for a stalemate (sufficiently low $\bar{K}_i$ for $(i = 1, 2)$) are met.

As mentioned on page **Fill In**, Fearon (1994) does not acknowledge the potential for a stalemate. This is because if a stalemate occured then all types would opt for a stalemate - since delay is not costly and there are no types who can credibly threaten war, no type would ever concede.[6] Thus Fearon's results are not robust to the introduction of stalemates. By contrast, in the model presented in the main text, the imposition of sunk costs (in the form of the penalty $\overline{K}$) mean that resolved types do not want to delay and that types sufficiently resolved not to concede may prefer war to stalemate.

D Modeling Extension: Introducing a Discount Factor

In the main text, I assume that states stop incurring sunk costs if the war of attrition enters a stalemate and incur a one-time penalty instead. In this section, I assume that countries pay sunk costs indefinitely and introduce a discount factor. To some readers this might allow for a more natural method with which to bound bound payoffs and allow for endogenous stalemates. I show that the results in the main text are robust to the introduction of a discount factor in that they are qualitatively similar to those in the main paper.

Let $e^{-r_i t}$ denote the discount factor incurred by a state who chooses to wait until time t with discount rate $r_i > 0$. We can rewrite the utility function in equation (1) as

$$U_i(t_i, \omega; w_i) = \int_0^{t_i} \int_{\{w_j | \tau_j(t,0)\}} f_j(w)[1 - k_i t]e^{-r_i t} dw + \int_{\{w_j | \tau_j(t,0)\}} f_j(w)[w_i - k_i t]e^{-r_i t} dw dt +$$
$$\mathbb{1}_{\{t_i \neq \infty\}} e^{-r_i t_i} \int_{\{w_j | \tau(t_j, \omega) > t_i\}} f_j(w) dw \left[\mathbb{1}_{\{\omega=1\}} w_i - \mathbb{1}_{\{\omega=0\}} a_i t - k_i t\right] dw - \mathbb{1}_{\{t_i \neq \infty\}} \frac{k_i}{r_i} \int_{\{w_j | \tau_j(\infty, \omega)\}} f_j(w) dw \tag{C. 1}$$

[6] This assumes that $\bar{K}_i = 0$ in Fearon's model, in the spirit of no costs for delay.

In what follows, I replicate each of the propositions in the main text. The lemmas replicate in a much more straightforward manner, and will not be restated.

D.1 The Peaceful Phase with a Discount Factor

As in the main text, the game begins with a peaceful phase during which neither country goes to war. Conceding types must play a mixed strategy $Q_i(t)$ during this phase and Lemma 2 from the main text continues to apply.

The expected utility function for a type that concedes during the peaceful phase is

$$U_i(t_i, 0; w_i | t < T^p) = F_j(\beta_j^p) \int_0^{t_i} q_j(t)[1 - k_i t]e^{-r_i t} dt - [1 - Q_j(t_i)F_j(\beta_j^p)])[a_i t_i + k_i t_i]e^{-r_i t} \quad \text{(C. 2)}$$

Taking the first-order condition, we can characterize the modified hazard rate as

$$\frac{F_j(\beta_j^p)q_j(t)}{1 - Q_j(t_i)F_j(\beta_j^p)} = \frac{[a_i + k_i][1 - r_i t]}{1 + a_i t} \quad \text{(C. 3)}$$

The hazard rate reveals that the introduction of the discount factor causes the rate of concession to slow down as the war of attrition progresses. This is required to adjust for the way in which the discount factor ameliorates the penalties from the audience costs and sunk costs that accumulate with delay.

To determine when the most resolved type of either country will go to war, we can rewrite equation (A. 7), the expected utility function for a type going to war during the peaceful phase, as

$$U_i(t_i, 1; \overline{w}_i | t \leq T^p) = F_j(\beta_j^p) \int_0^{t_i} q_j(t)[1 - k_i t]e^{-r_i t} dt + [1 - F_j(\beta_j^p)Q_j(t_i)][\overline{w}_i - k_i t_i]e^{-r_i t_i} \quad \text{(C. 4)}$$

Taking the first order condition we find that type $\overline{w}_i$'s utility will be maximized when

$$\frac{F_j(\beta_j^p)q_j(t_i)}{1 - Q_j(t_i)F_j(\beta_j^p)} = \frac{k_i[1 - r_i t_i] + r_i \overline{w}_i}{1 - \overline{w}_i} \quad \text{(C. 5)}$$

This mirrors the calculations in the main text with the addition of an $r\overline{w}_i$ term in the numerator on the right-hand side. This reflects that the discount factor makes the most resolved type more impatient and willing to exit earlier since the discount factor reduces the gains from a delayed war. Substituting equation (C. 3) into the expression, we find that the most resolved type of country i will choose to go to war when

$$\overline{w}_i = \frac{1 - \frac{k_i}{a_i + k_i}[1 + a_i t_i]}{1 + \frac{r_i[1 + a_i t_i]}{[a_i + k_i][1 - r_i t]}} \quad \text{(C. 6)}$$

This is similar to the expression in equation (4) in the main text, with the addition of the terms in the denominator on the right-hand side. These additional terms are strictly greater than 1 for an $r > 0$. Since the numerator is strictly decreasing in t, it follows that the discount factor causes the peaceful phase to end sooner.

As in the main text, the peaceful phase will end whenever there is a single type that is no longer willing to delay going to war. Defining T^p analogously to its definition in the main text, we can restate Propositions 1 and 2 as follows

Proposition C. 1 *Let* $T_1 = \min\{T^p, \overline{T}\}$

(i) Types $w_i \in [\underline{w}_i, \beta_i^p]$ *(*$i = 1, 2$*) concede during* $[0, T^1]$ *and play strategy* $Q_i(t)$ *as defined in equation (C. 3).*

(ii) No type goes to war during the interval $[0, T^1)$*.*

(iii) Countries posterior beliefs during this period are given by equation (A. 6).

Proof: To replicate the proof of Proposition 1, we need to show that the utility function in equation (C. 4) is concave in t. Taking the second derivative, we find that this will be true if

$$F_j(\beta_j^p)\frac{dq_j(t_i)}{dt}[1-\overline{w}_i] - k_i q_j(t_i) + r_i q_j(t_i)[\overline{w}_i - k_i t_i] + r_i k_i[1 - F_j(\beta_j^p)Q_j(t_i)] < 0$$

As in the main text, to find the value of $F_j(\beta_j^p)$ we must rearrange the hazard rate in equation (C. 3) to find that

$$q_j(t)[1 + a_i t_i] = [a_i + k_i][1 - rt][1 - F(\beta^p)Q_j(t)]$$

so that taking the derivative with respect to t, we find that

$$F_j(\beta_j^p)\frac{dq_j(t)}{dt} = \frac{-r_i[a_i + k_i][1 - F(\beta^p)Q_j(t)] - q_j(t)[2a_i + k_i] + r_i q_j(t)[a_i t + k_i t]}{1 + a_i t}$$

Which we can substitute back into the second order condition

$$[1-\overline{w}_i]\left[\frac{-r_i[a_i + k_i][1 - F_j(\beta_j^p)Q_j(t_i)] - q_j(t_i)[2a_i + k_i] + r_i q_j(t_i)[a_i t_i + k_i t_i]}{1 + a_i t_i}\right]$$
$$-k_i q_j(t_i) + r_i q_j(t_i)[\overline{w}_i - k_i t_i] + r_i k_i[1 - F_j(\beta_j^p)Q_j(t_i)] < 0$$

Rearranging and substituting for the hazard rate from equation (C. 3), we find that

$$[1-\overline{w}_i]\left[-r_i[a_i + k_i] - \frac{[a_i + k_i]^2[1 - r_i t_i]^2}{1 + a_i t_i} - a_i\frac{[a_i + k_i][1 - r_i t_i]}{1 + a_i t_i}\right]$$
$$-k_i[a_i + k_i][1 - r_i t_i] + r_i[\overline{w}_i - k_i t_i][1 - r_i t_i][a_i + k_i] + r_i k_i[1 + a_i t_i] < 0$$

Rearranging and substituing for $\bar{w}_i$ from equation (C. 6), we are left with

$$-k_i[a_i+k_i][1-r_it_i]-r_ik_it_i[1-r_it_i][a_i+k_i]-r_i^2k_it_i[1+a_it_i]$$
$$-k_ir_i[1+a_it_i]-k_i[a_i+k_i][1-r_it_i]-k_ir_i[1-r_it_i][1+a_it_i]$$
$$-a_ik_i[1-r_it_i]-r_ia_i-\frac{r_i^2a_i[1+a_it_i]}{[a_i+k_i][1-r_it_i]}<0$$

which, given that $1-r_it_i$ must be positive to satisfy Lemma 2, is negative as desired. ■

D.2 The First Screening Phase with a Discount Factor

As in the main text, the peaceful phase is followed by a first screening phase where the resolved types of one country are screened by sunk costs while the other country's resolved types continue to delay going to war. Without loss of generality let country j be the country whose types go to war during the first screening phase. As a result, the unresolved types of country i are screened by the threat of war, while the unresolved type of country j continue to play a mixed strategy.

As in the main text, the mixed strategy played by country i must continue to concede at the same rate as in the peaceful phase, but must now does so as a monotonically increasing pure strategy

$$\frac{f_i(\tau(t,0))\tau_i'(t,0)}{1-F_i(\tau_i(t,0))}=\frac{[a_j+k_j][1-r_jt]}{1+a_jt_j} \tag{C. 7}$$

That country i's rate of concession does not change, implies that resolved types of country j face an identical trade-off as to when to exit as they did in the peaceful phase, so that (C. 6) continues to determine when resolved types of country j escalate in the first screening phase. Taking the derivative of (C. 6) with respect to t_j we find that the inverse of the strategy function for resolved types of country j is given by

$$\frac{d\bar{w}_j^t}{dt}=\frac{-\frac{k_ja_j}{a_j+k_j}\left[1+\frac{r_j[1+a_jt_j]}{[a_j+k_j][1-r_jt_j]}\right]-\left[1-\frac{k_j}{a_j+k_j}[1+a_jt_j]\right]\times\frac{r_ja_j[a_j+k_j[1-r_jt_j]+r_j^2[a_j+k_j][1+a_jt_j]}{[a_j+k_j]^2[1-r_jt_j]^2}}{\left[1+\frac{r_j[1+a_jt_j]}{[a_j+k_j][1-r_jt_j]}\right]^2} \tag{C. 8}$$

which is still negative and represents a faster rate of decrease than in equation (A. 13), a result of the decrease in expected utility from delay that is attributable to the discount factor.

The utility function for an unresolved type of country i conceding during the first screening phase is given by

$$\begin{aligned}U_i(t_i,0;w_i|T^p\le t_i\le T^f)&=F_j(\beta_j^p)\int_0^{T^p}q_j(t)[1-k_it]e^{-rt}dt\\+[F_j(\beta_j^f)-F_j(\beta_j^p)]\int_{T^p}^{t_i}s_j(t)[1-k_it]e^{-rt}dt&-\int_{T^p}^{t_i}f_j(\tau_j(t,1))\tau_j'(t,1)[w_i-k_it]e^{-rt}dt\\-[F_j(\tau_j(t_i,1))-(F_j(\beta_j^f)-F_j(\beta_j^p))S_j(t)&-F_j(\beta_j^p)][a_it_i+k_it_i]e^{-rt_i}\end{aligned} \tag{C. 9}$$

Taking the first order condition, we find that the hazard rate for concessions for country j is given by

$$\begin{aligned} \frac{[F_j(\beta_j^f) - F_j(\beta_j^p)]s_j(t)}{F_j(\tau_j(t,1)) - (F_j(\beta_j^f) - F_j(\beta_j^p))S_j(t) - F_j(\beta_j^p)} &= \\ \frac{f_j(\tau_j(t,1))\tau_j'(t,1)}{F_j(\tau_j(t,1)) - (F_j(\beta_j^f) - F_j(\beta_j^p))S_j(t) - F_j(\beta_j^p)} \times \frac{\underline{w}_i^t + a_i t_i}{1 + a_i t_i} &+ \frac{[a_i + k_i][1 - r_i t_i]}{1 + a_i t_i} \end{aligned} \tag{C. 10}$$

Finally, we can rewrite the utility function for the most resolved type of country i as follows

$$\begin{aligned} U_i(t_i, 1; \overline{w}_i | T_p \leq t_i \leq T^f) &= F_j(\beta_j^p) \int_0^{T^p} q_j(t)[1 - k_i t]e^{-rt}dt \\ +[F_j(\beta_j^f) - F_j(\beta_j^p)] \int_{T^p}^{t_i} s_j(t)[1 - k_i t]e^{-rt}dt &- \int_{T^p}^{t_i} f_j(\tau_j(t,1))\tau_j'(t,1)[\overline{w}_i - k_i t]e^{-rt}dt \\ +[F_j(\tau_j(t_i,1)) - (F_j(\beta_j^f) &- F_j(\beta_j^p))S_j(t) - F_j(\beta_j^p)][\overline{w}_i - k_i t_i]e^{-rt_i} \end{aligned} \tag{C. 11}$$

Taking the first order condition, we find that the most resolved type of country i will go to war when country i concedes at the same rate as the right-hand side of equation (C. 5). Substituting for the left-hand side hazard rate from (C. 10), we find that the most resolved type of country i will go to war when

$$\overline{w}_i = \frac{1 - \frac{k_i'[1-r_i t_i][1+a_i t_i]}{\frac{f_j(\tau_j(t_i,1))\tau_j'(t_i,1)}{F_j(\tau_j(t_i,1)) - (F_j(\beta_j^f) - F_j(\beta_j^p))S_j(t_i) - F_j(\beta_j^p)}[\underline{w}_i + a_i t_i] + [a_i + k_i][1 - r_i t_i]}}{1 + \frac{r_i \overline{w}_i[1+a_i t_i)]}{\frac{f_j(\tau_j(t_i,1))\tau_j'(t_i,1)}{F_j(\tau_j(t_i,1)) - (F_j(\beta_j^f) - F_j(\beta_j^p))S_j(t_i) - F_j(\beta_j^p)}[\underline{w}_i + a_i t_i] + [a_i + k_i][1 - r_i t_i]}} \tag{C. 12}$$

We can therefore restate Proposition 3 and 4 as

Proposition C. 2 *Let $T^2 = \min\{T^f, \overline{T}\}$. If there exists a $T^p < \overline{T}$, then during $[T^p, T^2]$, the following must hold*

(i) Types w_i who concede play strategy $\tau_j(\cdot, 0)$ as defined in equation (C. 7).

(ii) Types w_j who concede must form a connected interval and play strategy $S_j(t)$ as given by equation (C. 10).

(iii) Types w_j who go to war play $\tau_j(\cdot, 1)$ as defined in equation (C. 8).

(iv) Country i does not go to war during $[T^p, T^2)$.

(v) County i's posterior beliefs posterior beliefs during this period are given by (A. 14) and beliefs for country j are given by (A. 15).

where T^f is the analogue to its definition in the main text. The proof of this claim is identical to that of Proposition 3 and 4 in the main text and is therefore omitted.

D.3 The Second Screening Phase with a Discount Factor

As in the main text, the final phase of the war of attrition sees the resolved types of both countries screened by sunk costs and the unresolved types of both countries screened by the threat of war. With the introduction of a discount factor we can we rewrite the utility function for a type that exits during the second screening phase as

$$\begin{aligned} U(t_i,\omega;w_i|T^f \leq t_i < \overline{T}) = F_j(\beta^p)\int_0^{T^P} q_j(\ell)[1-k_it]e^{-r_it}dt \\ +[F_j(\beta^f)-F_j(\beta^p)]\int_{T^p}^{T^f} s(t)[1-k_it]e^{-r_it}dt + \int_{T^f}^{t_i} f_j(\tau_j(t,0))\tau_j'(t,0)[1-k_it]e^{-r_it}dt \\ -\int_{T^p}^{t_i} f(\tau_j(t,1))\tau_j'(t,1)[w_i-k_it]e^{-r_It}dt \\ +(F_j(\tau_j(t_i,0))-F_j(\tau_j(t_i,1)))\max\{w_i-k_it_i,-a_it_i-k_it_i\}e^{-r_it_i} \end{aligned} \tag{C. 13}$$

Taking the first order condition for a type that goes to war we find that the hazard rate for concessions must be given by

$$\frac{f_j(\tau_j(t,0))\tau_j'(t,0)}{F_j(\tau_j(t,0))-F_j(\tau_j(t,1))} = \frac{k_i[1-r_it]+r_i\overline{w}_i^t}{1-\overline{w}_i^t} \tag{C. 14}$$

and that the hazard rate determining when resolved types go to war is given by

$$\frac{f(\tau_j(t,1))\tau_j'(t,1)}{F_j(\tau_j(i,1))-F_j(\tau_j(t,0))} = \frac{k_i[1-r_it_i][\overline{w}_i^t+a_it_i]+r\overline{w}_i[1+a_it_i]-a_i[1-r_it_i][1-\overline{w}_i^t]}{[1-\overline{w}_i^t][\underline{w}_i^t+a_it_i]} \tag{C. 15}$$

These closely resemble the results in the main text, though the hazard rates for both concession and escalation are accelerated. For resolved types to be willing to delay, the conceding types must compensate them for the decrease in payoff to escalation they will incur by not going to war. For conceding types, the rate of escalation must increase to account for the amelioration of the risks and costs associated with delay that result from the discount factor.

We can therefore restate proposition 5 as

Proposition C. 3 *If there exists a $T^f < \overline{T}$, then during $[T^f, \overline{T}]$ the following must hold,*

(i) Types w_i $(i = 1, 2)$ who concede play strategy $\tau_i(\cdot, 0)$ as defined in equation (C. 14).

(ii) Types w_i $(i = 1, 2)$ who go to war play strategy $\tau_i(\cdot, 1)$ as defined in equation (C. 15)

(iii) Country i's $(i = 1, 2)$ posterior beliefs during $[T^f, \overline{T})$ are given by (14).

Proof: To replicate Proposition 5, we need to show that the utility function in equation (C. 13) is concave in t_i for both a type going to war and a conceding type. Taking the second derivative we find that this will be true for a type going to war if

$$\frac{df_j(\tau_j(t_i,0))\tau_j'(t_i,0)}{dt}[1-\overline{w}_i] - f_j(\tau_j(t_i,1))\tau_j'(t_i,1))[k_i(1-r_it_i)+r\overline{w}_i]$$
$$+f_j(\tau_j(t_i,0))\tau_j'(t_i,0))[k_i(1-r_it_i)+r\overline{w}_i] + r_ik_i[F_j(\tau_j(t_i,0)) - F_j(\tau_j(t_i,1))] < 0$$

We cannot proceed without additional information regarding the $\frac{df_j(\tau_j(t_i,0))\tau_j'(t_i,0)}{dt}$ term. To solve for this term we can rearrange the hazard rate in equation (C. 14) as

$$f_j(\tau_j(t,0))\tau_j'(t,0)[1-\overline{w}_i] = k_i[1-r_it][F_j(\tau_j(t,0)) - F_j(\tau_j(t,1))]$$
$$+r_i\overline{w}_i^t[F_j(\tau_j(t,0)) - F_j(\tau_j(t,1))]$$

Taking the derivative we find that

$$\frac{df_j(\tau_j(t,0))\tau_j'(t,0)}{dt}[1-\overline{w}_i^t] = \frac{d\overline{w}_i^t}{dt}f_j(\tau_j(t,0))\tau_j'(t,0)$$
$$+r\frac{d\overline{w}_i^t}{dt}[F_j(\tau_j(t,0)) - F_j(\tau_j(t,1))] - r_ik_i[F_j(\tau_j(t,0)) - F_j(\tau_j(t,1))]$$
$$+f_j(\tau_j(t,1))\tau_j'(t,1)[k_i(1-r_it)+r_i\overline{w}_i^t] - f_j(\tau_j(t,0))[k_i(1-r_it_i)+r_i\overline{w}_i^t]$$

Substituting back into the second order condition we are left

$$\frac{d\overline{w}_i}{dt}f_j(\tau_j(t_i,0))\tau_j'(t_i,0) + r\frac{d\overline{w}_i}{dt}[F_j(\tau_j(t_i,0)) - F_j(\tau_j(t_i,1))] < 0$$

which is true given that $\overline{w}_i^t$ is strictly decreasing in t.

Similarly the second derivative for a conceding type will be given by

$$\frac{df_j(\tau_j(t_i,0))\tau_j'(t_i,0)}{dt}[1+a_it_i] - \frac{df_j(\tau_j(t_i,1))\tau_j'(t_i,1)}{dt}[\underline{w}_i+a_it_i]$$
$$+f_j(\tau_j(t_i,0))\tau_j'(t_i,0)\{a_i+[a_i+k_i][1-r_it_i]\} - f_j(\tau_j(t_i,1))\tau_j'(t_i,1)\{a_i+[a_i+k_i][1-r_it_i]\}$$
$$+r[F_j(\tau_j(t_i,1)) - F_j(\tau_j(t_i,0))][a_i+k_i] < 0$$

Note that the hazard rate that determines when resolved types go to war is first derived by taking the first order condition for a conceding type and rearranging to get

$$\frac{f_j(\tau_j(t,0))\tau_j'(t,0)}{F_j(\tau_j(t,1)) - F_j(\tau_j(t,0))} = \frac{f_j(\tau_j(t,1))\tau_j'(t,1)}{F_j(\tau_j(t,1)) - F_j(\tau_j(t,0))}\frac{\underline{w}_i^t+a_it}{1+a_it} + \frac{[a_i+k_i][1-r_it]}{1+a_it_i}$$

rather than substituting in for the left-hand side with equation (C. 15) and getting (C. 15), we rearrange to show that it is possible to express $f_j(\tau_j(t_i,1))\tau_j'(t_i,1)$ as

$$f_j(\tau_j(t,1))\tau_j'(t,1)[\underline{w}_i^t+a_it] = f_j(\tau_j(t,0))\tau_j'(t,0)[1+a_it]$$
$$-[a_i+k_i][1-r_it][F_j(\tau_j(t,1)) - F_j(\tau_j(t,0))]$$

Taking the derivative with respect to t we find that

$$\frac{f_j(\tau_j(t,1))\tau_j'(t,1)}{dt}[\underline{w}_i^t + a_i t] = \frac{dt f_j(\tau_j(t,0))\tau_j'(t,0)}{dt}[1 + a_i t]$$
$$-f_j(\tau_j(t,1))\tau_j'(t,1)\left[\frac{d\underline{w}_i^t}{dt} + a_i + [a_i + k_i][1 - r_i t]\right] + f_j(\tau_j(t,0))\tau_j'(t,0)\left[a_i + [a_i + k_i][1 - r_i t]\right]$$
$$+r_i[a_i + k_i][F_j(\tau_j(t,1)) - F_j(\tau_j(t,0))]$$

Substituting these terms back into the second order condition, we are only left with

$$f_j(\tau_j(t_i,1))\tau_j'(t_i,1)\frac{d\underline{w}_i}{dt} < 0$$

which holds given that $\tau_j'(t_i,1)$ is strictly negative and $\frac{d\underline{w}_i}{dt}$ is strictly positive ■

D.4 The Horizon Date with a Discount Factor

As in the main text, the game ends once all unresolved types have conceded. At this point, any type of either country that intends to escalate does so as they can no longer justify delay. If a forward the valuation that a country places on paying sunk costs forever is sufficiently low, then moderately resolved types may prefer to remain in the war of attrition and incur sunk costs than pay the audience costs required to concede or the costs of fighting required for going to war.

If the war of attrition ends in a stalemate, then a type that remains in the war of attrition forever can expect to pay $\int_{\overline{T}}^{\infty} k_i' e^{-r_i t} dt = \frac{k_i}{r_i}$ sunk costs. Therefore the least resolved type to escalate the crisis will be type $\alpha_i \equiv -\frac{k_i}{r_i}$ and the lest resolved type to remain in the war of attrition forever will be the type denoted β_i for which the following equation holds with equality

$$\frac{F_j(\overline{w}_j^{\overline{T}}) - F_j(\alpha_j)}{F_j(\overline{w}^t) - F_j(\beta_j)}\beta_i - \frac{F_j(\alpha_j) - F_j(\beta_j)}{F_j(\overline{w}_j^{\overline{T}}) - F_j(\beta_j)}\frac{k_i}{r} = -a_i t \qquad \text{(C. 16)}$$

If the war of attrition ends with all types exiting, then the assignment of types to exit strategies is exactly as it is in the main text. We can therefore restate Proposition 6 as follows. A proof of the Proposition is omitted as it follows the arguments in the main text.

Proposition C. 4 *The game can end at any phase with the following determining countries' choice of exit strategy*

(i) All types exit: Types $w_i \in [\underline{w}_i, \beta_i)$ concede and types $w_i \in [\beta_i, \overline{w}_i]$ escalate for $\beta_i = -a_i t$.

(ii) Some types remain in forever: If $\frac{k_i}{r} < a_i t$ for both $i = 1, 2$, then there exists in equilibrium where types $w_i \in [\underline{w}_i, \beta_i)$ concede, types $w_i \in [\beta_i, \alpha_i)$ remain in the war of attrition forever, and types $w_i \in [\alpha_i, \overline{w}_i]$ escalate for $\alpha_i = -\frac{k_i}{r}$ and β_i as defined in (C. 16).

D.5 Ruling Out Alternative Equilibria with a Discount Factor

Once again, it is possible to show that any equilibrium which has types exit after time $t = 0$ must have countries exit according to the sequence of phases described above. We can therefore restate Proposition B.1 as

Proposition C. 5 *Any equilibrium that does have the war of attrition end at $t = 0$, must have countries play strategies according to Propositions C.1, C.2, and C.3.*

Proof: As in the Proof of Proposition B.1, it is the monotonicity requirements established in Lemmas 3 and 4 which rule out any alternative equilibrium. We know that equation (C. 15) must be the hazard rate that determines when counties go to war during any interval in which both countries go to war. Note that the denominator in that expression

$$[1 - \overline{w}_i][\underline{w}_i + a_i t]$$

is negative. Therefore, for $\tau_j'(t_i, 1)$ to be decreasing it must be the case that the numerator is positive. Note that the numerator

$$k_i[1 - r_i t_i][\overline{w}_i^t + a_i(t_i)] + r\overline{w}_i^t[1 + a_i t_i] - a_i[1 - r_i t_i][1 - \overline{w}_i^t]$$

will be positive so long as

$$\overline{w}_i^t > \frac{1 - \frac{k_i}{a_i + k_i}[1 + a_i t_i]}{1 + \frac{r_i[1 + a_i(t_i)]}{[a_i + k_i][1 - r_i t_i]}}$$

However, from the discussion of the peaceful phase we know that this will not be true for any type before T^p. This rules out any equilibrium that has an interval where both countries have types going to war before T^p.

The remainder of the proof is identical to that of proposition B.1.■

E Modeling Extension: Allowing for Signaling Prior to $T = 0$:

This section explores an extension in which one of the two countries has the ability to engage in sunk cost signaling before entering the war of attrition. That is, I consider an extension in which one country is granted the ability to "instanteously" produce large amounts of sunk cost signals prior to $t = 0$, mirroring costly signaling as in Fearon (1997), so that any learning that occurs from this signaling occurs before the war of attrition begins. I show that the allowing for such signaling does not eliminate the war of attrition - delay must still occur with positive probability. This demonstrates that the dynamic screening results are not an artificial consequence of a choice to prohibit states from engaging costly signaling in the classical sense.

Specifically, I show that when granted the ability to engage in sunk cost signaling in this way, there must be at least some types of $w_i \in [\underline{w}_i, 0)$ who will prefer to mimick any message sent by types $w_i \in [0, \overline{w}_i]$. This is because less resolved states' decision to invest more in sunk costs in the model is not driven by the game form but by their weaker incentive to spend resources to avoid war. However, so long as types $w_i < 0$ participate in the war of attrition, then both countries have some hope for a concession if they delay and Proposition (B. 1) and (B. 2) apply.

For simplicity's sake we will begin by considering the case where country i can send some costless message $m \in M$ where M is some compact metric space. We will define a signal as effective it induces all types of country j to exit immediately at the start of the war of attrition.

Definition D. 1 *A signal is an effective signal if in response to country i's message, types $w_j \leq 0$ concede at $t = 0$ and types $w_j > 0$ fight at $t = 0$.*

If a signal is effective, then best response of a resolved type of country i is to delay exit until after country j has exited since they incur no cost for delaying war and may obtain a concession. In effect, when a signal is effective, then the war of attrition is not played and the game ends immediately at $t = 0$.

The following result demonstrates that a signal can only be effective if it removes all doubt from country i's mind that country i is unresolved.

Lemma D. 1 *A costless or sunk cost signal m is an effective signal if and only if it induces a posterior belief that*

$$G_j(w_i|m) = \begin{cases} 0 \ \textit{if} \ w_i \leq 0 \\ \frac{f(w_i)}{1-F(0)} \textit{if} w_i > 0 \end{cases} \tag{D. 1}$$

The intuition for this result is straightforward. If country i is resolved with certainty, then country j has no reason to delay exit and must do so immediately. However, if country i is possibly unresolved, then resolved types of country j have an incentive to delay and await its concession. By subgame perfection, any unresolved type of country i will use that opportunity to concede an arbitrarily short after time $t = 0$.

Note there can be a large number of signals that constitute an effective signal. For example, if there exist two messages m', m'' each sent only by types w'_i, w''_i respectively where $w'_i, w''_i \in (0, \overline{w}_i]$, then both will constitute effective signals. However, from the perspective of resolved types of country i the two messages are payoff equilivalent. It is without loss of generality to assume that an effective signal sent by country i therefore induces the beliefs listed in equation (D. 1).

It turns out that the countries cannot cause the war of attrition to end at time $t = 0$ and produce an effective signal with sunk cost signals.

Lemma D. 2 *No costless message can constitue an effective signal.*

Intuitively, this fails because there are types $w_i < 0$ who will become locked into conflict should they not send the effective signal and have to play the war of attrition game. These types are better off mimicking the signal rather than paying the costs of delay.

Now assume that messages are not costless and that sending a message $m \in \mathbb{R}^+$. Such that a country's expected utility from sending message is now given by equation (1) with an additional

$-m$ for sending message m prior to the war of attrition. The following result demonstrates that this cannot prove an effective message.

Proposition D. 1 *No sunk cost message can constitue an effective signal.*

The intuition for this result, is similar to that for Lemma (D. 2). Types $w_i < 0$ who will become locked into conflict by sending an ineffective signal would be more likely to fight by sending the ineffective signal. These types are more incentivized to avoid war since they stand more to lose from fighting, therefore there exist types $w_i < 0$ would become locked into fighting under an alternate message who would prefer to spend more to produce an effective signal.

E.1 Proof of Lemma (D. 1)

We will begin by proving that a signal is effective only if it induces the posterior in equation (D. 1). That is, we will show that if Country j believes that it is possible that Country i may be a type $w_i < 0$, there will be unresolved and resolved types of country j remaining in the crisis after $t = 0$.

First, observe that if a signal is effective, then there is a unique equilibrium in which Country i chooses to go to war at some time $t > 0$, and all types of country j exit immediately. Resolved types of country j have no profitable deviation since they can expect no concession by delaying and will only pay additional sunk costs. Similarly unresolved types have no benefit to delay concession as they will only pay sunk costs and accumulate additional audience costs. Moreover, if they delay too long than Country i will fight. Finally, country i's specific choice of exit time is irrelevant so long as $t_i > 0$, since country i will exit before its choice is realized. Therefore, the only relevant deviation to consider is one in which Country i chooses to go to war at time $t = 0$, in which case it will have an expected utility of

$$U_i(0, 1|w_i) = F_j(0)\frac{1}{2}(1 + w_i) + (1 - F_j(0))w_i$$

which is strictly smaller than

$$U_i(t_i, 1|w_i) = F_j(0) + (1 - F_j(0))w_i$$

which is its expected utility for an exit time $t_i > 0$. This is sufficient to prove that a message that induces the belief in equation (D. 1) must be an effective signal.

Next, we will prove that if a message does not induce the beliefs in equation j, then it cannot be an effective signal and must have types of country j exit after $t = 0$. Suppose not. That is suppose that country j believed that there was some probability $z > 0$ such that country $i's$ type was $w_i < 0$. Moreover, suppose that all types of country j exited at time $t = 0$ either by going to war or conceding. As above, resolved types of country i with $w_i > 0$ would choose to exit by going to war after $t = 0$ and have no incentive to deviate from this action. Types $w_i < 0$ must make a choice. They can either choose to concede at time $t = 0$ for an expected utility of

$$U_i(0, 0|w_i) = \frac{1}{2}F_j(0) + (1 - F_j(0))\frac{w_i}{2}$$

or they can choose to exit after time $t > 0$ in exchange for a utility of

$$U_i(t_i, 1|w_i) = F_j(0) + (1 - F_j(0))w_i$$

Lt $\hat{w}_i$ denote the cutoff type that is indifferent between the two options. If there is a nondegnerate set of types $w_i < \hat{w}_i$ that prefers to concede at time $t = 0$. Then exiting at time $t = 0$ is no longer an equilibrium as all types of country j have a strictly profitable deviation to delaying exit by some arbitrarily small $\epsilon > 0$ and checking to see if country i concedes. This would be sufficient to prove that the message is not an effective signal in this instance.

Therefore, all that remains to show is that there will exist types of country j with an incentive to deviate and concede at a time $t > 0$ when they believe that there exist a nondegenerate set of types $w_i \in [\hat{w}_i, 0]$ in the war of attrition at time $t = 0$. First, observe that types of country i have selected a time to exit $t_i > 0$. If we require that i's strategies be subgame perfect, then this cannot be an equilibrium. Suppose that all resolved types of country j go to war at time t delay war until some arbitrary time $t_j > 0$ that may be arbitrarily small. Then there must exist some non-degenerate subset of types $w_i < 0$ who would prefer to concede immediately after $t = 0$, once a mass of types of country j have conceded. Therefore, this cannot be an equilibrium. ■

E.2 Proof of Lemma (D. 2)

Suppose not. That is, suppose that there existed a costless message that constituted an effective signal. The expected utility for type $w_i = 0$ from sending that signal is simply $F_j(0)$. Unresolved types of country i who do not send the effective message must either concede immediately or play a war of attrition game that goes past $t = 0$, per Lemma (D. 1). Note that there must exist a type $w_i = \epsilon$ for some ϵ that is arbitrarily small, that must get locked into fighting by audience costs in a war of attrition with positive length. This type's expected utility from participating in the war of attrition must be strictly less than from mimicking the costless signal since some types of country j will become locked into fighting as well and it must pay for the costs of delay. Thus all types in the interval $w_i \in [-\epsilon, 0)$ must also send the effective signal. A contradiction. ■

E.3 Proof of Proposition (D. 1)

Suppose that there existed a costly signal m^* that constituted an effective signal. Moreover, let m' denote the ineffective signal sent by a type $w_i = -\epsilon$ for some artbitarily small positive ϵ. Per lemma (D. 1), message m' must lead to a war of attrition with positive delay. Following the logic of Reich (2023) for this to be incentive compatible, the most type $w_i = 0$ would be willing to spend on the effective signal is given by

$$B(0) \equiv m^* - m' = F_j(0) - U_i(m'|0)$$

where $U_i(m'|0)$ is type $w_i = 0$ expected utility from sending signal m'. Note that because the equilibrium following signal m' has a higher probability of war since types $w_j < 0$ must be locked in by audience costs. It follows that type $B(-\epsilon)$, how much type $w_i = -\epsilon$ would be willing to spend on signal m^* instead of m' is strictly greater than $B(0)$. To see this simply observe that for a type who is going to become locked in by audience costs

$$\frac{U_i(m^*|w_i)}{\partial w_i} = 1 - F_j(0)$$

and that

$$\frac{U_i(m^*|0)}{\partial w_i} = 1 - F_j(\beta_j)$$

It follows that there must exist a type $w_i = -\epsilon$ who must strictly prefer to deviate from message m_i' to m^*, since they have a lower payoff to fighting and message m^* produces a lower probability of war than message m'.

References

Ashworth, Scott and Ethan Bueno de Mesquita (2006). "Monotone Comparative Statics for Models of Politics". *American Journal of Political Science* 50.1, pp. 214–231.

Fearon, James D. (1994). "Domestic Political Audiences and the Escalation of International Disputes". *American Political Science Review* 88.3, pp. 577–592.

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

Hendricks, Ken, Andrew Weiss, and Charles Wilson (1988). "The War of Attrition in Continuous Time with Complete Information". *International Economic Review* 29.4, pp. 663–680.

Reich, Noam (Mar. 2023). "When Can States Signal with Sunk Costs?" *Unpublished Manuscript.*

Takahashi, Yuya (2015). "Estimating a War of Attrition: The Case of the US Movie Theater Industry". *American Economic Review* 105.7, pp. 2204–2241.