Political Science Research and Methods Vol 3, No. 2, 281–307 May 2015

doi:10.1017/psrm.2014.23

War with Crazy Types

AVIDIT ACHARYA AND EDOARDO GRILLO

This article introduces a model of war and peace in which leaders believe that their adversaries might be crazy types who always behave aggressively, regardless of whether it is strategically optimal to do so. In the model, two countries are involved in a dispute that can either end in a peaceful settlement or escalate into 'limited war' or 'total war.' If it is common knowledge that the leaders of the countries are strategically rational, then the only equilibrium outcome of the model is peace. Yet if a leader believes that there is a chance that her adversary is a crazy type, then even a strategically rational adversary may have an incentive to adopt a madman strategy in which he pretends to be crazy. This leads to limited war with positive probability, even when both leaders are strategically rational. The article shows that despite having two-sided incomplete information, the model has a generically unique equilibrium. Moreover, the model identifies two countervailing forces that drive equilibrium behavior: a reputation motive and a defense motive. When the prior probability that a leader is crazy decreases, the reputation motive promotes less aggressive behavior by that leader, while the defense motive pushes for more aggressive behavior. These two forces underlay several comparative statics results. For example, the study analyzes the effect of increasing the prior probability that a leader is crazy, and the effect of changing the relative military strengths of the countries, on the equilibrium behavior of both leaders. The analysis also characterizes the conditions under which the madman strategy is profitable (or not), which contributes to the debate in the literature about its effectiveness.

"Come egli è cosa sapientissima simulare in tempo la pazzia."
["It is wise to sometimes pretend to be crazy."]
–Niccolò Machiavelli in *Discourses on Livy*

In the opening paragraph of his classic article, Fearon (1995, 379) lists three explanations for the occurrence of wars:

> First, one can argue that people (and state leaders in particular) are sometimes or always irrational. They are subject to biases and pathologies that lead them to neglect the costs of war or to misunderstand how their actions will produce it. Second, one can argue that the leaders who order war enjoy its benefits but do not pay the costs, which are suffered by soldiers and citizens. Third, one can argue that even rational leaders who consider the risks and costs of war may end up fighting nonetheless.

Avidit Acharya is Assistant Professor of Political Science, Stanford University, Encina Hall West, Room 406, Stanford, CA 94305-6044 USA (email: avidit@stanford.edu). Edoardo Grillo is Unicredit and Universities Foscolo Fellow, Collegio Carlo Alberto, Via Real Collegio, 30, 10024 Moncalieri (Torino), Italy (email: edoardo.grillo@carloalberto.org). We are grateful to Roland Bénabou, James Fearon, Mark Fey, Adam Meirowitz, John Londregan, Juan Ortner, Satoru Takahashi, the editor, two anonymous referees, and especially Stephen Morris and Kristopher Ramsay, for valuable comments and discussions. We also thank audiences at Vanderbilt, Princeton and the University of Rochester for their comments. We are responsible for any remaining errors.

Fearon proceeds to focus his attention on the third perspective, which he calls the rationalist explanation of war. Under this perspective, war is an outcome of strategic actions taken by two rational (and unitary) states that have fully considered its costs, benefits and uncertainty.

Indeed, with very few exceptions, the formal literature on war addresses only this rationalist explanation.[1] Moreover, despite the plausibility of Fearon's first two explanations, the bulk of this literature makes the idealized assumption that it is common knowledge that all leaders are rational and behave strategically. Yet both historical and contemporary evidence suggests that this assumption is more an idealization of reality rather than a reflection of it.[2] There is ample evidence that key decision makers have, at various times, tried to build reputations of being crazy or to take advantage of their adversaries' concerns that they may be irrational. Consider, for example, White House Chief of Staff Bob Haldeman's (1978, 122) recollection of president Richard Nixon explaining to him the 'madman theory' at the height of the Vietnam War:

> I call it the Madman Theory, Bob. I want the North Vietnamese to believe I've reached the point where I might do anything to stop the war. We'll just slip the word to them that, for God's sake, you know Nixon is obsessed about communism. We can't restrain him when he's angry—and he has his hand on the nuclear button' and Ho Chi Minh himself will be in Paris in two days begging for peace.

White House papers released in the early 2000s made it clear that the Nixon administration used this madman strategy. On October 10, 1969, Nixon ordered a secret operation called Giant Lance in which 18 B-52 bombers loaded with nuclear weapons flew toward the Soviet Union to get the Soviets to think that he was possibly mad, and that he might be willing to use nuclear weapons in Vietnam (see, for example, Sagan and Suri 2003).

There is also evidence spanning much of history that other leaders have used the madman strategy, or that they were actually crazy. For example, Kimball (2004) argues that the Hittite King Mursli used this strategy in antiquity when demanding the release of a hostage. Much more recently, the *New York Times* reported that in late 2005, US General John Abizaid expressed concern "that Iran's new President Ahmedinejad seemed unbalanced, crazy even."[3] Given the use of the madman strategy in history, and the doubts that contemporary political and military leaders express about the sanity of their adversaries, the idealized assumption of common knowledge of rationality appears at odds with reality.

This article develops a model of war that relaxes the assumption of common (in fact, mutual) knowledge of strategic rationality. Our model builds on the existing crisis bargaining framework, but assumes that there exist types of both countries that have behavioral commitments to particular actions, including how much they are willing to concede in bargaining. In our model, these types are "crazy" in a particular way: whenever they are confronted with a choice between two actions, they always choose the more aggressive one, and at the time of bargaining they only make or accept offers that would give them an unreasonably large payoff. Consequently, one can view our model as bringing together Fearon's (1995) three explanations for war. It introduces crazy types who are "subject to biases and pathologies that lead them to neglect the

[1] A notable exception is Jackson and Morelli (2007), which models the agency problem that arises in Fearon's (1995) second explanation. See Jackson and Morelli (2009), Reiter (2003) and Powell (2002) for surveys of the remaining literature.

[2] Fearon (1995, 409) himself concludes his article with the following disclaimer: "I am not saying that explanations for war based on irrationality or 'pathological' domestic politics are less empirically relevant. Doubtless they are important, but we cannot say how so or in what measure if we have not clearly specified the causal mechanisms making for war in the 'ideal' case of rational unitary states."

[3] http://www.nytimes.com/2010/11/29/world/middleeast/29iran.html.

costs of war" or "who enjoy its benefits but do not pay the costs," and it explores the effect of crazy types on the behavior of the strategically rational types "who [fully] consider the risks and costs… [but] may end up fighting nonetheless."[4] More importantly, by introducing crazy types into the standard crisis bargaining model, we are able to analyze the considerations of leaders like Nixon, Abizaid and Ahmedinejad, discussed above, and to assess the effectiveness of the so-called madman theory as a strategy in crisis bargaining.

Our model has three distinctive features. First, despite having two-sided incomplete information, it is tractable and makes (generically) unique equilibrium predictions. In equilibrium, the strategic types of both countries mimic the behavior of the crazy types with positive probability and, as a result, conflict takes place with positive probability (which, of course, is inefficient).

Second, we exploit the uniqueness of our equilibrium predictions to provide new comparative static results relating the probabilities of crazy types, and the countries' military strengths, to the probability of war. At the heart of these comparative statics are two distinct equilibrium forces that arise due to mutual doubts of strategic rationality: a *reputation motive* and a *defense motive*. The reputation motive describes a leader's incentive to pretend to be crazy in order to get a better settlement, while the defense motive describes a leader's incentive to make larger demands that risk escalation in order to deter his adversary from pretending to be crazy too often. Our results provide a precise characterization of the combined effect of the reputation and defense motives. This enables us to derive the comparative statics of equilibrium behavior and payoffs with respect to the model's parameters.

Third, and finally, our article contributes to a long-standing debate about whether the madman strategy 'works,' by characterizing the costs and benefits of using this strategy, and the conditions under which it is optimal. We show that these conditions depend largely on prior beliefs—in particular, on the effect that these beliefs have on the balance between the reputation motive for one country and the defense motive for the other. When the prior probability of the crazy type is sufficiently small for each country, both pretend to be crazy with some probability. However, because each country is mixing between escalation and concession, the madman strategy ends up being payoff-neutral. (Pretending to be crazy with a higher probability would result in a lower payoff; committing to pretend with a lower probability would raise payoffs, but would not be credible). This happens because the defense motive pushes a country to adopt more aggressive behavior in order to limit the benefits that its opponent can obtain by also playing aggressively. As the prior probability of the crazy type increases for a particular country, that country puts less probability on the concessional action, and eventually has a strict incentive to pretend to be crazy. Then, the strategy of always pretending to be crazy becomes profitable as the reputation motive is no longer in play (that is, the country no longer seeks to preserve the reputational content of aggressive behavior). Yet when the prior probability with which the other country is crazy exceeds a certain threshold, the madman strategy may not be adopted at all, as the defense motive can no longer compensate for the increase in the expected aggressiveness of the opponent. Therefore, whether or not the madman strategy works depends on prior beliefs about *both* countries, and on how each country is expected to respond to aggressions. In particular, even if a country has an incentive to occasionally pretend to be crazy, this may be only payoff neutral.

In the remainder of the introduction, we illustrate the most salient features of our approach. Then we review the literature and present the model, along with its comparative statics. The final section concludes.

[4] The assumption that the crazy type always chooses the more aggressive action can be weakened to assume that this type chooses the more aggressive action with sufficiently high probability.

An Illustration of our Approach

Consider the game tree depicted in Figure 1. Countries A and B are engaged in a dispute. Country A moves first: the choice is between attacking Country B and resolving the dispute peacefully. If Country A attacks, then Country B chooses between surrender and retaliate. If Country B retaliates, then Country A can either end the war with an armistice or escalate the conflict by choosing total war. Since all actions are uniquely labeled, we can identify terminal nodes with the actions that lead to them. We assume that $1 < w < 3$, so that Country A's preference over outcomes is Surrender $\succ_A$ Peace $\succ_A$ Armistice $\succ_A$ Total War, while Country B's preference is Peace $\succ_B$ Armistice $\succ_B$ Surrender $\succ_B$ Total War. Also note that the payoffs in Figure 1 are consistent with the idea that war is costly: with each aggressive action—attack and retaliate—one unit of total payoff is lost, and with total war an extra three units are lost. By backward induction, one can show that under complete information the unique subgame-perfect equilibrium outcome of the model is peace.

However, suppose that Country B believes that Country A is a strategic type that plays according to seqential rationality only with probability $p \in (0, 1)$; for our purposes, we assume that p is close to 1. With complementary probability $1 - p$, Country B believes that Country A is a crazy type that always chooses attack, and always chooses total war in the event that Country B chooses to retaliate. For simplicity, assume that Country A is certain (that is, believes with probability 1) that Country B is a strategic type that plays according to sequential rationality. Finally, suppose that these prior beliefs are common knowledge. It is easy to show that this game has a unique sequential equilibrium in which the strategic type of Country A attacks Country B with positive probability, and Country B retaliates with positive probability. The following is a sketch of the argument.

Sequential rationality requires the strategic type of Country A to always choose armistice over total war when confronted with this decision. Let a denote the equilibrium probability with which Country B believes that Country A is the strategic type, conditional on Country A choosing attack. If Country A attacks, then Country B's expected payoff from retaliating is aw, while its payoff from surrendering is 1. Consequently, the equilibrium probability with which Country B retaliates is 1 if $a > 1/w$ and 0 if $a < 1/w$. Country B mixes between surrender and retaliate only if $a = 1/w$.

Now, it is clear that there is no equilibrium in which the strategic type of Country A chooses peace with probability 1. If there were such an equilibrium, then conditional on Country A choosing to attack, Country B would believe with certainty that Country A is crazy; that is, $a = 0$. But since the crazy type always chooses total war, Country B would surrender for sure. Therefore, the strategic type of Country A would want to deviate to attack. Similarly, there is no

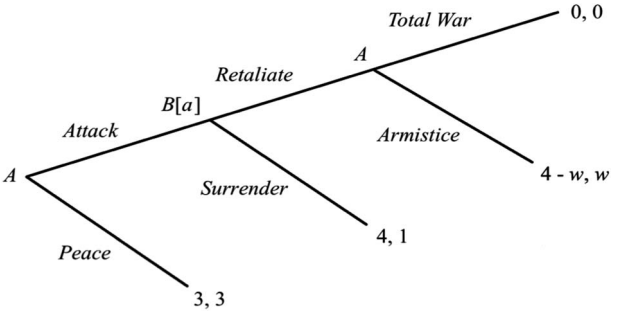

Fig. 1. Game tree for the illustrative example

equilibrium in which the strategic type of Country A definitely attacks. Otherwise, given that both types of Country A attack, Country B's posterior would be the same as its prior; that is, $a = p$. Now, we assumed that p is close to 1; in particular we assume $p > 1/w$. But in this case, Country B retaliates with probability 1, which in turn implies that the strategic type of Country A would want to deviate to peace.

Equilibrium must then involve the strategic type of Country A mixing between peace and attack, and thus it must be indifferent between these two actions. This indifference pins down the probability with which Country B retaliates. It is easy to show that this probability is simply $1/w$. Since Country B mixes between retaliate and surrender, it must be indifferent between these actions, so $a = 1/w$. This describes the equilibrium probability with which the strategic type of Country A chooses to attack: $(1-p)/[(w-1)p] > 0$. This concludes the characterization of the sequential equilibrium.

Now, define the equilibrium probability of war as the probability that the equilibrium outcome will be armistice or total war:

$$\underbrace{\left[(1-p) + \frac{1-p}{p} \cdot \frac{1}{w-1}\right]}_{\Pr[\text{A attacks}]} \times \underbrace{\left[\frac{1}{w}\right]}_{\Pr[\text{B retaliates}]} \tag{1}$$

Therefore, the probability with which Country A attacks is increasing in its prior probability of being the crazy type. Also, the probability with which Country A attacks and the probability with which Country B retaliates are both decreasing in w. Since w measures the split of total payoff after an armistice, we conclude that the probability of war is decreasing in the relative strength of Country B.

Equation 1 implies that the equilibrium probability of war goes to 0 as p goes to 1. Nevertheless, even a relatively small probability that Country A is the crazy type can lead to war with significantly higher probability. For example, suppose that $w = 1.1$ so that Country A is relatively stronger than Country B, and there is a 1 percent prior chance that Country A is the crazy type; that is, $p = 0.99$. In this case, the equilibrium probability of war is slightly over 10 percent. This is because the probability with which the *strategic* type of Country A attacks is slightly above 10 percent, while Country B retaliates with probability slightly larger than 90 percent. Therefore, even a small chance that Country A is crazy may have an amplified effect on its equilibrium behavior. Note also that in this example, slightly over 90 percent of wars are fought between strategic types.

The example above highlights some of the salient features of our approach. However, several questions remain: What happens when *both* countries have positive prior probability of being crazy? What happens when these probabilities are not as small as we have assumed above? What happens when one country is able to make offers to the other country to avoid war or cease hostilities? In this case, what can one assume about the behavior of the crazy type at the negotiating table? Which incentives determine the probability with which strategic types participate in costly conflicts? Our model addresses all of these questions. It handles two-sided incomplete information with ease, its equilibrium predictions are almost always unique, it generalizes the comparative statics results above, and it identifies the incentives that lie behind them.

RELATED LITERATURE

Our model builds upon the crisis bargaining literature, which goes back to the work of Powell (1987), Banks (1990) and Fearon (1995). The chief insight in this literature is that war cannot be

an equilibrium outcome when both parties are able to locate a Pareto superior negotiated settlement. By contrast, our model shows that because of the presence of aggressive crazy types, bargaining cannot resolve conflict—even between the strategic types—when they have incentives to pretend to be crazy types that are unreasonably demanding. In this way, our model is most closely related to crisis bargaining models with private information in which the parties involved have incentives to misrepresent their information, for example Fey and Ramsay (2010), Leventoglu and Tarar (2008), Slantchev (2005) and Schultz (1999). However, the key difference is that these previous articles incorporate incomplete information by assuming that types can be either 'tough' or 'lenient' rather than 'crazy' or 'strategic.' In particular, even the tough types are strategic, and will accept offers that are larger than any payoff they can hope to achieve by rejecting. Besides being qualitatively different, our model facilitates a more tractable analysis and, in contrast to previous articles, yields unique equilibrium predictions, which enables us to deliver a number of comparative statics results relating the probability of war and the payoffs received by the strategic types with the prior probability of the crazy types.

Substantively, our article contributes to the debate about the effectiveness of the madman theory, as articulated by Richard Nixon. Kaplan (1991) explains how scholars in the 1950s and 1960s debated the merits of using irrationality and unpredictability in nuclear policy. While some scholars, including Schelling (1963), did not consider the madman strategy to be effective in the age of nuclear weapons, others, notably Kissinger (1969), argued in favor of the madman strategy (claiming that the Soviets were also using it) as well as the merits of limited war (which in our model can be thought of as a war ending with the action we label armistice). Indeed, Kissinger became an ardent supporter of the madman strategy while working for Nixon, and oversaw much of its use in Vietman. For example, Sherry (1995) argues that the Nixon administration's decision to indiscriminately bomb Cambodia was a manifestation of the madman theory, while Sagan and Suri (2003) recount the use of the madman strategy in Operation Giant Lance, mentioned above. Whether or not the madman strategy works has been debated, however. Jeffrey Kimball writes, for example, that "with or without nuclear threats, the madman theory has worked for some decision makers, leaders, statesmen, tyrants, aggressors, and conquerors during the long course of history, but it has not always worked, and it did not work for Nixon and Kissinger during the Vietman War" (Kimball 2005).

Although the debate over the effectiveness of the madman strategy is yet unresolved, it is possible, and perhaps likely, that the strategy has been used even in contemporary politics. For example, Harvey Simon (2013) suggests that Kim Jong Un might be using the madman strategy, and Lake (2011) writes that the George W. Bush administration thought of Saddam Hussein as "uniquely evil," suggesting that they might have viewed him as a potential madman. Lake (2011) also advocates a theory of war that accounts for the possibility of cognitive biases and irrationality. As a step in this direction, our model introduces the possibility of a crazy type in the standard crisis bargaining model, and contributes to the debate about the effectiveness of the madman strategy. In particular, we characterize the conditions under which the madman strategy yields a higher expected payoff than any other strategy (making it an equilibrium strategy) and the conditions under which it does not. In doing so, we exploit the uniqueness of equilibrium predictions to derive comparative static results with respect to the aggressiveness of the countries and their relative military strength.

Our article is also closely related to an incisive article by Patty and Weber (2006), who argue that war cannot arise under the assumption of common knowledge of strategic rationality, but they do not model what happens when this assumption is relaxed. We, on the other hand, explicitly relax the assumption of common knowledge of rationality, and are thus able to

characterize the effect of crazy (or, in the language of Patty and Weber, 'irrational') types on the equilibrium behavior of strategic types.

Ours is not the first article on international security to study the effect of incorporating behavioral types into an otherwise rational framework. In an early article, Alt, Calvert and Humes (1988) studied deterrence by a hegemonic power against a series of short-run challengers in which the hegemon could possibly be a dominant strategy type as in Kreps and Wilson (1982). These articles, and others in the reputation literature (for example, Fudenberg and Levine 1989), show that the opportunity to build a reputation is payoff improving to the player who takes the opportunity to build it. In contrast to these articles, in our model strategic agents pool with the behavioral type, even though this is (in expectation) payoff-neutral for them. In this way, our article is most closely related to the work of Abreu and Gul (2000) on reputational bargaining that shows that a slight possibility of irrationality on either side has a pooling effect that produces inefficient delays in bargaining.[5]

Two compelling foundations for the crazy type that appears in our model emerge from the work of Weisiger (2013) in international relations theory, and Bénabou and Tirole (2009) in behavioral economics. Weisiger (2013) argues that uncertainty about leaders' actual intentions could lead to costly conflict, and generate the belief that some leaders may possess a dispositional inability to commit to peace. Bénabou and Tirole (2009) study belief distortions created by pride, dignity and wishful thinking about future outcomes, especially as they relate to intransigence in bargaining.

MODEL

Consider the game tree depicted in Figure 2. Countries A and B are engaged in a dispute. Country A begins by deciding between peace and attack. In the case of peace, the countries receive payoffs (z_A, z_B). If Country A attacks, then Country B makes an offer $x_A \in X \equiv [0, 1]$, where x_A is the payoff it is offering to Country A and $x_B \equiv 1-x_A$ is the payoff that it is proposing for itself. Country A can either accept the offer or escalate the conflict by rejecting it. If it rejects, then Country B either signs an armistice that leads to payoffs (y_A, y_B), or it chooses total war, which results in payoffs (0, 0). We assume that war is costly for both sides.

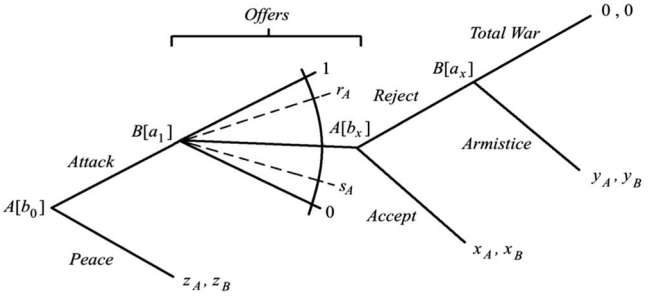

Fig. 2. The game tree

[5] Recent work by Ely and Välimäki (2003) and Canes-Wrone, Herron and Shotts (2001) shows that reputation effects may even be payoff-decreasing when the incentive for strategic types to separate from 'bad types' results in them receiving lower payoffs than they would achieve in the absence of bad types.

ASSUMPTION 1. War is costly:

$$(a)\, z_A > y_A > 0,\ (b)\, z_B > y_B > 0, \text{ and } (c)\, z_A + z_B > 1 > y_A + y_B.$$

Under Assumption 1, it is easy to see that with complete information, the game has a unique subgame perfect equilibrium, the only outcome of which is peace. However, instead of assuming complete information, suppose that at the beginning of the game, Country B believes that Country A is strategically rational only with probability $a_0 \in (0,1)$; with complementary probability $1-a_0$, Country B believes that Country A is a crazy type that always attacks, and accepts an offer x_A if and only if $x_A \geq r_A$ for some $r_A < 1$. Similarly, assume that at the beginning of the game Country A believes that Country B is strategically rational only with probability $b_0 \in (0,1)$; with complementary probability $1-b_0$, Country A believes that Country B is a crazy type that always makes the offer $s_A \in X$ for some $s_A > 0$, and always chooses total war.[6] We call this game $G(a_0, b_0)$, and we make the following assumptions on r_A and s_A.

ASSUMPTION 2. Crazy types are greedy:

$$(a)\ 1-y_B > r_A > z_A \text{ and } (b)\ \min\{1-z_B, y_A\} > s_A > 0.$$

Assumption 2a states that the crazy type of Country A seeks a payoff greater than the peaceful payoff z_A. It also states that this type is not too demanding: the agreement that it seeks is better for Country B than the outcome under an armistice. This means that Country B has the opportunity to reach a settlement that is better for it than a war that ends in armistice, even when Country A is crazy.[7] Assumption 2b states that the crazy type of Country B makes an offer that is worse for Country A than the outcome under armistice, and worse than giving Country B the payoff it would have received under peace. Combining Assumptions 1 and 2 yields:

$$1 > r_A > y_A > s_A > 0. \tag{2}$$

It is useful to point out that our assumptions give strategic types the opportunity to prevent either the start or the escalation of a costly war at each node (Country A can choose peace or accept the offer of Country B, while Country B can make the concessional offer r_A or choose armistice). We make this modeling choice primarily because we are interested in understanding how the existence of crazy types may lead to costly conflicts between strategic types, even though there are agreements that would be Pareto superior for the strategic types. In other words, we are analyzing a model in which the assumptions are stacked against producing war through the introduction of crazy types.[8] Nevertheless, we will show that costly wars may take place.

A behavioral strategy profile for the game $G(a_0, b_0)$ is denoted $\langle(\alpha, \alpha_x),\ (\beta, \beta_{TW})\rangle$, where α is the probability with which the strategic type of Country A chooses to attack; $\alpha_x : X \to [0,1]$ is a mapping where $\alpha_x(x_A)$ denotes the probability with which the strategic type of Country A rejects the offer x_A; $\beta \in \Delta(X)$ is a probability measure of the set of feasible offers made by the

[6] Our assumption about the behavior of irrational types in the bargaining phase of the game adapts Myerson's (1991) notion of an *r-insistent* type (see also Abreu and Gul 2000).

[7] If, instead, we were to assume that $1-y_B < r_A$, then Country B would never offer r_A. As a result, the strategic type of Country A would never mimic the behavior of the crazy type, as its final payoff of choosing attack (s_A or y_A) would never exceed the payoff from choosing peace at the initial node (z_A). If we also relax Assumption 1a by assuming $y_A > z_A$, then the strategic type of Country A could still behave aggressively in an attempt to increase its payoff by the amount $y_A - z_A$. In this case, the equilibrium would be similar to the one we characterized above, with y_A replacing w_A (though, here we would have to also account for the incentives of Country B to mimic its crazy type). As a result, the main insights of the article would still hold.

[8] By changing the *ex ante* likelihood of crazy types ($1-a_0$ and $1-b_0$) and the size of their demands (captured by r_A and s_A), we can make the damage imposed by crazy types quite substantial.

strategic type of Country B; and $\beta_{TW} : X \to [0,1]$ is a mapping where $\beta_{TW}(x_A)$ is the probability with which the strategic type of Country B chooses total war following the rejection of offer x_A by Country A. If the strategic types of the two countries play the behavioral strategy profile $\langle(\alpha, \alpha_x), (\beta, \beta_{TW})\rangle$, then Country B's updated belief that Country A is strategic, conditional on an attack, is:

$$a_1 \equiv \frac{\alpha a_0}{1-a_0+\alpha a_0}. \tag{3}$$

Country A's updated belief that Country B is strategic, conditional on receiving an offer (x_A, x_B), is given by the function $b_x : X \to [0,1]$ such that:

$$b_x(x_A) = \begin{cases} 1 & \text{if } x_A \neq s_A \\ \frac{\beta(s_A)b_0}{1-b_0+\beta(s_A)b_0} & \text{if } x_A = s_A. \end{cases} \tag{4}$$

We denote Country B's belief that Country A is strategic at the node at which it chooses between total war and armistice by a_x. Although we can characterize a_x using Bayes rule, its value will not matter at any information set. This is because Assumption 1b implies that in any perfect equilibrium of the game, the strategic type of Country B will choose armistice.

DEFINITION 1. A *sequential equilibrium* (or simply equilibrium) of game $G(a_0, b_0)$ is

(i) a behavioral strategy profile $\langle(\alpha, \alpha_x), (\beta, \beta_{TW})\rangle$, and
(ii) an associated Bayesian belief system $(a_0, a_1, a_x, b_0, b_x)$
such that (α, α_x) and (β, β_{TW}) are sequentially rational given $(a_0, a_1, a_x, b_0, b_x)$.

Before stating our first proposition, we define the following thresholds:

$$\underline{a} = \frac{1-r_A-y_B}{1-s_A-y_B} \qquad \bar{a} = \frac{1-r_A-y_B}{1-y_A-y_B}$$

$$\underline{b} = \frac{z_A-s_A}{r_A-s_A} \qquad \overline{b} = \frac{s_A}{y_A} + \frac{(y_A-s_A)(z_A-s_A)}{y_A(r_A-s_A)}.$$

It is easy to verify from Assumptions 1 and 2 that $1 > \bar{a} > \underline{a} > 0$ and $1 > \overline{b} > \underline{b} > 0$. These thresholds are depicted in Figure 3, which divides the parameter space $\mathcal{P} = (0,1)^2$ into five regions, labeled (*i*) through (*v*). We now characterize the equilibria of the game $G(a_0, b_0)$ in these five regions, except on the boundaries.[9]

PROPOSITION 1. The equilibria of the game $G(a_0, b_0)$ in Regions (*i*) to (*v*) are characterized as follows. In every equilibrium of the game we have $\beta_{TW}(x_A) = 0$ for all $x_A \in X$. Furthermore:

(i) If $b_0 < \underline{b}$, then in every equilibrium, we have $\alpha = 0$,

$$\alpha_x(x_A) \in \begin{cases} \{0\} & \text{if } x_A = s_A \text{ or } x_A > y_A \\ [0,1] & \text{if } x_A = y_A \\ \{1\} & \text{if } x_A < y_A \text{ and } x_A \neq s_A, \end{cases}$$

$\beta(r_A) = 1$ and $\beta(x_A) = 0$ for all $x_A \neq r_A$.

[9] The knife-edge cases for which equilibrium is not characterized are cases of indifference. We ignore these cases in the interest of substantive emphasis, as none of our conclusions depends on what happens in these cases.

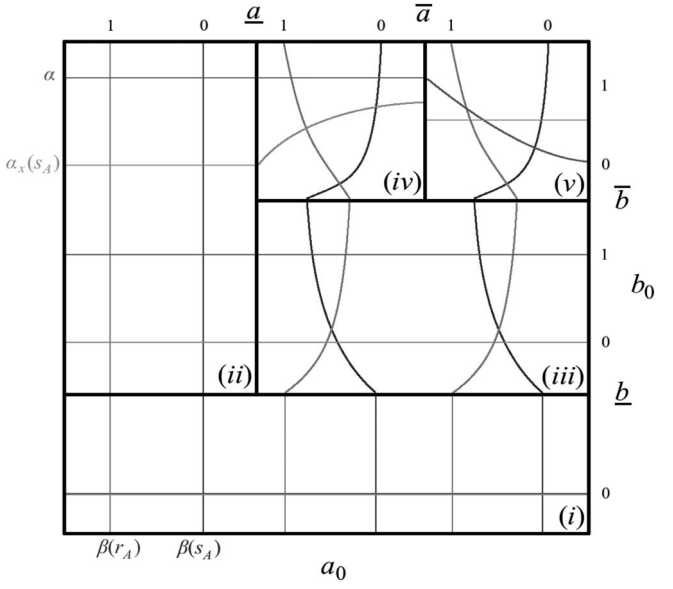

Fig. 3. Equilibrium behavior in the five regions defined in Proposition 1
Note: $\alpha_x(s_A)$ in orange, β (r_A) in blue and β (s_A) in brown.

(ii) If $b_0 > \underline{b}$ and $a_0 < \underline{a}$, then in every equilibrium, we have $\alpha = 1$, α_x is given by (*) above, $\beta(r_A) = 1$ and $\beta(x_A) = 0$ for all $x_A \neq r_A$.

(iii) If $\overline{b} > b_0 > \underline{b}$ and $a_0 > \underline{a}$, then in every equilibrium, we have $\alpha = \frac{1-a_0}{a_0} \cdot \frac{1-r_A-y_B}{r_A-s_A}$, α_x is given by (*) above, $\beta(s_A) = 1 - \frac{z_A-s_A}{b_0(r_A-s_A)}$, $\beta(r_A) = \frac{z_A-s_A}{b_0(r_A-s_A)}$ and $\beta(x_A) = 0$ for all $x_A \neq s_A, r_A$.

(iv) If $b_0 > \overline{b}$ and $\overline{a} > a_0 > \underline{a}$, then in every equilibrium we have $\alpha = 1$,

$$\alpha_x(x_A) \in \begin{cases} \{1\} & \text{if } x_A < y_A \text{ and } x_A \neq s_A \\ \{1 - \frac{1-r_A-y_B}{a_0(1-s_A-y_B)}\} & \text{if } x_A = s_A \\ [0,1] & \text{if } x_A = y_A \\ \{0\} & \text{if } x_A > y_A \end{cases}$$

$\beta(s_A) = \frac{1-b_0}{b_0} \frac{s_A}{y_A-s_A}$, $\beta(r_A) = 1 - \frac{1-b_0}{b_0} \frac{s_A}{y_A-s_A}$ and $\beta(x_A) = 0$ for all $r_A \neq s_A, r_A$.

(v) If $a_0 > \overline{a}$ and $b_0 > \overline{b}$, then in the unique equilibrium $\alpha = \frac{1-a_0}{a_0} \frac{1-r_A-y_B}{r_A-y_A}$,

$$\alpha_x(x_A) = \begin{cases} 1 & \text{if } x_A < y_A \text{ and } x_A \neq s_A \\ \frac{y_A-s_A}{1-s_A-y_B} & \text{if } x_A = s_A \\ 0 & \text{if } x_A \geq y_A, \end{cases}$$

$\beta(s_A) = \frac{1-b_0}{b_0} \frac{s_A}{y_A-s_A}$, $\beta(y_A) = 1 - \frac{z_A-b_0y_A}{b_0(r_A-y_A)} - \frac{(1-b_0)s_A}{b_0(y_A-s_A)}$, $\beta(r_A) = \frac{z_A-b_0y_A}{b_0(r_A-y_A)}$ and $\beta(x_A) = 0$ for all $x_A \neq s_A, y_A, r_A$.

PROOF. See Appendix A.

We summarize the main features of equilibrium as follows. If the prior probability that Country B is strategic is very low, $b_0 < \underline{b}$, then the strategic type of Country A will not attack to avoid

having to bargain with what is likely to be the crazy type of Country B (Case i). However, the strategic type of Country A may attack when $b_0 > \underline{b}$. If Country A is likely to be crazy, $a_0 < \underline{a}$, while Country B is believed to be strategic with probability $b_0 > \underline{b}$, then the strategic type of Country A will attack for sure, and the strategic type of Country B will try to settle the dispute early by making the concessional offer r_A. This offer will be accepted by both the strategic and crazy types of Country A (Case ii). If Country A is likely to be strategic, $a_0 > \underline{a}$, and Country B is moderately likely to be strategic, $\overline{b} > b_0 > \underline{b}$, then the strategic type of Country A attacks with probability $\alpha \in (0,1)$. If it attacks, then the strategic type of Country B mixes between the concessional offer r_A, which would be accepted for sure, and the greedy offer s_A, which the crazy type of Country B would make. Therefore, the strategic type of Country B sometimes pretends to be crazy (Case iii). This is also what happens when Country B is very likely to be strategic, $b_0 > \overline{b}$, and Country A is moderately likely to be strategic, $\bar{a} > a_0 > \underline{a}$, except that in this case the strategic type of Country A attacks for sure (Case iv). Finally, when both countries are very likely to be strategic, $a_0 > \bar{a}$ and $b_0 > \overline{b}$, then Country A mixes between attacking and taking the peaceful outcome; following an attack, Country B mixes between the concessional, intermediate and greedy offers, r_A, y_A and s_A (Case v).

An important feature of Proposition 1 is that its predictions are unique. As mentioned above, our model differs from other crisis bargaining models with incomplete information in that it does not lead to multiplicity of equilibrium predictions. The reason for this is as follows. The behavior of the strategic type of each country is determined by two countervailing forces: (1) the incentive to build a reputation for being crazy by mimicking the crazy type and (2) the incentive to deter the opponent from pretending to be crazy too often. For a wide range of values of prior beliefs, a_0 and b_0, equilibrium requires that these two forces exactly offset each other. This can happen only if the strategic type is indifferent between its actions. In turn, this indifference requires that the strategic type assigns a particular probability to its opponent being the crazy type. This probability pins down equilibrium behavior and delivers uniqueness. Put differently, uniqueness stems from the fact that, in equilibrium, the strategic type has to mimic the crazy type with a particular probability, which results in an equilibrium level of reputation that leaves the strategic type of the opponent indifferent between its equilibrium actions.

Furthermore, although Country B has infinitely many possible offers, in equilibrium it only makes one of three offers: a low (greedy) offer s_A that corresponds to what the crazy type of Country B would ask for, an intermediate offer y_A that is equal to what Country B would offer in the complete information game, and a high (concessional) offer r_A that even the crazy type of Country A would accept. Other offers are neither helpful in building a reputation for being crazy, nor optimal against one of the two types of Country A. They reveal that Country B is the strategic type, but conditional on Country A knowing this, Country B could do better.

Also notice that in Region (*v*) the strategic type of Country B makes offers y_A and r_A with positive probability, even though both of these offers are accepted with probability 1 by the strategic type of Country A *and* $r_A > y_A$. To understand this feature of equilibrium, notice that these offers yield different payoffs if Country B is facing the crazy type of Country A. Indeed, the crazy type accepts r_A and rejects y_A with certainty; consequently, Country B's payoff from making these offers to a crazy type of Country A would be $1 - r_A$ and y_B, respectively, with $1 - r_A > y_B$ by Assumption 2a. Proposition 1 implies that the equilibrium-updated belief a_1 will be such that the strategic type of Country B is indifferent between offering r_A (and receiving $1 - r_A$ with certainty) and offering y_A (and receiving $1 - y_A$ with probability a_1 and y_B with probability $1 - a_1$).

Finally, Proposition 1 contains the result that uncertainty concerning a country's type leads to conflict with positive probability, and that conflict is hard to settle. To see this, note that a_0, the

probability with which the strategic type of Country A decides to attack, is positive in all regions except (*i*), and the equilibrium behavior further implies that armistice arises with positive probability in Regions (*iv*) and (*v*). Thus not only does the strategic type of Country A initiate conflicts, but the strategic type of Country B may also fail to make offers that would settle them without creating further inefficiency.

We end this section by noting that, as in the illustrative example above, the majority of wars may be fought between strategic types. To see this, define the equilibrium probability of war as the probability with which the game play reaches the node in which Country B must choose between armistice or total war. Then note that the fraction of wars fought between two strategic types is simply:

$$\%\text{of Wars between Strategic Types} = \frac{\Pr[\text{War between Strategic Types}]}{\Pr[\text{War}]}$$

$$= \frac{[a_0 \cdot \alpha \cdot \alpha_x(s_A)] \cdot [b_0 \cdot \beta(s_A)]}{[1-a_0 + a_0 \cdot \alpha \cdot \alpha_x(s_A)] \cdot [1-b_0 + b_0 \cdot \beta(s_A)]}, \tag{5}$$

where α, $\alpha_x(s_A)$ and $\beta(s_A)$ are the equilibrium values of these choice variables given in Proposition 1. This follows because the unconditional probability that the game play reaches the node in which Country B chooses between armistice and total war is the expression in the denominator, while the same probability (conditional on both types of countries being strategic) is the expression in the numerator. Then, assume that the payoff parameters of the model are given by:

$$z_A = z_B = 0.55, \;\; r_A = 0.555, \;\; s_A = 0.45, \;\; y_A = 0.545, \;\; y_B = 0.1,$$

so that the payoff that a strategic type of Country B can get from conflict (y_B) is lower than the payoff that Country A can get (y_A). Furthermore, let $a_0 = b_0 = 0.99$ so that countries are believed to be strategic with 99 percent certainty. In this case, wars fought among strategic types represent approximately 75 percent of all wars. Thus even though the fraction of crazy types in the population is relatively small, uncertainty may lead to excessive aggression by the strategic types, to the point where most wars are fought among strategic types.

COMPARATIVE STATICS

In this section we report some of the comparative statics of the model with respect to parameters that have particular interpretations. First we report the comparative statics of equilibrium behavior, and then we report the comparative statics of the *ex ante* expected payoffs of the strategic types.

Comparative Statics of Equilibrium Behavior

Below, we report the comparative statics of equilibrium behavior with respect to a_0 and b_0, which measure the prevalence of crazy types; with respect to r_A and $1 - s_A$, which we interpret as measuring the 'aggressiveness' of the crazy types during the bargaining phase of the game; and, finally, with respect to the payoff split from armistice (y_A, y_B), which we interpret as reflecting the relative military strength of the two countries during war.

Prevalence of Crazy Types. Proposition 2 below reports the comparative statics of equilibrium behavior (specifically, the choice variables α, $\alpha_x(x_A)$ and $\beta(x_A)$) with respect to a_0 and b_0.

(The comparative statics of $\beta_{TW}(x_A)$ are trivial, since it is always equal to 0). Its proof follows from the equilibrium characterization in Proposition 1, and its visual representation is provided in Figure 3. For ease of reading, we state the result referring to Regions (*i*) to (*v*) depicted in the figure, rather than repeating the thresholds that define these regions.

PROPOSITION 2. The comparative statics with respect to a_0 and b_0 are as follows:

(1) α is continuous and weakly decreasing in a_0 for all b_0. Furthermore, it is strictly decreasing in a_0 only in Regions (*iii*) and (*v*).
(2) $\beta(r_A)$ is constant (and equal to 1) in Regions (*i*) and (*ii*), strictly decreasing in b_0 in Regions (*iii*) and (*v*) and strictly increasing in b_0 in Region (*iv*).
(3) $\beta(s_A)$ is constant (and equal to 0) in Regions (*i*) and (*ii*), strictly increasing in b_0 in Region (*iii*) and strictly decreasing in b_0 in Regions (*iv*) and (*v*).
(4) $\beta(y_A)$ is constant (and equal to 0) in Regions (*i*) to (i*v*) and strictly increasing in Region (*v*).
(5) $\alpha_x(s_A)$ is constant (and equal to 0) in Regions (*i*) to (*iii*), strictly increasing in a_0 in Region (*iv*) and constant (strictly between 0 and 1) in Region (*v*).

PROOF. Follows immediately from Proposition 1.

According to Proposition 2, the probability with which the strategic type of Country A attacks is strictly decreasing in Regions (*iii*) and (*v*) and constant everywhere else. The intuition for this is similar to the one we provided in the example above. Consider, for instance, Region (*v*), and recall that Proposition 1 states that Country B mixes between three offers: the concessional offer r_A that the crazy type of Country A accepts, the intermediate offer y_A and the greedy offer s_A. This is possible only if Country B is indifferent between these offers. Now, suppose that a_0 increases while a remains constant. In this case, after an attack, Country B believes that Country A is irrational with lower probability. This, in turn, makes the concessional offer less attractive. So, to keep Country B indifferent between the three offers, the equilibrium value of a must adjust down in order to maintain a fixed posterior probability of Country A being the crazy type. We call this the *reputation motive*. Intuitively, the reputation motive leads the strategic type of Country A to compensate changes in the *exogenous probability* of attacking, a_0, with changes in the *endogenous probability* of attacking, α, in order to maintain a constant level of reputation, a_1. The same intuition holds in Region (*iii*). In the remaining regions, the incentives for the strategic Country A to mimic the crazy type are either totally absent (Region (*i*)) or overwhelmingly strong (Regions (*ii*) and (*iv*), leading to the result that α is constantly equal to 0 and 1 in these respective cases.

The relationship between $\beta(\cdot)$ and b_0 is more interesting. Here, there are two competing forces. On the one hand, as we increase b_0, the strategic type of Country B must decrease the probability with which it mimics the crazy type, which is a consequence of a reputation motive analogous to the one we described in the previous paragraph. On the other hand, Country B has to 'protect' itself from the possibility of aggressive behavior by Country A. This can be done in two ways: *exogenously*, by relying on its reputation for being a crazy type, $1-b_0$, or *endogenously*, by decreasing the probability $\beta(r_A)$ of the concessional offer. In order to prevent the strategic type of Country A from attacking too often, the strategic type of Country B has to substitute an increase in the exogenous parameter b_0 with a decrease in the endogenous probability $\beta(r_A)$. We call this the *defense motive*. Obviously, this motive is stronger if there is a high probability that Country A is playing strategically (a_0 high). As a result of these two

opposing forces, $\beta(r_A)$ is increasing in b_0 in Region (*iv*), where the reputation motive prevails, and decreasing in b_0 in Regions (*iii*) and (*v*), where the defense motive prevails. Furthermore, the probability mass lost by offer r_A in Country B's equilibrium strategy as a consequence of the defense motive is reallocated either to the greedy offer s_A or to the intermediate offer y_A that would arise in the complete information game. In particular, s_A receives more of this mass in Region (*iii*) when the prior probability of Country B being crazy is high, while y_A receives more of the mass in Region (*v*), where this probability is relatively low and it is harder for the strategic type of Country B to mimic the crazy type.

Finally, consider the comparative statics of the probability with which Country A rejects *B*'s offer, $\alpha_x(\cdot)$. Since the strategic type of Country A will always accept offers r_A and y_A, the only relevant comparative static here is the one of $\alpha_x(s_A)$ with respect to a_0. This probability is constantly equal to 0 in Regions (*i*) to (*iii*), increasing in a_0 in Region (*iv*), and equal to a positive constant in Region (*v*). This happens because at this late node in the game, the reputation motive disappears, but the defense motive is still in play: Country A has to substitute its exogenous reputation with endogenous choices to avoid being exploited. It does this by increasing the probability of rejecting the greedy offer s_A.

To sum up, on the one hand, the reputation motive pushes for a decrease in the aggressiveness of strategic types as a result of a decrease in their prior probability of being crazy; on the other hand, because of the defense motive, the strategic types of both countries will react to a decrease in the exogenous probability of being crazy with an increase in the endogenous probability of behaving aggressively.

Aggressiveness of crazy types. The defense motive is also at play when we study how the countries' probabilities of concession vary as we increase the aggressiveness of the crazy type of the opponent. We measure the aggressiveness of Country A's crazy type by r_A and the aggressiveness of Country B's crazy type by $1 - s_A$. (Here, we identify aggressiveness as the minimum share of surplus that the crazy type seeks during the bargaining part of the game.) As we increase r_A, the probability $\beta(r_A)$ with which the strategic type of Country B makes the concessional offer decreases. Similarly, as we decrease s_A, the probability $\alpha_x(s_A)$ with which the strategic type of Country A rejects the greedy offer s_A goes up. The intuition for these results is that in both of these cases, the defense motive pushes strategic types to react to an exogenous increase in the aggressiveness of the opponent by lowering the probability with which they submit to aggressive behavior. This result is summarized in the following proposition.

PROPOSITION 3. In equilibrium:

(1) The probability $\beta(r_A)$ with which Country B makes the concessional offer is weakly decreasing in r_A.
(2) The probability $\alpha_x(s_A)$ with which Country A rejects the greedy offer of Country B is weakly decreasing in s_A.

PROOF. See Appendix B.

Military strength. We now consider the comparative statics of equilibrium behavior with respect to changes in y_A while holding the sum $y_A + y_B$ constant. We are interested in these comparative statics because we interpret the payoff split (y_A, y_B) as a measure of the relative military strength of the two countries. The stronger a country is, the larger the share of payoffs it

can expect to receive following a war that ends in armistice. To state the result of this section, we make the following assumption:

ASSUMPTION 3. $h = (z_A, z_B, s_A, r_A, y_A, y_B)$ and $h' = (z_A, z_B, s_A, r_A, y'_A, y'_B)$ are payoff profiles, each satisfying Assumptions 1 and 2. Furthermore, $y_A + y_B = y'_A + y'_B = \bar{y}$, and $y'_A > y_A$.

Our objective is to study the effect of a change from payoff profile h to h' on the equilibrium behavior characterized in Proposition 1. Let α, α_x and β denote the equilibrium quantities evaluated at payoff profile h, and let α', α'_x and β' denote the same quantities evaluated at payoff profile h'. Similarly, $\bar{a}$, $\underline{a}$, $\bar{b}$ and $\underline{b}$ are the thresholds in Part 5 evaluated at h, while $\bar{a}'$, $\underline{a}'$, $\bar{b}'$ and $\underline{b}'$ are the thresholds evaluated at h'. It is straightforward to verify the following (in)equalities:

$$\underline{a}' > \underline{a}, \quad \bar{a}' > \bar{a}, \quad \underline{b} = \underline{b}', \quad \bar{b}' < \bar{b}.$$

Now, define

$$\tilde{a} = \left(\frac{1-s_A-y_B}{1-s_A-y'_B}\right)\bar{a}'.$$

One can verify that $\tilde{a} \in (\bar{a}, \bar{a}'')$. Given this, the following proposition characterizes the comparative statics of the equilibrium behavior with respect to the payoff split (y_A, y_A).[10]

PROPOSITION 4. The following are true:

(1) If $a_0 < \underline{a}$ or if $b_0 < \underline{b}$ or if $a_0 \in (\underline{a}, \bar{a})$ and $b_0 > \overline{b}$, then $\alpha' = \alpha$; otherwise, $\alpha' > \alpha$.
(2) If $a_0 < \underline{a}$ or if $b_0 < \underline{b}$ or if $a_0 > \underline{a}'$ and $b_0 \in (\underline{b}, \bar{b}')$, then $\beta'(s_A) = \beta(s_A)$; otherwise, $\beta'(s_A) < \beta(s_A)$.
(3) If $a_0 < \underline{a}$ or if $b_0 < \underline{b}$ or if $a_0 > \underline{a}$ and $b_0 \in (\underline{b}, \bar{b}')$, then $\beta'(r_A) = \beta(r_A)$; if $a_0 > \bar{a}'$ and $b_0 > \bar{b}'$ then $\beta'(r_A) < \beta(r_A)$; otherwise, $\beta'(r_A) > \beta(r_A)$.
(4) If $a_0 > \bar{a}'$ and $b_0 > \bar{b}'$ then $\beta'(y_A) > \beta(y_A)$; if $a_0 \in (\bar{a}, \bar{a}')$ and $b_0 > \overline{b}$ then $\beta'(y_A) < \beta(y_A)$; otherwise, $\beta'(y_A) = \beta(y_A)$.
(5) If $a_0 > \underline{a}'$ and $b_0 \in (\bar{b}', \bar{b})$ or if $a_0 \in (\tilde{a}, 1)$ and $b_0 > \bar{b}$ then $\alpha'_x(s_A) > \alpha_x(s_A)$; if $a_0 \in (\underline{a}, \tilde{a})$ and $b_0 > \bar{b}$ then $\alpha_{x'}(s_A) < \alpha_x(s_A)$; otherwise, $\alpha'_x(s_A) = \alpha_x(s_A)$.

PROOF. See Appendix C.

The proof of the proposition in Appendix C utilizes the fact that a discrete change in y_A keeping $\bar{y}$ fixed has two effects. First, it modifies the boundaries of four out of the five regions defined in Proposition 1. Second, it affects the equilibrium behavior of countries within each of the five new regions. The comparative statics of the equilibrium behavior must take into account the combination of these two effects, since equilibrium strategies may differ across the boundaries of the five regions.

The result of Proposition 4 can be understood as follows. Part 1 of the proposition states the intuitive result that an increase in the relative military strength of Country A makes it behave (weakly) more aggressively. In particular, the probability of an attack increases unless Country A was attacking either with probability 1 or with probability 0; in the latter case, Country B's reputation for being crazy is so high that it discourages Country A from initiating a conflict.

[10] As before, we ignore the knife-edge cases $a_0 = \underline{a}, \bar{a}, \underline{a}', \bar{a}'$ and $b_0 = \underline{b}, \bar{b}, \bar{b}'$.

Part 2 states that the probability of making the greedy offer weakly decreases with Country A's relative military strength, because an increase in y_A reduces Country B's equilibrium expected payoff from mimicking the crazy type. (Recall that an increase in y_A is compensated by a decrease in y_B.) Part 3 states that an increase in Country A's relative military strength has an ambiguous effect on the equilibrium probability with which Country B makes the concessional offer. Intuitively, an increase in y_A has two effects. First, for the same reason as before, it makes the concessional offer more appealing for Country B. Second, by increasing the expected payoff from attacking, it makes the strategic type of Country A more aggressive. Thus due to the defense motive described above, the strategic type of Country B must decrease the probability of concession. Depending on which of these forces prevails, the probability of making the concessional offer could either increase or decrease. Similar reasoning lies behind the intuition of Part 4, which states that an increase in the relative military strength of Country A has an ambiguous effect on the probability with which Country B makes the intermediate offer y_A. (Note that $\beta(y_A) = 1-\beta(s_A)-\beta(r_A)$). Finally, Part 5 states that the effect of an increase in the military strength of Country A has an ambiguous effect on whether Country A accepts or rejects Country B's greedy offer. Once more, the reason for this is that the defense motive for Country A serves as a countervailing force vis-à-vis the increase in expected payoff associated with the rejection of the greedy offer.

Therefore, due to the defensive motive, an increase in the military strength of Country A may lead Country B to be less accommodating and to test Country A's true type by making the intermediate offer y_A. Notice that this result holds when there is little uncertainty that Country A is strategically rational; that is, when $a > \bar{a}'$. This is exactly the parameter range in which the strategic type of Country B is more concerned about defending itself against attacks from the strategic type of Country A.

Comparative Statics of Equilibrium Payoffs

In this section, we study the comparative statics of *ex ante* expected payoffs with respect to the prevalence and aggressiveness of crazy types in our model. Proposition 1 immediately implies that Country A's *ex ante* expected payoff is equal to

$$V_A = \begin{cases} z_A & \text{in regions (i), (iii), (v)} \\ b_0 r_A + (1-b_0)s_A & \text{in region (ii)} \\ b_0\beta(r_A)r_A + [1-b_0\beta(r_A)]s_A & \text{in region (iv)} \end{cases},$$

where the equilibrium value of $\beta(r_A)$ in Region (*iv*) is given in Proposition 1. Furthermore, since Proposition 1 says that Country B always puts positive probability on offer r_A, Country B's expected payoff at the node at which it makes an offer is always $1-r_A$. This implies that Country B's *ex ante* expected payoff is always given by

$$\begin{aligned} V_B &= a_0[(1-\alpha)z_B + \alpha(1-r_A)] + (1-a_0)(1-r_A) \\ &= z_B - (1-a_0+a_0\alpha)[z_B-(1-r_A)] \end{aligned},$$

where α takes its equilibrium value in Proposition 1 depending on which of the five regions the parameters (a_0, b_0) fall in. Thus Country B's expected payoff is always strictly less than z_B since $z_B > 1-r_A$, which follows from Assumptions 1c and 2a.

Prevalence of crazy types. It is straightforward to verify that V_A is larger than z_A in Regions (*ii*) and (*iv*). Therefore, recalling that $r_A > z_A > s_A$, we know that Country A's payoff is always weakly decreasing in the prior probability that Country B is the crazy type, $1 - b_0$, and strictly decreasing in Regions (*ii*) and (*iv*). The reason for this is that a higher probability of facing the crazy type reduces Country A's incentive to play aggressively. Yet Country A's payoff is piecewise constant in $1 - a_0$, with upward jumps when a_0 crosses $\bar{a}$ and $\underline{a}$ in the region where $b_0 > \underline{b}$. Indeed, if its own prior reputation for being crazy exceeds a critical value, Country A will be able to ignore the reputation motive and adjust its equilibrium behavior in order to obtain a higher payoff.

For similar reasons, V_B is always decreasing in the prior probability, $1 - a_0$, with which its opponent is crazy, and strictly decreasing in Regions (*i*), (*iii*) and (*v*). Furthermore, it is piecewise constant in 1–b_0 with discrete upward jumps when b_0 crosses the boundaries of the equilibrium regions. We summarize the above discussion as follows.

PROPOSITION 5.

(1) Country A's *ex ante* expected payoff V_A is:
   (a) weakly increasing in b_0 and strictly increasing in Regions (*ii*) and (*iv*).
   (b) piecewise constant in a_0 with downward jumps at the boundaries of Regions (*ii*), (*iii*), (*iv*) and (*v*).
(2) Country B's *ex ante* expected payoff V_B is:
   (a) weakly increasing in a_0 and strictly increasing in Regions (i), (iii) and (v).
   (b) piecewise constant in b_0 with downward jumps at the region boundaries.

PROOF. Follows from the characterization of Proposition 1.

Aggressiveness of crazy types. Unlike the comparative statics of payoffs with respect to the prevalence of crazy types, the comparative statics of payoffs with respect to the aggressiveness of crazy types may be ambiguous. The reason behind this ambiguity hinges on the joint effect of the reputation and defense motives.

First, consider Country A. Changes in r_A and $1 - s_A$ affect equilibrium behavior both because these changes may affect behavior within the various regions and because they may affect the boundaries of these regions. As a result, V_A will sometimes be increasing, sometimes be decreasing and sometimes remain constant in r_A and s_A. Indeed, as r_A increases, the strategic type of Country A gains more from mimicking the crazy type. Thus, keeping equilibrium behavior constant, a rise in r_A would increase Country A's payoff. However, due to the reputation motive for Country A, the strategic type of Country A will have to lower its probability of attacking in order to preserve its reputation. Furthermore, due to the defense motive of Country B, the probability with which the strategic type of Country B makes the concessional offer has to decrease. Since these two forces result in a decrease in V_A, the *ex ante* payoff of Country A may end up being constant or decreasing in r_A. Yet as s_A decreases (so that $1-s_A$ increases), mimicking the crazy type becomes less profitable for the strategic type of Country A and more profitable for the strategic type of Country B. As a result, the reputation motives of both countries may result in an increase in α and a decrease in β (s_A); when these effects are particularly strong (as is the case in Region (*iv*)), V_A may increase the aggressiveness of Country B's crazy type.

Now, consider Country B. The *ex ante* payoff of the strategic type of Country B varies ambiguously with r_A, because a rise in r_A has two opposing effects. On the one hand, it

decreases the payoff that Country B can get by making the concessional offer and, consequently, lowers its expected payoff. On the other hand, because of Country A's reputation motive, it lowers the probability, α, of an attack, which has the effect of increasing Country B's payoff. Depending on which effect dominates, V_B could either decrease or increase in r_A. Finally, a change in the aggressiveness of Country B's crazy type, $1-s_A$, can also have ambiguous effects on V_B through its effect on α. Once more, this is the result of two countervailing forces. First, a decrease in s_A directly lowers α, as it decreases the payoff that the strategic type of Country A can get from mimicking the crazy type. Second, for similar reasons as before, it relaxes the constraints of the reputation motive, and pushes for an increase in α.

The characterization of these comparative statics results follows almost immediately from Proposition 1, but the formal statement of our result, which we present next, requires us to take into account the effects of changes in r_A or s_A in several regions of the parameter space. To that end, let $h = (z_A, z_B, s_A, r_A, y_A, y_B)$, $h^* = (z_A, z_B, s_A, r'_A, y_A, y_B)$ and $h^\dagger = (z_A, z_B, s'_A, r_A, y_A, y_B)$ be three parameter profiles, each satisfying Assumptions 1 and 2, and assume that $r'_A > r_A$ and $s_A > s'_A$. Let $\underline{a}$, $\bar{a}$, $\underline{b}$ and $\overline{b}$ define the boundaries of the regions at profile h; $\underline{a}^*$, $\bar{a}^*$, $\underline{b}^*$ and $\bar{b}^*$ define the boundaries at profile h^*; and $\underline{a}^\dagger$, $\overline{a}^\dagger$, $\underline{b}^\dagger$ and $\overline{b}^\dagger$ define the boundaries at profile $h^\dagger$. Similarly, let V_A and V_B be the *ex ante* expected payoffs under h; V_A^* and V_B^* be the same payoffs under h^*; and $V_A^\dagger$ and $V_B^\dagger$ be the same payoffs under $h^\dagger$. We then have the following result:

PROPOSITION 6. *Country A's payoffs satisfy*

$$V_A^* \begin{cases} > V_A & \text{if } a_0 \le \underline{a}^* \text{ and } b_0 \ge \underline{b}^*; \text{ or if } a_0 \in [\underline{a}, \bar{a}^*] \text{ and } b_0 \ge \overline{b}^* \\ < V_A & \text{if } a_0 \in [\underline{a}^*, \underline{a}] \text{ and } b_0 \ge \underline{b}; \text{ or if } a_0 \in [\bar{a}^*, \bar{a}] \text{ and } b_0 \ge \bar{b} \\ = V_A & \text{elsewhere} \end{cases}$$

$$V_A^\dagger \begin{cases} < V_A & \text{if } a_0 \le \underline{a} \text{ and } b_0 \ge \underline{b}; \text{ or if } a_0 \in [\underline{a}, \bar{a}] \text{ and } b_0 \ge \bar{b} \\ > V_A & \text{if } a_0 \in [\underline{a}, \bar{a}] \text{ and } b_0 \in [\overline{b}^\dagger, \bar{b}] \\ = V_A & \text{elsewhere} \end{cases}.$$

For some threshold $\bar{z} > 0$ (derived in Appendix D), Country B's payoffs satisfy

$$V_B^* \begin{cases} > V_B & \text{if } a_0 \in [\underline{a}^*, \underline{a}] \text{ and } b_0 \in [\underline{b}, \overline{b}^*]; \text{ or if } a_0 \in [\bar{a}^*, \bar{a}] \text{ and } b_0 \ge \overline{b}, \\ & \text{or if } z_B > 1-y_A, a_0 \ge \bar{a} \text{ and } b_0 \ge \bar{b}; \text{ or if } z_B > \bar{z}, a_0 \ge \bar{a}^* \text{ and } b_0 \in [\bar{b}^*, \bar{b}] \\ < V_B & \text{if } a_0 \in [\underline{a}^*, \underline{a}] \text{ and } b_0 \in [\underline{b}, \bar{b}^*]; \text{ or } a_0 \in [\bar{a}^*, \bar{a}] \text{ and } b_0 \ge \bar{b}; \\ & \text{or if } z_B > 1-y_A, a_0 \ge \bar{a} \text{ and } b_0 \ge \bar{b}; \text{ or } \textit{if } z_B > \bar{z}, \ a_0 \ge \bar{a}^* \text{ and } b_0 \in [\bar{b}^*, \bar{b}] \\ = V_B & \text{elsewhere} \end{cases}$$

$$V_B^\dagger \begin{cases} < V_B & \text{if } a_0 > \underline{a} \text{ and } b_0 \in [\overline{b}^\dagger, \bar{b}] \\ > V_B & \text{if } b_0 \in [\underline{b}, \underline{b}^\dagger]; \text{ or if } a_0 > \underline{a}^\dagger \text{ and } b_0 \in [\underline{b}^\dagger, \overline{b}^\dagger] \\ = V_B & \text{elsewhere} \end{cases}.$$

PROOF. See Appendix D.

FINAL REMARKS

We constructed a model of international conflict in which war arises as a result of uncertainty about whether countries behave strategically. This uncertainty may lead even strategic countries to behave according to Machiavelli's dictum that "it is sometimes wise to pretend to be crazy." By characterizing the exact conditions under which this happens, the article contributes to a long-standing debate in the literature about the effectiveness of the so-called madman strategy, which has been used by leaders throughout history.

Unlike the previous literature, our model delivers unique equilibrium predictions, which enable us to derive a number of new, but natural, comparative static results. In particular, our model identifies two countervailing effects that play a role in determining how often countries pretend to be crazy: the reputation motive and the defense motive. Whereas the defense motive provides a country with incentives to behave aggressively (to shield itself from aggression by the opponent), the reputation motive limits the extent to which a country can pretend to be crazy. Depending on which of these two forces prevails, conflict may arise and persist, exacerbating the inefficiencies associated with war.

Furthermore, we show how the balance between the reputation and defense motives is essential in determining the profitability of the madman strategy. Indeed, a country's incentives to pretend to be crazy are limited both by its own reputation motive and by the defense motive of the other country. Thus changes in leaders' perceived aggressiveness (measured in our model either by the prevalence or intransigence of their crazy types) may have ambiguous and non-obvious effects on the probability of conflict.

One of the peculiar features of the model is the treatment of the two countries as being asymmetric, in the sense that Country A is a first mover and can unilaterally impose peace on Country B. This assumption is built into the game form. Given this rigidity, the most natural way to interpret the model is to assume that Country A, the first mover, is possibly dissatisfied with the status quo and may seek to change it with force. Country B, on the other hand, is not dissatisfied, but has to entertain the possibility that Country A may use force. One simple remedy for this extreme asymmetry is to assume that there are two countries, say P and Q, which are chosen with equal probability to play the game in the role of Country A with the other country playing in the role of Country B. This is the case of extreme symmetry. More generally, one could assume that P is chosen to play in the role of Country A with some probability $\lambda \in [0, 1]$ while Q is chosen to play in the role of Country A with complementary probability $1-\lambda$. In this case, all actions, payoffs and comparative statics (that is, derivatives of equilibrium actions and payoffs) for each country would be weighted by λ and $1-\lambda$.

Finally, by assuming that the crazy type of Country A always accepts offers that are at least as large as r_A, and by assuming that the bargaining stage of the game is a static interaction, we are ruling out the possibility that crazy types can keep coming back to demand more each time they are appeased. An interesting extension to the model would be to introduce the possibility that crazy types become more demanding whenever their current demands are met. To analyze this extension, we would need to first develop a dynamic model, and then model the crazy type in this way. We think this would be a fruitful avenue of future research. Nevertheless, the current article is a first step toward relaxing the strong common-knowledge-of-rationality assumptions that pervade crisis bargaining theory.

REFERENCES

Abreu, Dilip, and Faruk Gul. 2000. 'Bargaining and Reputation'. *Econometrica* 68(1):85–117.

Alt, James E., Randall L. Calvert, and Brian D. Humes. 1988. 'Reputation and Hegemonic Stability: A Game Theoretic Analysis'. *American Political Science Review* 82(2):445–66.

Banks, Jeffrey S. 1990. 'Equilibrium Behavior in Crisis Bargaining Games'. *American Journal of Political Science* 33(3):599–614.
Bénabou, Roland, and Jean Tirole. 2009. 'Over My Dead Body: Bargaining and the Price of Dignity'. *American Economic Review. Papers and Proceedings* 99(2):459–65.
Canes-Wrone, Brandice, Michael C. Herron, and Kenneth W. Shotts. 2001. 'Leadership and Pandering: A Theory of Executive Policymaking'. *American Journal of Political Science* 45(3):532–50.
Ely, Jeffrey, and Juuso Välimäki. 2003. 'Bad Reputation'. *Quarterly Journal of Economics* 118(3): 785–814.
Fearon, James D. 1995. 'Rationalist Explanations for War'. *International Organization* 49(3):379–414.
Fey, Mark, and Kristopher W. Ramsay. 2010. 'Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict'. *American Journal of Political Science* 55(1):149–69.
Fudenberg, Drew, and David K. Levine. 1989. 'Reputation and Equilibrium Selection in Games with Patient Players'. *Econometrica* 57(4):759–78.
Haldeman, Harry R. 1978. *The Ends of Power*. New York: Times Books.
Jackson, Matthew O., and Massimo Morelli. 2007. 'Political Bias and War'. *American Economic Review* 97(4):1353–73.
——. 2009. 'The Reasons for Wars – an Updated Survey'. In *The Handbook on the Political Economy of War*, edited by C. Coyne. Cheltenham, 34–57. UK: Edward Elgar Publishing.
Kaplan, Fred M. 1991. *The Wizards of Armageddon*. Palo Alto, CA: Stanford University Press.
Kimball, Jeffrey. 2004. *The Vietnam War Files: Uncovering the Secret History of Nixon-Era Strategy*. Lawrence: University Press of Kansas.
——. 2005. 'Did Thomas C. Schelling Invent the Madman Theory?' *History News Network*. Available at http://hnn.us/article/17183.
Kissinger, Henry A. 1969. *American Foreign Policy: Three Essays*. New York: W. W. Norton & Company.
Kreps, David S., and Robert Wilson. 1982. 'Reputation and Imperfect Information'. *Journal of Economic Theory* 27(2):253–79.
Lake, David A. 2011. 'Two Cheers for Bargaining Theory: Assessing Rationalist Explanations of the Iraq War'. *International Security* 35(3):7–52.
Leventoglu, Bahar, and Ahmer Tarar. 2008. 'Does Private Information Lead to Delay or War in Crisis Bargaining?' *International Studies Quarterly* 52(3):533–53.
Myerson, Roger B. 1991. *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard University Press.
Patty, John W., and Roberto A. Weber. 2006. 'Agreeing to Fight: An Explanation of the Democratic Peace'. *Politics, Philosophy and Economics* 5(3):305–20.
Powell, Robert. 1987. 'Crisis Bargaining, Escalation and MAD'. *American Political Science Review* 81(3):717–36.
——. 2002. 'Bargaining Theory and International Conflict'. *Annual Review of Political Science* 5:1–30.
Reiter, Dan. 2003. 'Exploring the Bargaining Model of War'. *Perspectives on Politics* 1:27–43.
Sagan, Scott D., and Jeremi Suri. 2003. 'The Madman Nuclear Alert: Secrecy, Signaling, and Safety in October 1969'. *International Security* 27(4):150–83.
Schelling, Thomas C. 1963. *The Strategy of Conflict*. New York: Oxford University Press.
Schultz, Kenneth A. 1999. 'Do Democratic Institutions Constrain of Inform? Contrasting Two Institutional Perspectives on Democracy and War'. *International Organization* 53(2):233–66.
Sherry, Michael S. 1995. *In the Shadow of War*. New Haven, CT: Yale University Press.
Simon, Harvey. 2013. 'Kim Jong-un and the 'Madman Theory' of Diplomacy.' *History News Service*. Available at http://historynewsservice.org/2013/04/kim-jong-un-and-the-madman-theory-of-diplomacy/.
Slantchev, Branislav L. 2005. 'Military Coercion in Interstate Crises'. *American Political Science Review* 99(4):533–47.
Weisiger, Alex. 2013. *Logics of War: Sources of Limited and of Unlimited Conflicts*. Ithaca, NY: Cornell University Press.

APPENDIX

A. Proof of Proposition 1

For any belief a_x, sequential rationality requires $\beta_{TW}(x_A) = 0$. The remaining assertions of the proposition are an immediate consequence of the following lemmata. The first three are preliminary results. The latter five each characterize the equilibria in one of the five regions of the parameter space.

LEMMA 1:

(*i*) In any equilibrium of the game, we have

$$\alpha_x(x_A) \in \begin{cases} \{0\} & \text{if } b_x < x_A/y_A \\ [0,1] & \text{if } b_x = x_A/y_A \\ \{1\} & \text{if } b_x > x_A/y_A. \end{cases} \tag{A1}$$

Consequently, the strategic Country A must accept any offer $x_A > y_A$.

(*ii*) If $\beta(s_A) = 0$ in equilibrium, then

$$\alpha_x(x_A) \in \begin{cases} \{0\} & \text{if } x_A = s_A \text{ or } x_A > y_A \\ [0,1] & \text{if } x_A = y_A \\ \{1\} & \text{if } x_A < y_A \text{ and } x_A \neq s_A; \end{cases} \tag{*}$$

(*iii*) If in equilibrium $\alpha = 0$, then $\beta(r_A) = 1$.

(*iv*) If in equilibrium $\beta(y_A) > 0$, then $\alpha_x(y_A) = 0$.

PROOF.

(*i*) At the node labeled $A[b_x]$, rejecting the offer (x_A) gives the strategic Country A a payoff of y_A with probability b_x and 0 with probability $1-b_x$. On the other hand, accepting produces a payoff of x_A for sure. Therefore, it accepts for sure if $b_x < x_A/y_A$, rejects for sure if $b_x > x_A/y_A$, and is indifferent between accepting and rejecting if $b_x = x_A/y_A$. Since it must always be that $b_x \in [0,1]$, Country A will accept any offer $x_A > y_A$.

(*ii*) Observe that if $\beta(s_A) = 0$ then from Equation 4 in the main text,

$$b_x(x_A) = \begin{cases} 1 & \text{if } x_A \neq s_A. \\ 0 & \text{if } x_A = s_A \end{cases} \tag{A2}$$

This, along with (A1), immediately gives (*).

(*iii*) If $\alpha = 0$, then $a_1 = 0$. Therefore, Country B believes that Country A will reject any offer $x_A < r_A$. So by making such an offer it expects to receive a payoff of y_B. On the other hand, the offer r_A would be accepted for sure and give Country B a payoff of $1-r_A > y_B$. Therefore, the strategic Country B will make offer the offer r_A with certainty.

(*iv*) By the result of (*iii*), if Country B makes the offer y_A with positive probability, then it must be that $\alpha > 0$, consequently $a_1 > 0$. Suppose, for the sake of contradiction, that the strategic Country A rejects this offer with probability $\delta > 0$. Then the expected payoff to Country B of making this offer is

$$a_1(\delta y_B + (1-\delta)(1-y_A)) + (1-a_1)y_B. \tag{A3}$$

But by making the offer $y_A + \varepsilon$, where $0 < \varepsilon < \delta(1-y_A-y_B)$, Country B would have an expected payoff of

$$a_1(1-y_A-\varepsilon) + (1-a_1)y_B,$$

since by Lemma 1(*i*) the offer $y_A + \varepsilon$ is accepted by the strategic Country A. One can then use the assumption that $\varepsilon < \delta(1-y_A-y_B)$ to verify that the payoff in Inequality A4 is greater than the payoff in Inequality A3.

LEMMA 2: In any equilibrium of the game, the support of β is a subset of the following set of three offers: $\{s_A, y_A, r_A\}$.

PROOF. Recall that Equation 19 in the main text states that $1 > r_A > y_A > s_A > 0$. So what we must show is that β does not put any probability mass on the intervals $[0, s_A)$, (s_A, y_A), (y_A, r_A) and $(r_A, 1]$. Next, observe that if $a_1 = 0$, then $\alpha = 0$ and by Lemma 1(*iii*), only r_A is in support of β. Therefore, we can assume throughout the remainder of this proof that $a_1 > 0$.

Suppose β puts positive mass on $(r_A, 1]$. By Lemma 1(*i*), any offer $x_A \in (r_A, 1]$ will be accepted by both the strategic type and the crazy type; so it will be accepted with certainty. But so will the offer r_A. Since $1-r_A > 1-x_A$ for all $x_A \in (r_A, 1]$, Country B has a profitable deviation to the offer r_A. Therefore, β must put zero probability mass on $(r_A, 1]$.

Suppose that β puts positive mass on (y_A, r_A). Then, there exists $x_A \in (y_A, r_A)$ in support of β. If such an offer x_A is made, then by Lemma 1(*i*) it is accepted with probability a_1 and rejected with probability $1-a_1$. But again by Lemma 1(*i*), the offer $\frac{x_A + y_A}{2}$ will also be accepted with probability a_1 and rejected with probability $1-a_1$. Moreover, deviating to this offer is profitable for Country B. Therefore it cannot be that x_A is in support of β. Consequently, β cannot put positive probability on the interval (y_A, r_A).

Now suppose that β puts positive mass on $[0, s_A) \cup (s_A, y_A)$. Then if Country B makes an offer $x_A \in [0, s_A) \cup (s_A, y_A)$, we have $b_x(x_A) = 1$ by Equation 4 in the main text. Therefore, Country B's expected payoff of making the offer x_A is equal to y_B. But if Country B deviates to the offer $y_A + \varepsilon$, where $0 < \varepsilon < 1-y_A-y_B$, the strategic Country A will accept, giving Country B an expected payoff of

$$a_1(1-y_A-\varepsilon) + (1-a_1)y_B > y_B, \tag{A4}$$

where the inequality holds by Assumption 1c and $\varepsilon < 1-y_A-y_B$. In other words, Country B has a profitable deviation to the offer $y_A + \varepsilon$. Consequently, β cannot put positive probability mass on $[0, s_A) \cup (s_A, y_A)$.

LEMMA 3:

(*i*) There is no equilibrium with $\beta(r_A) = 0$.
(*ii*) There is no equilibrium with $\beta(y_A) > 0$ and $\beta(s_A) = 0$.
(*iii*) If $a_0 < \underline{a}$, then in equilibrium we must have $\beta(r_A) = 1$. If $a_0 > \underline{a}$ and $b_0 > \underline{b}$, then in equilibrium we must have $\beta(r_A) < 1$.
(*iv*) If $b_0 < \underline{b}$, then in equilibrium we must have $\alpha = 0$. If $b_0 > \underline{b}$, then in equilibrium we must have $\alpha > 0$.
(*v*) If $b_0 > \overline{b}$ and $a_0 > \underline{a}$, then in equilibrium we must have $\alpha_x(s_A) > 0$.

PROOF.

(*i*) Suppose there is an equilibrium with $\beta(r_A) = 0$. Then by Lemma 2, β puts positive probability only on a subset of $\{s_A, y_A\}$. But because $0 < s_A < y_A < z_A$ by Assumptions 1a and 2b, Country A's expected payoff from attacking must be less than z_A, its payoff to peace. Thus $\alpha = 0$, which by Lemma 1(*iii*) implies $\beta(r_A) = 1$. Contradiction.

(*ii*) Suppose there is an equilibrium with $\beta(y_A) > 0$ and $\beta(s_A) = 0$. By Lemma 1(*iv*), we must have $\alpha_x(y_A) = 0$. Since Lemma 3(*i*) implies $\beta(r_A) > 0$, Country B's expected payoff from the offer y_A must equal its expected payoff from the offer r_A:

$$1-r_A = a_1(1-y_A) + (1-a_1)y_B, \tag{A5}$$

which reduces to $a_1 = \bar{a}$. But note that by (*) we must have $\alpha_x(s_A) = 0$. Consequently, by deviating to the offer s_A, Country B can receive the expected payoff

$$a_1(1-s_A) + (1-a_1)y_B = \bar{a}(1-s_A) + (1-\bar{a})y_B > 1-r_A, \tag{A6}$$

where the inequality follows from substituting the expression for $\bar{a}$, simplifying and using Assumptions 2a and 2b. Thus the deviation is profitable to Country B. Contradiction.

(*iii*) Suppose $a_0 < \underline{a}$. Country B's maximum expected payoff from making the offer y_A or s_A is $a_1(1-s_A) + (1-a_1)y_B$. It is easily verified that this expected payoff is strictly less than $1-r_A$ when

$a_1 < \underline{a}$. Combining this with Lemma 2 and the fact that $a_1 \equiv \frac{\alpha a_0}{1-a_0+\alpha a_0} \leq a_0$ for all $\alpha \in [0,1]$ yields $\beta(r_A) = 1$.

On the other hand if $a_0 > \underline{a}$ and $\beta(r_A) = 1$, then $\alpha_x(s_A) = 0$ by (*). Consequently, by making the offer s_A, Country B has an expected payoff of $a_1(1-s_A)+(1-a_1)y_B$, which is strictly less than $1-r_A$ whenever $a_1 < \underline{a}$. The payoff to Country A of attacking is therefore $b_0 r_A + (1-b_0)s_A > z_A$, since $b_0 > \underline{b}$. Consequently, $\alpha = 1$ and $a_1 = a_0 < \bar{a}$, establishing a contradiction.

(*iv*) If Country A chooses to attack, then its maximum expected payoff is $(1-b_0)s_A + b_0 r_A$. One can easily verify that its expected payoff from peace, z_A, is strictly greater than this payoff whenever $b < \underline{b}$. Therefore, $\alpha = 0$.

On the other hand if $b_0 > \underline{b}$ and $\alpha = 0$, then by Lemma 1(*iii*) we need $\beta(r_A) = 1$. Therefore, by attacking, Country A can get an expected payoff of $b_0 r_A + (1-b_0)s_A$. Since $b_0 > \underline{b}$, this expected payoff is greater than z_A, establishing a contradiction.

(*v*) Suppose for the sake of contradiction that $\alpha_x(s_A) = 0$. By Lemma 1(*i*), this implies $b_x(s_A) \leq s_A/y_A$, or equivalently

$$\beta(s_A) \leq \frac{1-b_0}{b_0}\frac{s_A}{y_A - s_A}, \tag{A7}$$

which follows from noting that $b_x(s_A)$ is given by Equation 6 in the main text. Since $b_0 > \overline{b} > \underline{b}$, Lemma 3(*iv*) implies $\alpha > 0$; consequently $a_1 > 0$. So, if the strategic Country B makes the offer s_A, it gets $a_1(1-s_A)+(1-a_1)y_B > a_1(1-y_A)+(1-a_1)y_B$ by Assumption 2b. Therefore, $\beta(y_A) = 0$. By Lemma 3(*i*) and (*iii*), we know that $\beta(r_A), \beta(s_A) > 0$. This implies the indifference condition

$$1-r_A = a_1(1-s_A)+(1-a_1)y_B, \tag{A8}$$

or, equivalently $a_1 = \underline{a}$. Substituting a_1 from Equation 2 in the main text, this reduces to $\alpha = \frac{1-a_0}{a_0}\frac{\underline{a}}{1-\underline{a}} \in (0,1)$, where the strict inclusion holds because $a_0 > \underline{a}$ by assumption. But $\alpha \in (0,1)$ implies the indifference condition

$$b_0[(1-\beta(s_A))r_A + \beta(s_A)s_A] + (1-b_0)s_A = z_A$$
$$\Leftrightarrow \beta(s_A) = 1 - \frac{z_A - s_A}{b_0(r_A - s_A)}. \tag{A9}$$

Combining $\beta(s_A)$ in Inequality A9 with Inequality A7 implies $b_0 \leq \overline{b}$. This contradicts our assumption that $b_0 > \overline{b}$.

LEMMA 4: If $b_0 < \underline{b}$, then the equilibrium set is characterized by $\alpha = 0$, α_x given by (*) above, and $\beta(r_A) = 1$.

PROOF. Lemma 3(*iv*) implies $\alpha = 0$. Then Lemma 1(*iii*) implies $\beta(r_A) = 1$. Then Lemma 1(*ii*) implies that α_x is given by (*). Moreover, it is easy to verify that the given specifications for α, β and α_x are all sequentially rational given the starting beliefs, a_0 and b_0, and the updated beliefs, a_1 and b_x, that they imply.

LEMMA 5: If $a_0 < \underline{a}$ and $b_0 > \underline{b}$ then the equilibrium set is characterized by $\alpha = 1$, α_x given by (*) above, and $\beta(r_A) = 1$.

PROOF. Lemma 3(*iii*) implies that $\beta(r_A) = 1$. Then by Lemma 1(*ii*), α_x is given by (*). Finally, it is easy to verify that Country A's expected payoff from attack is $b_0 r_A + (1-b_0)s_A$, which is strictly greater than its payoff to peace, z_A, whenever $b_0 > \underline{b}$. Thus $\alpha = 1$. Moreover, it is easy to verify that the given specifications for α, β and α_x are all sequentially rational given the starting beliefs, a_0 and b_0, and the updated beliefs, a_1 and b_x, that they imply.

LEMMA 6: If $a_0 > \underline{a}$ *and* $\underline{b} < b_0 < \overline{b}$ then the following describes the set of equilibrium behavioral strategy profiles:

$\alpha = \frac{1-a_0}{a_0} \cdot \frac{1-r_A-y_B}{r_A-s_A}$, α_x is given by (*) above, $\beta(r_A) = \frac{z_A-s_A}{b_0(r_A-s_A)}$, $\beta(s_A) = 1 - \frac{z_A-s_A}{b_0(r_A-s_A)}$, and $\beta(y_A) = 0$.

PROOF.

Step (1): First we show that $\beta(y_A) = 0$. Suppose, for the sake of contradiction, that $\beta(y_A) > 0$. Lemma 1 (*iv*) implies that $\alpha_x(y_A) = 0$. Lemma 3(*ii*) implies $\beta(s_A) > 0$ as well. Therefore, we need the indifference condition

$$a_1(1-y_A) + (1-a_1)y_B = a_1[(1-\alpha_x(s_A))(1-s_A) + \alpha_x(s_A)y_B] + (1-a_1)y_B. \tag{A10}$$

Since Lemma 3(*iv*) implies $\alpha > 0$, which in turn implies $a_1 > 0$, we solve Inequality A10 for

$$\alpha_x(s_A) = \frac{y_A - s_A}{1-s_A-y_B} \in (0,1), \tag{A11}$$

where the strict inclusion follows from Assumptions 2a and 2b. This then implies $b_x(s_A) = s_A/y_A$ by Lemma 1(*i*). Using Equation 4 in the main text, we solve for

$$\beta(s_A) = \frac{1-b_0}{b_0} \cdot \frac{s_A}{y_A - s_A}. \tag{A12}$$

Since $\alpha > 0$ and Inequality A11 implies that Country A must have the same expected payoff from accepting and rejecting the offer s_A, we need

$$z_A \geq b_0[\beta(r_A)r_A + \beta(s_A)s_A + \beta(y_A)y_A] + (1-b_0)s_A, \tag{A13}$$

in which we can substitute (A12) and $\beta(r_A) = 1-\beta(y_A)-\beta(s_A)$, and solve to get

$$\beta(y_A) \leq \frac{b_0 y_A(r_A - s_A) - s_A(r_A - z_A) - y_A(z_A - s_A)}{b_0(y_A - s_A)(r_A - y_A)}. \tag{A14}$$

This expression on the right hand side of Inequality A14 is non-negative if and only if $b_0 \geq \overline{b}$. But this contradicts the premise of the Lemma.

Step (2): We now show that $\alpha_x(s_A) = 0$. Suppose $\alpha_x(s_A) > 0$. Then we need $b_x(s_A) \geq s_A/y_A$ by Lemma 1(*i*). If $\alpha_x(s_A) = 1$ then the expected payoff to Country B from making the offer s_A would be y_B. But by Lemma 1(*i*), Country B's expected payoff of offering r_A is $1-r_A > y_B$ by Assumption 2a. Therefore, we need $\beta(r_A) = 1$, which contradicts $b_x(s_A) \geq s_A/y_A$.

Now, suppose that $\alpha_x(s_A) \in (0,1)$. This implies $b_x(s_A) = s_A/y_A$, so that $\beta(s_A)$ is given by Inequality A12. Since we showed in Step (1) that $\beta(y_A) = 0$, and we know from Lemma 3(*iv*) that $\alpha > 0$, we need

$$z_A \leq b_0[(1-\beta(s_A))r_A + \beta(s_A)s_A] + (1-b_0)s_A. \tag{A14}$$

But this inequality reduces to $b_0 \geq \overline{b}$, which is again a contradiction.

Step (3): We now calculate the equilibrium values of α, β and α_x. We have shown that $\alpha_x(s_A) = 0$, and $\beta(y_A) = 0$, which implies that α_x is given by (*) above. Since $a_0 > \underline{a}$, Lemma 3(*iii*) implies $\beta(r_A) < 1$. Therefore, by Lemmas 2 and 3(*i*), we need $\beta(r_A), \beta(s_A) > 0$. These results imply the indifference condition Inequality A8, which in turn implies $a_1 = \underline{a}$. This implies that $\alpha = \frac{1-a_0}{a_0}\frac{\underline{a}}{1-\underline{a}} \in (0,1)$, where the strict inclusion follows from the fact that $a_0 > \underline{a}$ by assumption. Next, $\alpha \in (0,1)$ requires the indifference condition in Inequality A9 to be satisfied. The expressions for $\beta(s_A)$ and $\beta(r_A)$ follow. Moreover, it is easy to verify that any behavioral strategy profile satisfying the specifications in the statement of Lemma 6 constitutes an equilibrium, given the assumptions on a_0 and b_0, and the updated beliefs a_1 and b_x implied by the behavioral strategy profile.

LEMMA 7: If $b_0 > \overline{b}$ and $\overline{a} > a_0 > \underline{a}$, then in every equilibrium we have

$$\alpha = 1,\ \beta(r_A) = 1 - \tfrac{1-b_0}{b_0}\tfrac{s_A}{y_A - s_A},\ \beta(s_A) = \tfrac{1-b_0}{b_0}\tfrac{s_A}{y_A - s_A},\ \beta(y_A) = 0$$

$$\text{and } \alpha_x(x_A) \in \begin{cases} \{1\} & \text{if } x_A < y_A \text{ and } x_A \neq s_A \\ \{1 - \frac{1-r_A-y_B}{a_0(1-s_A-y_B)}\} & \text{if } x_A = s_A \\ [0,1] & \text{if } x_A = y_A \\ \{0\} & \text{if } x_A < y_A \end{cases}.$$

PROOF.

Step (1): We first show that $\beta(s_A) = \frac{1-b_0}{b_0}\frac{s_A}{y_A - s_A}$. By Lemma 3($v$), we need $\alpha_x(s_A) > 0$. Observe that $\alpha_x(s_A) = 1$ implies $\beta(s_A) = 0$. (The argument is the same as in Step (2) of Lemma 6). Thus, according to Lemma 3(ii), $\beta(y_A) = 0$, and this contradicts Lemma 3(iii). Therefore, we conclude that $\alpha_x(s_A) \in (0,1)$. By Lemma 1(i), this implies that $b_x(s_A) = s_A/y_A$, from which $\beta(s_A) = \frac{1-b_0}{b_0}\frac{s_A}{y_A - s_A}$ follows.

Step (2): We now show that $\beta(y_A) = 0$ in equilibrium. Suppose that $\beta(y_A) > 0$. In Step (1) we showed that $\beta(s_A) > 0$. Therefore, the indifference condition in Inequality A5 must be satisfied; thus $a_1 = \bar{a}$ and $\alpha = \frac{1-a_0}{a_0}\frac{\bar{a}}{1-\bar{a}}$. Since $a_0 < \bar{a}$, by assumption, we have $\alpha > 1$, which is absurd.

Step (3): We now establish the values of the other choice variables in equilibrium. By Step (2), we have $\beta(y_A) = 0$. By Lemma 3(v) and the argument in Step (1), we have $\alpha_x(s_A) \in (0,1)$ and $\beta(s_A) = \frac{1-b_0}{b_0}\frac{s_A}{y_A - s_A}$. Therefore, Country A is indifferent between accepting the offer s_A and rejecting it. These observations imply that the expected payoff to Country A from attacking is

$$b_0[(1-\beta(s_A))r_A + \beta(s_A)s_A] + (1-b_0)s_A, \tag{A15}$$

which is greater than z_A whenever $b_0 > \overline{b}$. Therefore, $\alpha = 1$. Thus $a_1 = a_0$. Then Country B must be indifferent between the offers r_A and s_A; that is

$$1 - r_A = a_0[(1-\alpha_x(s_A))(1-s_A) + \alpha_x(s_A)y_B] + (1-a_0)y_B. \tag{A16}$$

This implies that $\alpha_x(s_A)$ takes the value stated in the Lemma. Moreover, it is easy to verify that any behavioral strategy profile satisfying the specifications in the statement of Lemma 6 constitutes an equilibrium, given the assumptions on a_0 and b_0, and the updated beliefs a_1 and b_x implied by the behavioral strategy profile.

LEMMA 8: *If* $a_0 > \bar{a}$ and $b_0 > \overline{b}$, then the unique equilibrium is characterized by $\alpha = \frac{1-a_0}{a_0}\frac{1-r_A-y_B}{r_A-y_A}$, $\beta(s_A) = \frac{1-b_0}{b_0}\frac{s_A}{y_A-s_A}$, $\beta(y_A) = 1 - \frac{z_A - b_0 y_A}{b_0(r_A - y_A)} - \frac{(1-b_0)s_A}{b_0(y_A - s_A)}$, $\beta(r_A) = \frac{z_A - b_0 y_A}{b_0(r_A - y_A)}$, and

$$\alpha_x(x_A) = \begin{cases} 1 & \text{if } x_A < y_A \text{ and } x_A \neq s_A \\ \frac{y_A - s_A}{1 - s_A - y_B} & \text{if } x_A = s_A \\ 0 & \text{if } x_A \geq y_A \end{cases}$$

PROOF.

Step (1): We begin by showing that $\beta(y_A) > 0$. Suppose for the sake of contradiction that $\beta(y_A) = 0$. Lemma 3(iii) implies $\beta(s_A), \beta(r_A) > 0$. Then, the exact argument as in Step (1) of Lemma 7 establishes that $\beta(s_A) = \frac{1-b_0}{b_0}\frac{s_A}{y_A - s_A}$. The exact argument as in Step (3) of Lemma 7 establishes that $\alpha = 1$, hence $a_1 = a_0$. Now, the assumption that $a_0 > \bar{a}$ can be re-written as $a_0(1-y_A-y_B) > 1-r_A-y_B$. Observe from this that we can find $\varepsilon > 0$ small enough so that

$$\begin{aligned} & a_0(1-y_A-y_B-\varepsilon) > 1-r_A-y_B \\ \Leftrightarrow\ & a_0(1-y_A-\varepsilon) + (1-a_0)y_B > 1-r_A. \end{aligned} \tag{A17}$$

Since we stated above that $a_1 = a_0$, and we know that Lemma 1(i) states that the strategic Country A must accept any offer greater than y_A, the term on the left-hand side of Inequality A17 is Country B's expected payoff from making the offer $y_A + \varepsilon$, while the term on the right-hand side is its expected payoff from the offer r_A. Since we need $\beta(r_A) > 0$ by Lemma 3(i), we have a contradiction. Therefore $\beta(y_A) > 0$.

Step (2): We now establish the value of the choice variables in equilibrium. Step (1) shows that $\beta(y_A) > 0$, and by Lemma 3(i) and (ii), we need $\beta(r_A), \beta(s_A) > 0$ as well. These imply a number of indifference conditions as follows. With the help of Lemma 1(iv) and Lemma 3(i), we need the indifference condition Inequality A5 to be met. This implies $a_1 = \bar{a}$, thus $\alpha = \frac{1-a_0}{a_0}\frac{1-r_A-y_B}{r_A-y_A} \in (0,1)$, where the strict inclusion follows from $a_0 > \bar{a} > \underline{a}$. We also need the indifference condition Inequality A10, which implies $\alpha_x(s_A) = \frac{y_A - s_A}{1-s_A-y_B} \in (0,1)$ as in Inequality A11. Obviously, the stated expression for $\alpha_x(x_A)$ when $x_A \neq s_A$ follows from Lemma 1(i) and (iv). This in turn implies $\beta(s_A) = \frac{1-b_0}{b_0}\frac{s_A}{y_A - s_A}$ as in Inequality A12. Finally, because we showed that $\alpha \in (0,1)$ and Country A is indifferent between accepting and rejecting the offer s_A (recall that $\alpha_x(s_A) \in (0,1)$), we also need the indifference condition

$$z_A = b_0[\beta(r_A)r_A + \beta(y_A)y_A + \beta(s_A)s_A] + (1-b_0)s_A, \tag{A18}$$

which we can solve using $\beta(s_A) = \frac{1-b_0}{b_0}\frac{s_A}{y_A - s_A}$ and $\beta(r_A) + \beta(y_A) + \beta(s_A) = 1$ to get

$$\beta(r_A) = \frac{z_A - b_0 y_A}{b_0(r_A - y_A)} \qquad \beta(y_A) = 1 - \frac{z_A - b_0 y_A}{b_0(r_A - y_A)} - \frac{(1-b_0)s_A}{b_0(y_A - s_A)}. \tag{A19}$$

Furthermore, it is easy to verify that because $b_0 > \overline{b}$, the expressions for $\beta(y_A)$ and $\beta(r_A)$ given by Inequality A19 are strictly positive. Moreover, it is easy to verify that these choice variables are sequentially rational given the assumptions on a_0, b_0, and the implied updated beliefs a_1 and b_x.

B. Proof of Proposition 3

Let $h = (z_A, z_B, s_A, r_A, y_A, y_B)$, $h^* = (z_A, z_B, s_A, r'_A, y_A, y_B)$ and $h^\dagger = (z_A, z_B, s'_A, r_A, y_A, y_B)$ be three parameter profiles, each satisfying Assumptions 1 and 2, and suppose that $r'_A > r_A$ and $s'_A > s_A$. Notice that a change from h to h^* or from h to $h^\dagger$ will affect both the equilibrium behavior within the five regions described in Proposition 1 as well as the boundaries of these regions. Let $\underline{a}, \bar{a}, \underline{b}$ and $\overline{b}$ denote the boundaries of the regions under parameter profile h; $\underline{a}^*, \bar{a}^*, \underline{b}^*$ and $\overline{b}^*$ the boundaries under parameter profile h^*; and $\underline{a}^\dagger, \bar{a}^\dagger, \underline{b}^\dagger$ and $\overline{b}^\dagger$ the boundaries under parameter profile $h^\dagger$.

First consider Statement (1). Notice that $\underline{a}, \bar{a}, \underline{b}$ and $\overline{b}$ are all decreasing in r_A, holding other parameters constant. From the characterization of Proposition 1, it follows immediately that the result holds as long as the change r_A does not result in a change in the region in which a_0 and b_0 fall; and, in particular, the relationship is strict within regions (*iii*) and (*v*). Now, consider the case in which a change from h to h^* leads to a change in the region in which a_0 and b_0 fall. Several cases are possible. If $b_0 \in [\underline{b}^*, \underline{b}]$, then $\beta(r'_A) \leq 1$, while $\beta(r_A) = 1$. The same is true if $a_0 \in [\underline{a}^*, \underline{a}]$ and $b_0 > \underline{b}^*$ hold simultaneously. Now consider the case in which $a_0 \in [\underline{a}, \bar{a}^*]$ and $b_0 \in [\overline{b}^*, \overline{b}]$. This corresponds to the case in which $\beta(r_A) = \frac{z_A - s_A}{b_0(r_A - s_A)}$ and $\beta(r_A) = 1 - \frac{1-b_0}{b_0}\frac{s_A}{y_A - s_A}$; indeed, the characterization of equilibrium behavior is given by region (*iii*) under h and by region (*iv*) under h^*. In this case, $\beta(r_A) \geq \beta(r'_A)$ follows immediately from noticing that the highest value of b_0 is given by $\overline{b}$ and from rearranging terms. If, instead, $a_0 \geq \bar{a}^*$ and $b_0 \in [\overline{b}^*, \overline{b}]$, then $\beta(r_A) = \frac{z_A - s_A}{b_0(r_A - s_A)}$ and $\beta(r'_A) = \frac{z_A - b_0 y_A}{b_0(r_A - y_A)}$. From the fact that $b_0 \geq \overline{b}^*$, we can conclude that $\beta(r_A) \geq \beta(r'_A)$ if and only if $\frac{r'_A - s_A}{r_A - s_A} \geq \frac{r'_A - y_A}{r'_A - y_A}$, which holds since $r'_A > r_A$. Finally, suppose that $a_0 \in [\bar{a}^*, \bar{a}]$ and $b_0 \geq \overline{b}$. In this case, $\beta(r_A) = \frac{z_A - s_A}{b_0(r_A - s_A)}$ and $\beta(r'_A) = \frac{z_A - b_0 y_A}{b_0(r_A - y_A)}$. The result holds, as the lowest possible value that b_0 can take is $\overline{b}$.

Now consider Statement (2). Note that $\underline{a}$ and $\overline{b}$ are increasing in s_A, while $\underline{b}$ and $\bar{a}$ are respectively decreasing and constant in s_A. From Proposition 1, if an increase in s_A (that is, a decrease in $1 - s'_A$) does not change the relevant region, $\alpha_x(s_A)$ will weakly increase and will strictly increase in regions (*iv*) and (*v*). If the relevant region changes, the equilibrium value of $\alpha_x(s_A)$ will either stay constant at 0 or increase from 0 to some positive value (such an increase will happen when either $a_0 \geq \underline{a}^\dagger$ and $b_0 \in [\overline{b}, \overline{b}^\dagger]$, or when $b_0 \geq \overline{b}^\dagger$ and $a_0 \in [\underline{a}, \underline{a}^\dagger]$). This concludes the proof.

C. Proof of Proposition 4

Note that $y_B = \bar{y} - y_A$, so changes in y_A will affect y_B when keeping $\bar{y}$ fixed. So, in what follows, whenever we refer to an "increase in y_A" (for instance), we mean an "increase in y_A holding $\bar{y}$ fixed," which actually results in an equal decrease in y_B.

To prove the proposition, we must take into account the fact that a change in y_A may simultaneously change the thresholds $\underline{a}$, $\bar{a}$, $\overline{b}$ and $\underline{b}$ and the equilibrium behavior characterized by α, α_x and β. Note that, conditional on remaining in the interior of each of the five regions defined in Proposition 1 (and Figure 3), the comparative statics of the equilibrium behavior with respect to marginal changes in y_A are as follows:

(B1) In regions (*i*) and (*ii*), the equilibrium behavior is constant in y_A.

(B2) In region (*iii*), α is increasing in y_A, while $\beta(x_A)$ and $\alpha_x(x_A)$ are constant with respect to y_A for all $x_A \neq y_A$.

(B3) In region (*iv*), α is constant in y_A, $\beta(r_A)$ is increasing in y_A, $\beta(s_A)$ is decreasing in y_A, $\beta(x_A)$ is constant in y_A for all $x_A \neq r_A, s_A$, $\alpha_x(s_A)$ is increasing in y_A and $\alpha_x(x_A)$ is constant in y_A for all $x_A \neq y_A, s_A$.

(B4) In region (*v*), α and $\alpha_x(s_A)$ are increasing in y_A, $\alpha_x(x_A)$ is constant in y_A for all $x_A \neq s_A, y_A$, $\beta(s_A)$ and $\beta(r_A)$ are decreasing in y_A, $\beta(y_A)$ is increasing in y_A and $\beta(x_A)$ is constant in y_A for all $x_A \neq r_A, y_A, s_A$.

We prove each of the five results stated in the proposition separately. Some results are straightforward and follow immediately from the equilibrium characterization in Proposition 1 and the observations in (B1)–(B4). So, we focus on cases that are not trivially implied by these results.

(1) Given the change in the thresholds resulting from an increase in y_A, and since α is non-decreasing in y_A in each of the five regions, it is easy to verify that $\alpha' \geq \alpha$ with strict inequality when $\alpha \neq 0, 1$.

(2) The result is straightforward in all cases except the case in which $a_0 > \underline{a}'$ and $b_0 \in [\overline{b}', \overline{b}]$. In this case, we have $\beta(s_A) = 1 - \frac{z_A - s_A}{b_0(r_A - s_A)}$, while $\beta'(s_A) = \frac{1-b_0}{b_0}\frac{s_A}{y'_A - s_A}$. One can verify that in the case we are analyzing, $(\beta(s_A) - \beta'(s_A))$ is increasing in b_0, and converges to 0 as $b_0 \to \underline{b}'$. Therefore, it must be that $\beta(s_A) > \beta'(s_A)$.

(3) The result is straightforward in all cases except the case in which $a_0 \in (\underline{a}', \overline{a}')$ and $b_0 \in (\overline{b}', \overline{b})$, and the case in which $a_0 > \bar{a}'$ and $b_0 \in (\overline{b}', \overline{b})$. In the first case, $\beta(r_A) = \frac{z_A - s_A}{b_0(r_A - s_A)}$ and $\beta(r'_A) = \frac{b_0 y'_A - s_A}{b_0(y'_A - s_A)}$. The result follows from noticing that the difference $(\beta'(r_A) - \beta(r_A))$ increases in b_0 and it is equal to 0 when $b_0 = \overline{b}'$. In the second case, $\beta(r_A) = \frac{z_A - s_A}{b_0(r_A - s_A)}$ and $\beta'(r_A) = \frac{z_A - b_0 y'_A}{b_0(r_A - y'_A)}$. One can verify that in this case, $(\beta(r_A) - \beta'(r_A))$ is increasing in b_0, and converges to 0 as $b_0 \to \overline{b}'$. Therefore, it must be that $\beta(r_A) > \beta'(r_A)$.

(4) Since $\beta(y_A) = 0$ except when $a_0 > \bar{a}$ and $b_0 > \overline{b}$, the result is straightforward in all cases.

(5) The result is straightforward in all cases except the case in which $a_0 \in (\bar{a}, \bar{a}'')$ and $b_0 > \overline{b}$. In this case, $\alpha_x(s_A) = \frac{y_A - s_A}{1 - s_A - y_B}$ and $\alpha'_x(s_A) = 1 - \frac{1 - r_A - y'_B}{a_0(1 - s_A - y'_B)}$. One can verify that the quantity $\alpha_x(s_A) - \alpha'(s_A)$ is strictly decreasing in a_0 on the interval $(\underline{a}, 1)$ and is equal to 0 when $a_0 = \tilde{a}$. The result follows instantly.

D. Proof of Proposition 6

Recall from the proof of Proposition 3 that $\underline{a}, \bar{a}, \underline{b}, \bar{b}$ are all decreasing in r_A; $\underline{a}$ and $\overline{b}$ are increasing in s_A; and $\underline{b}$ and $\overline{b}$ are, respectively, decreasing and constant in s_A. Furthermore, our assumptions imply $r_A > s_A$, so that that V_A is larger in region (*ii*) than in region (*iv*), and larger in region (*iv*) than in regions (*i*), (*iii*) and (*v*).

Consider first the payoffs of Country A. V_A is constant in both r_A and s_A in regions (*i*), (*iii*) and (*v*), and increasing in both of these parameters in regions (*ii*) and (*iv*). Therefore the statement of the proposition follows immediately from these results and the changes in the boundaries of the five regions reported above.

Now consider Country B and recall that $V_B = z_B - (1 - a_0 + a_0\alpha)[z_B - (1 - r_A)]$. Therefore, V_B depends on r_A both directly and indirectly through the effect that r_A may have on α. As a result, we can immediately claim that $V_B^* < V_B$ whenever $\alpha^* \geq \alpha$. This happens if either $a_0 \leq \underline{a}$, or $b_0 \leq \underline{b}^*$, or $a_0 \in [\underline{a}^*, \bar{a}^*]$ and $b_0 \geq \overline{b}^*$. Instead, if $a_0 \in [\underline{a}^*, \underline{a}]$ and $b_0 \in [\underline{b}, \overline{b}^*]$, $V_B = 1 - r_A$ and $V_B^* = z_B - (1 - a_0)\left[1 + \frac{1 - r'_A - y_B}{r'_A - s_A}\right][z_B - (1 - r_A)]$. Thus using the lower bound for a_0, we can conclude that $V_B^* > 1 - r_A = V_B$. A similar reasoning leads to the same conclusion if $a_0 \in [\bar{a}^*, \bar{a}]$ and $b_0 \geq \overline{b}$. If instead, $a_0 \geq \underline{a}$ and $b_0 \in [\underline{b}, \overline{b}^*]$, one can check that V_B is decreasing in r_A, as $V_B = z_B - (1 - a_0)\left(\frac{1 - s_A - y_B}{r_A - s_A}\right)[z_B - (1 - r_A)]$, which is decreasing in r_A by Assumption 2b. If $a_0 \geq \bar{a}$ and $b_0 \geq \overline{b}$, $V_B = z_B - (1 - a_0)\left(\frac{1 - y_A - y_B}{r_A - y_A}\right)[z_B - (1 - r_A)]$, which is increasing or decreasing in r_A, depending on whether $1 - z_B < y_A$ or $1 - z_B \geq y_A$, respectively. Finally, if $a_0 \geq \bar{a}^*$ and $b_0 \in [\overline{b}^*, \overline{b}]$, we have that $V_B = z_B - (1 - a_0)\frac{1 - s_A - y_B}{r_A - s_A}[z_B - (1 - r_A)]$ and $V_B^* = z_B - (1 - a_0)\frac{1 - y_A - y_B}{r'_A - y_A}[z_B - (1 - r'_A)]$. Thus it is immediate to verify that V_B^* will be greater or lower than V_B depending on whether z_B is greater or lower than

$$\bar{z} := \frac{(1 - s_A - y_B)(1 - r_A)(r'_A - y_A) - (1 - y_A - y_B)(1 - r'_A)(r_A - s_A)}{(1 - s_A - y_B)(r'_A - y_A) - (1 - y_A - y_B)(r_A - s_A)}. \tag{A20}$$

Finally, consider a change in the aggressiveness of Country B as measured by $1 - s_A$. Notice that V_B will increase (respectively, decrease) as α decreases (respectively, increases) after we change from h to $h^\dagger$. The result follows immediately by characterizing the regions in which these changes take place.