

# Collegio Carlo Alberto



## War with Crazy Types

Avidit Acharya

Edoardo Grillo

**No. 356**

**June 2014**

# Carlo Alberto Notebooks

[www.carloalberto.org/research/working-papers](http://www.carloalberto.org/research/working-papers)

# War with Crazy Types\*

Avidit Acharya<sup>†</sup>      Edoardo Grillo<sup>‡</sup>

April 1, 2014

## Abstract

We model a situation in which two countries are involved in a dispute. The dispute can end in a peaceful settlement, or it can escalate to war. If it is common knowledge that the countries are strategically rational, then the only equilibrium outcome of the model is peace. If, on the other hand, each country believes that there is some chance that its adversary is a crazy type that always behaves aggressively, then even a strategically rational country may have an incentive to *pretend* to be crazy. This leads to war with positive probability.

In addition to being qualitatively different from the existing literature, our model (i) enables a more tractable analysis of two-sided incomplete information, (ii) has a generically unique equilibrium prediction, and (iii) yields several new comparative statics results. For example, we analyze the effect of increasing the prior probability that the countries are crazy types, as well as the effect of changing the relative military strengths of the countries, on equilibrium behavior. In studying these comparative statics, our model identifies two countervailing forces that arise when the prior probability that a country is crazy decreases: a *reputation motive* that promotes less aggressive behavior by that country, and a *defense motive* that promotes more aggressive behavior by the other country.

**JEL Classification Codes:** C7, F5, N4

**Key words:** war, conflict, bargaining, reputation

---

\*We are grateful to Roland Bénabou, James Fearon, Mark Fey, Adam Meirowitz, John Londregan, Juan Ortner and especially Stephen Morris and Kristopher Ramsay for many useful comments and discussions. We also thank participants at the Political Economy Graduate Student Seminar Series at Princeton University, and the Watson Center for Conflict and Cooperation Seminar Series at the University of Rochester for their comments. We are responsible for any remaining errors.

<sup>†</sup>Assistant Professor of Political Science and Economics at the University of Rochester, Harkness Hall 327, Rochester NY 14627-0158 (email: aachary3@z.rochester.edu)

<sup>‡</sup>Unicredit & Universities Foscolo Fellow at Collegio Carlo Alberto, Via Real Collegio, 30, 10024 Moncalieri (Torino), Italy (email: edoardo.grillo@carloalberto.org).

“Come egli è cosa sapientissima simulare in tempo la pazzia.”  
[“It is wise to sometimes pretend to be crazy.”]  
–Niccolò Machiavelli

## 1 Introduction

In the opening paragraph to his classic article, Fearon (1995) 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.” (pg. 379)

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, it is the rationalist explanation that the formal literature on war addresses.<sup>1</sup> 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 players are rational and behave strategically. Yet, it is not only the theoretical plausibility of Fearon’s first two explanations that raises questions about this assumption.<sup>2</sup> Recent Wikileaks documents suggest that key decision

---

<sup>1</sup>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.

<sup>2</sup>We should note that even rational choice theorists view the assumptions of common knowledge of rationality as an idealization of reality, rather than a reflection of it. For example, Fearon (1995) 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.” (pg. 409)

makers in the real world have serious doubts about the sanity of their adversaries. For example, the *New York Times* reported that in late 2005, Gen. John Abizaid of the United States Central Command expressed concern “that Iran’s new President Ahmedinejad seemed unbalanced, crazy even.”<sup>3</sup>

This paper develops a model of war that relaxes the assumption of common (in fact, mutual) knowledge of strategic rationality. We build on the existing crisis bargaining framework, assuming 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 action, 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 that are “subject to biases and pathologies that lead them to neglect the costs of war,” or “who enjoy its benefits but do not pay the costs” and, it explores the effect of these types on the behavior of the strategically rational types “who [fully] consider the risks and costs ... [but] may end up fighting nonetheless.”<sup>4</sup>

Despite two-sided incomplete information, our model yields a (generically) unique equilibrium prediction. In equilibrium, the strategic types of both countries mimic the behavior of the crazy types with positive probability; as a result, inefficient conflicts may arise (Proposition 1).

Although this (and a few other) features of our equilibrium have already been highlighted in the existing literature, we exploit the uniqueness of our equilibrium predictions to provide new comparative static results that were previously unattained. In particular, we identify two different forces that determine the equilibrium behavior of strategic types: a *reputation motive* and a *defense motive*. To get an intuition for these forces, consider what happens when the prior probability of a country being crazy decreases. Holding the behavior of this country fixed, this change in the prior probability of being crazy will have two effects. First, it gives the strategic type of that country a greater incentive to be less aggressive, so as to preserve the reputational content of aggressive behavior. (This is the reputation motive.) On the other hand,

---

<sup>3</sup><http://www.nytimes.com/2010/11/29/world/middleeast/29iran.html>.

<sup>4</sup>The assumption that the crazy type always chooses the more aggressive action can be weakened to assuming that this type chooses the more aggressive action with sufficiently high probability.

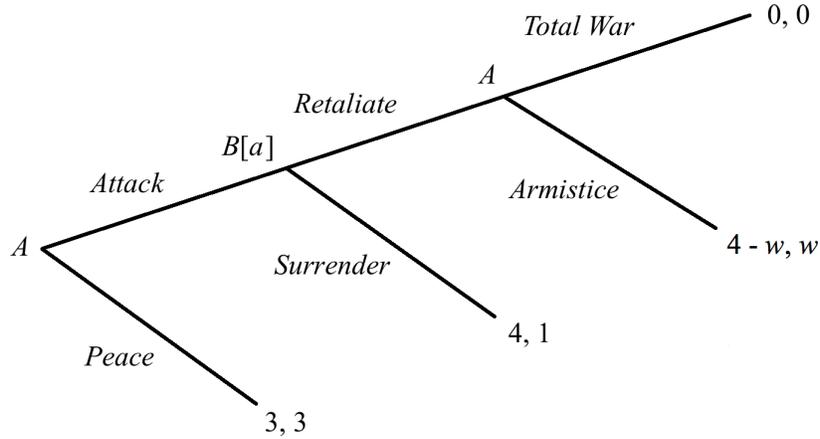
it gives the other country a greater incentive to behave more aggressively, in order to discourage its opponent from attacking it. (This is the defense motive.) Depending on how these two forces combine, the equilibrium behavior of strategic types will adjust differently to changes in the prior probabilities of the two countries being crazy (Proposition 2). Our results precisely characterize the net effects of the reputation and defense motives. This enables us to characterize comparative statics with respect to the model’s parameters, including payoffs (Proposition 3).

Our comparative static results are new, and provide important insights for the analysis of inter-state relationships. Our model is amenable to several extensions that might explain the determinants of the prior probability that a country is crazy. An example of such a determinant could be the country’s political institutions. For example, if democracies are perceived (perhaps incorrectly) to behave less aggressively than dictatorships (for instance, because people believe that democratic rulers represent the preferences of the median voter better than do dictators), our model would generate the Democratic Peace almost as a “self-fulfilling” phenomenon. We note, however, that there are a number of other factors that might influence a country’s prior probability of being perceived as crazy, besides political institutions. These include the country’s history, culture and other attributes of its ruling class. One of our most important results is the general finding that all of these factors affect the probability of war in the same direction as they affect the prior probability of the crazy type.

In the remainder of the introduction, we construct a simple model to illustrate our approach. Section 2 overviews the related literature. We present the model in Section 3. In Section 4, we study the comparative statics of the model. Section 5 concludes. All proofs omitted from the main text are in the Appendix.

## 1.1 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 it can escalate the conflict by choosing total war. Since all actions are uniquely labeled, we can identify terminal nodes with the actions that lead



**Figure 1.** Game Tree for the Illustrative Example

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. Note also 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 sequential rationality only with probability  $p \in (0, 1)$ ; for our purposes, we assume that  $p$  is close to 1. With complementary probability  $1 - p > 0$  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 (i.e., 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. 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, i.e.  $a = 0$ . But since the crazy type always chooses total war, country  $B$  would surrender for sure. But then the strategic type of country  $A$  would want to deviate to attack. Similarly, there is no equilibrium in which the strategic type of country  $A$  attacks for sure. Otherwise, given that both types of country  $A$  attack, country  $B$ 's posterior would be the same as its prior, i.e.  $a = p$ . 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. But then 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  $a = 1/w$ . But this pins down the equilibrium probability with which the strategic type of country  $A$  chooses to attack. That probability is  $(1 - p)/[(w - 1)p] > 0$ . This concludes the characterization of the sequential equilibrium.

Now, define the equilibrium probability of war to be the probability that the equilibrium outcome will be armistice or total war. This probability is

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

Note that the equilibrium probability of war goes to 0 as  $p$  goes to 1. Nevertheless, a relatively small probability that country  $A$  is the crazy type can lead to war with

significantly higher probability.<sup>5</sup> In addition, by inspecting (1), we see that the probability with which country  $A$  attacks is increasing in its prior probability of being the crazy type. Also notice that the probabilities with which country  $A$  attacks and 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  $B$  during a conflict.

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? The model that we build in this paper addresses all of these questions. It handles two-sided incomplete information with ease; its equilibrium predictions are almost always unique; it shows how the comparative statics results above easily generalize; and it identifies the strategic incentives behind them.

## 2 Related Literature

International relations scholars, at least from the time of Waltz (1959), have recognized the difficulty that sovereign states have in sustaining peace. A large literature, that includes contributions by Schelling (1963), Jervis (1978) and Kydd (1997) studies the extent to which security threats can be controlled in political anarchy. While Mearsheimer (2001) takes the pessimistic Hobbesian view that conflict is the natural state of affairs among nations, other authors such as Glaser (1992) express more cautious optimism. Some recent contributions to this debate are Baliga and Sjöström (2004), Chassang and Padró i Miquel (2008) and Acharya and Ramsay (2011), all of whom argue that fears and misperceptions may cause even rational states to abandon a peaceful arrangement. Despite very obvious modeling differences, these papers are

---

<sup>5</sup>For example, suppose that  $w = 1.1$  so that country  $A$  is relatively stronger than country  $B$ , and that there is a 1% prior chance that country  $A$  is the crazy type, i.e.  $p = 0.99$ . In this case, the equilibrium probability of war is slightly over 10%. This is because the probability with which the *strategic* type of country  $A$  attacks is slightly above 10% while country  $B$  retaliates with probability slightly larger than 90%. Therefore, even a small chance that country  $A$  is crazy may have an amplified effect on its equilibrium behavior.

actually closely related to our paper in their incorporation of dominant strategy types that parallel the role of our crazy types. However, while these papers model contagion due to higher order uncertainty, we focus on first-order uncertainty as modeled in the reputation literature.

In a separate class of models, war is viewed as an outcome of bargaining failure. Early papers include Powell (1987), Banks (1990) and Fearon (1995). These contributions have had a lasting impact on the study of war. One of the important insights of this literature is that war cannot be an equilibrium outcome when both parties are able to locate a Pareto superior negotiated settlement. With this caveat in mind, our model shows that because of the presence of aggressive crazy types, bargaining cannot terminate hostilities even between the strategic types. Thus, 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). All of these previous papers 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 concessions 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 papers, yields unique equilibrium predictions enabling us to deliver a number of comparative statics results relating the probability of war with the prior probability of the crazy types, and the payoffs received by the strategic types.

Our paper is also closely related to an incisive paper by Patty and Weber (2006). These authors 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 as a result we are able to characterize the effect of crazy (or “irrational,” in the language of Patty and Weber) types on the equilibrium behavior of strategic types. This enables us to derive comparative statics results that can be used to formalize the qualitative implications of Patty and Weber’s (2006) claims about the democratic peace.

We note that ours is not the first paper on international security to study the effect of incorporating behavioral types into an otherwise rationalist model. In an early paper, 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). However, in contrast to this paper, our paper uses behavioral types to explain the occurrence, rather than deterrence, of war.<sup>6</sup> Nevertheless, in using behavioral types, our model is closely related to the literature on reputation. Early work in this literature (e.g. Fudenberg and Levine 1989; Kreps and Wilson 1982) introduces behavioral types to show that strategic types can benefit by pooling with the behavioral types. Specifically, strategic types achieve higher payoffs than they would absent the opportunity to build reputations. More recent work by Ely and Välimäki (2003) and Canes-Wrone, Herron and Shotts (2001) shows, on the other hand, that reputation effects may not always be payoff-improving. In these papers, strategic “good types” separate from “bad types,” which results in them receiving lower payoffs than they would achieve in the case where the bad-reputation effects were absent. In contrast to all of these papers, in our model strategic agents pool with the behavioral type, even though this is (in expectation) payoff-neutral for them, but is overall Pareto inferior to the case where pooling incentives are absent. In this way, our paper is most closely related to the work of Abreu and Gul (2000) on reputational bargaining that shows that a slight possibility of irrationality for either side has a pooling effect (similar to ours) that produces inefficient delays in bargaining.<sup>7</sup> More importantly, our paper exploits the uniqueness of equilibrium predictions associated with the use of behavioral types to derive interesting new comparative static results, identify the forces at work behind them, and characterizing the conditions under which such forces are qualitatively relevant.

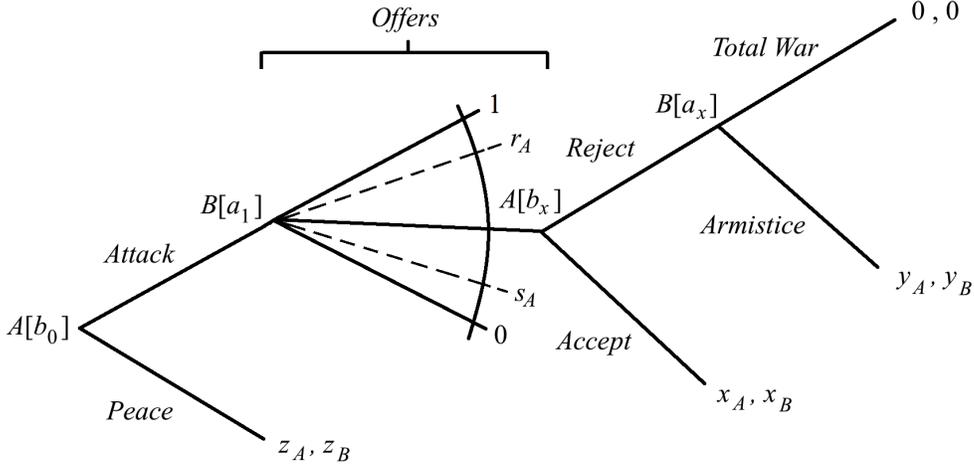
### 3 Model

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

---

<sup>6</sup>Incorporating behavioral types in otherwise complete information games has been part of an important research agenda in the game theoretic literature (see, e.g., Mailath and Samuelson 2006).

<sup>7</sup>A particularly compelling behavioral foundation for the crazy type that appears in our model is provided by Bénabou and Tirole (2009), who study belief distortions created by pride, dignity and wishful thinking about future outcomes, especially as they relate to intransigence in bargaining.



**Figure 2.** The Game Tree

the offer or escalate the conflict by rejecting. 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.

**Assumption 1.** War is costly:

- (i)  $z_A > y_A > 0$ , (ii)  $z_B > y_B > 0$ , and (iii)  $z_A + z_B > 1 > y_A + y_B$ .

Under Assumption 1, it is easy to see that with complete information, the game has a unique subgame perfect equilibrium whose only outcome 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.<sup>8</sup> We call this game  $G(a_0, b_0)$ , and we make the following assumptions on  $r_A$  and  $s_A$ .

<sup>8</sup>Our assumption on 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).

**Assumption 2.** Greedy crazy types:

$$(i) 1 - y_B > r_A > z_A \text{ and } (ii) y_A > s_A > 0.$$

Assumption 2(i) 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. Assumption 2(ii) states that the crazy type of country  $B$  makes an offer that is worse for country  $A$  than the outcome under armistice. Combining Assumptions 1 and 2,

$$1 > r_A > y_A > s_A > 0. \quad (2)$$

A behavioral strategy profile for the game  $G(a_0, b_0)$  is denoted  $\langle (\alpha, \alpha_x), (\beta, \beta_{TW}) \rangle$ , where  $\alpha$  is the probability with which the strategic type of country  $A$  chooses to attack;  $\alpha_x : X \rightarrow [0, 1]$  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 over the set of feasible offers made by the strategic type of country  $B$ ; and  $\beta_{TW} : X \rightarrow [0, 1]$  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

$$a_1 \equiv \frac{\alpha a_0}{1 - a_0 + \alpha a_0}. \quad (3)$$

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

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

We denote country  $B$ 's belief that country  $A$  is strategic at the node where 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 1(ii) 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  $(\alpha, \alpha_x)$  and  $(\beta, \beta_{TW})$  are sequentially rational given  $(a_0, a_1, a_x, b_0, b_x)$ .

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

$$\begin{aligned} \underline{a} &= \frac{1 - r_A - y_B}{1 - s_A - y_B} & \bar{a} &= \frac{1 - r_A - y_B}{1 - y_A - y_B} \\ \underline{b} &= \frac{z_A - s_A}{r_A - s_A} & \bar{b} &= \frac{s_A}{y_A} + \frac{(y_A - s_A)(z_A - s_A)}{y_A(r_A - s_A)}. \end{aligned}$$

It is easy to verify from Assumptions 1 and 2 that  $1 > \bar{a} > \underline{a} > 0$  and  $1 > \bar{b} > \underline{b} > 0$ . 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.<sup>9</sup>

**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  $\alpha = 0$ ,

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

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

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

- (iii) If  $\bar{b} > b_0 > \underline{b}$  and  $a_0 > \underline{a}$ , then in every equilibrium, we have  $\alpha = \frac{1 - a_0}{a_0} \cdot \frac{1 - r_A - y_B}{r_A - s_A}$ ,  $\alpha_x$  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  $x_A \neq s_A, r_A$ .

---

<sup>9</sup>It is clear from the proof of the proposition that the knife-edge cases for which the equilibria are not characterized are cases of indifference. We ignore these cases in the interest of substantive emphasis, as none of our qualitative results depend on what happens in these cases.

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

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

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

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

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

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

*Proof.* See Appendix A.  $\square$

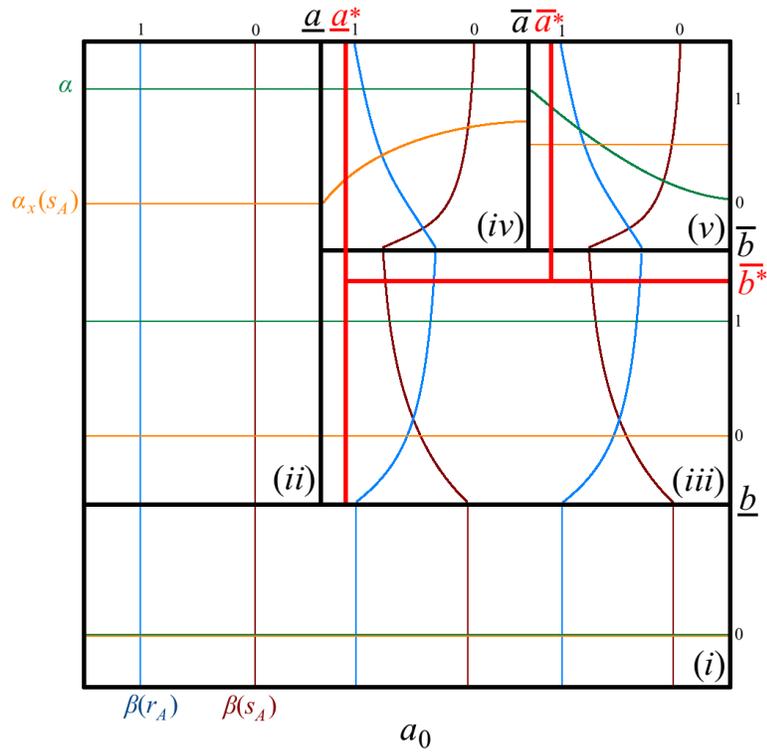
We summarize the main features of equilibrium as follows. If the prior probability that country  $B$  is strategic is very low,  $b_0 < \underline{b}$ , then the strategic type of country  $A$  will not attack to avoid having to bargain with a crazy type of country  $B$  (case (i)). However, the strategic type of country  $A$  will 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 country  $B$  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 very likely to be strategic,  $a_0 > \underline{a}$ , and country  $B$  is moderately likely to be strategic,  $\bar{b} > b_0 > \underline{b}$ , 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  $s_A$ , that 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,  $b_0 > \bar{b}$ , 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  $b_0 > \bar{b}$ , then country  $A$  mixes between attacking

and taking the peaceful outcome, and 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 we mentioned in Section 2, our model differs from other models in that incomplete information 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: (i) the incentive to build a reputation for being crazy by mimicking the crazy type (which deters the opponent from behaving aggressively), and (ii) the incentive to take an action that is optimal against one of the two types of the opponent. For a wide range of values of prior beliefs,  $a_0$  and  $b_0$ , equilibrium requires that these two forces to 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 and such a probability pins down equilibrium behavior and yields uniqueness. Put differently, uniqueness stems from the fact that in equilibrium the strategic type has to mimic the crazy type with a probability that results in an equilibrium level of reputation that leaves the strategic type of the opponent indifferent between its future 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, 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 the crazy type, nor optimal against one of the two types of country  $A$ . Therefore, no other offer is ever made in equilibrium.

Finally, Proposition 1 contains the important insight that uncertainty concerning a country's type leads to inefficient conflict with positive probability, and that these conflicts are not easily settled. To see this, note that  $\alpha$ , 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 prescribes that armistice arises with positive probability in regions  $(iv)$  and  $(v)$ . Thus, not only the strategic type of country  $A$  initiates conflicts, but also the strategic type of country  $B$  may fail to make offers that would settle these conflict without creating greater inefficiencies.



**Figure 3.** Equilibrium behavior in the five regions defined in Proposition 1:  $\alpha_x(s_A)$  in orange,  $\beta(r_A)$  in blue and  $\beta(s_A)$  in brown. The figure also depicts, in red, changes in the thresholds that will arise in the comparative statics analysis of Section 4.2.

## 4 Comparative Statics

In Section 4.1 below, we report the comparative statics of the equilibrium values of  $\alpha$ ,  $\alpha_x(x_A)$  and  $\beta(x_A)$  with respect to the prior probabilities  $a_0$  and  $b_0$ . (The comparative statics of  $\beta_{TW}(x_A)$  are trivial since it is always equal to 0.) Then, in Section 4.2, we study the comparative statics of the equilibrium behavior with respect to the payoff split from armistice  $(y_A, y_B)$ . We are interested in these comparative statics because we interpret this payoff split as reflecting the relative military strength of the two countries during war.

### 4.1 Comparative Statics with Respect to $a_0$ and $b_0$

Proposition 2 contains the comparative static results for the equilibrium behavior with respect to  $a_0$  and  $b_0$ . Its proof follows from the equilibrium characterization provided in Proposition 1 and its visual representation is provided in Figure 3. For ease of reference (and reading), we state the result referring to the regions  $(i)$  to  $(v)$  depicted in the figure, rather than repeating the thresholds that define these regions.

**Proposition 2.** The comparative static with respect to  $a_0$  and  $b_0$  are as follows:

- (1)  $\alpha$  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 region  $(iii)$  and  $(v)$  and strictly increasing in  $b_0$  in regions  $(iv)$ .
- (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  $(iv)$  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.  $\square$

Proposition 2 is an immediate corollary of the equilibrium characterization in Proposition 1. Among other things, it states that the probability with which strategic 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 of Section 1.1. 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  $\alpha$  remains the same. In this case, after an attack, country  $B$  believes that country  $A$  is irrational with lower probability than before. This in turn makes the concessional offer less attractive. So, to keep country  $B$  indifferent between the three offers, the equilibrium value of  $\alpha$  must in fact adjust downward 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,  $\alpha$ , in order to maintain the level or reputation,  $a_1$ , constant. The same intuition holds for 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  $\alpha$  is constantly equal to 0 and 1 in these respective cases.

The comparative static of  $\beta(\cdot)$  with respect to  $b_0$  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 of the concessional offer,  $\beta(r_A)$ . So, 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 will be greater 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 it is increasing in  $b_0$  in regions (iii) and (v), where the defense motive prevails. Furthermore, the probability mass lost by  $r_A$  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  $B$  being crazy is high, while  $y_A$  receives more of the mass when this probability is relatively low and it is harder for the strategic type of country  $B$  to mimic the crazy type (regions (iv) and (v)).

Finally, the strategic type of country  $A$  will always accept offers  $r_A$  and  $y_A$ . Thus, 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. This goal is achieved by increasing the probability of rejecting the greedy offer  $s_A$ .

## 4.2 Comparative Statics with Respect to the Payoff Splits

So far we have analyzed the model keeping the payoffs fixed. Here, we study the comparative statics of the equilibrium predictions with respect to changes in  $y_A$  while holding the sum  $y_A + y_B$  constant. The reason that we are interested in these comparative statics is because we interpret the payoff split  $(y_A, y_B)$  as a measure of the relative military strength of the two countries. The more powerful a country is, the larger the share of payoffs it can expect to receive following a war that ends in armistice. Throughout 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,

$$y_A + y_B = y_A^* + y_B^* = \bar{y}, \text{ and } y_A^* > y_A.$$

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  $\alpha$ ,  $\alpha_x$  and  $\beta$  denote the equilibrium quantities evaluated at payoff profile  $h$ , and let  $\alpha^*$ ,  $\alpha_x^*$  and  $\beta^*$  denote the same quantities evaluated at payoff profile  $h^*$ . Similarly,  $\bar{a}$ ,  $\underline{a}$ ,  $\bar{b}$  and  $\underline{b}$  are the thresholds in (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 inequalities:  $\underline{a}^* > \underline{a}$ ,  $\bar{a}^* > \bar{a}$ ,  $\underline{b} = \underline{b}^*$ , and finally  $\bar{b}^* < \bar{b}$ . Now, define  $\tilde{a} = \bar{a}^* \left( \frac{1-s_A-y_B}{1-s_A-y_B^*} \right)$ . It is easy to verify that  $\tilde{a} \in (\bar{a}, \bar{a}^*)$ .

The following proposition characterizes the comparative statics of the equilibrium behavior with respect to the payoff split  $(y_A, y_B)$ .<sup>10</sup>

**Proposition 3.** 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  $b_0 > \bar{b}$ , 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  $b_0 > \bar{b}$  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 B.  $\square$

The proof of the proposition in Appendix B 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, as depicted in Figure 3. 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 3 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, the reputation of country  $B$  for being crazy is so high that it discourages country  $A$  from initiating a conflict. Part (2) states that the probability of making the greedy offer is weakly decreasing in country  $A$ 's

---

<sup>10</sup>As before, we ignore the knife-edge cases  $a_0 = \underline{a}, \bar{a}, \underline{a}^*, \bar{a}^*$  and  $b_0 = \underline{b}, \bar{b}, \bar{b}^*$ .

relative military strength. The reason for this is that an increase in  $y_A$  reduces country  $B$ 's equilibrium expected payoff from mimicking the greedy 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 that we described previously, the strategic type of country  $B$  must decrease the probability of concession. Depending on which of these two 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 the defense motive for country  $A$  serves as a countervailing force vis-a-vis the increase in expected payoff associated with the rejection of the greedy offer.

Thus, due to the defensive motive, an increase in the military strength of country  $A$  may lead country  $B$  to be less accommodating and to actually test which of the two types country  $A$  is, by making the intermediate offer  $y_A$ . Notice that this result holds when there is little uncertainty that country  $A$  is strategically rational ( $a > \bar{a}^*$ ). Intuitively, this is exactly the range of parameters in which the strategic type of country  $B$  is more concerned about defending itself against attacks from the strategic type of country  $A$ .

### 4.3 Implications for Applications and Extensions

Before concluding the paper, we suggest ways in which variation in the prior probabilities  $a_0$  and  $b_0$  can be viewed as arising from a variety of unmodeled factors.

For example, the literature on international conflict stresses the link between the behavior of countries in conflict (and crisis) situations and their political institutions, culture and characteristics or motivation of leaders. Such linkage may generate inter-

esting corollaries to our comparative statics results above. In particular, suppose that democratic political institutions are commonly perceived (perhaps even incorrectly) to better represent the preferences of citizens. In this case, it would be reasonable to assume that the decision-makers in democracies are more likely than their autocratic counterparts to internalize the costs of wars. In our model, this would result in systematic differences in the prior probabilities  $a_0$  and  $b_0$  across countries with different political institutions. Specifically, these probabilities would be higher for democratic regimes than for autocratic ones. Given the comparative static results highlighted in Proposition 2, this would lead to the (somewhat self-fulfilling) conclusion that the democratic peace follows from nothing more than the popular belief that democracies are more likely to internalize the costs of conflicts (regardless of whether or not this belief is well-founded). Thus, if we associate democratic regimes with higher prior probabilities  $a_0$  and  $b_0$ , our model predicts that the democratic peace can arise through two different channels. On the one hand, the reputation motive for country  $A$  described above implies that a rise in  $a_0$  increases the probability of country  $A$  choosing peace,  $1 - \alpha$ . On the other hand, an increase in  $b_0$  above  $\bar{b}$  strengthens the reputation motive for country  $B$  and increases the probability it makes an offer that settles the dispute, avoiding escalation into an actual conflict (that is, it reduces the probability with which the second decision node of candidate  $A$  is reached).

Similarly, a change in leadership may result in changes in a country's perceived likelihood of being the crazy type, causing significant changes in foreign policy. For example, when the young Kim Jong-un replaced his father Kim Jong-il, opinion was divided on whether North Korea could be expected to behave more or less aggressively than before. On the one hand, some analysts argued that the new leader would be more likely to behave aggressively in order to solidify his power domestically. On the other hand, others suggested that he may try to initiate a new conciliatory course in North Korea's relationships with Japan, South Korea and the US. Our model highlights how these alternative explanations can be justified not only by a change in the exogenous probability of the North Korean leader being crazy, but also by a change in his strategically rational behavior determined by the net combined effect of the reputation and the defense motives. We regard the exact characterization of the net effect of these motives, described in regions  $(i)$  to  $(v)$  of Figure 3, to be the central contribution of our paper.

## 5 Conclusion

This paper constructs 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."

Unlike the previous literature, our model yields unique equilibrium predictions, which enable us to derive a number of new, but natural, comparative static results. In particular, we identify two countervailing effects that play a role in determining how often countries pretend to be crazy: the reputation motive, and the defense motive. On the one hand, the defense motive pushes a country to behave aggressively in order to shield itself from aggressive behavior by the opponent. On the other hand, the extent to which a country can pretend to be crazy is limited by its incentive to maintain its ability to establish a reputation. Depending on which force prevails, conflicts may arise and persist, exacerbating the inefficiencies associated with war.

# 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

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

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

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

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

(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,

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

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  $1 - r_A > y_B$ . 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  $B$  from making this offer is

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

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

$$a_1(1 - y_A - \varepsilon) + (1 - a_1)y_B, \quad (\text{A4})$$

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 the the payoff in (A4) is greater than the payoff in (A3).  $\square$

**Lemma 2.** In any equilibrium of the game, the support of  $\beta$  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  $\beta$  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 the support of  $\beta$ . Therefore, we can assume throughout the remainder of this proof that  $a_1 > 0$ .

Suppose  $\beta$  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,  $\beta$  must put zero probability mass on  $(r_A, 1]$ .

Suppose that  $\beta$  puts positive mass on  $(y_A, r_A)$ . Then, there exists  $x_A \in (y_A, r_A)$  in the support of  $\beta$ . 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 the support of  $\beta$ . Consequently,  $\beta$  cannot put positive probability on the interval  $(y_A, r_A)$ .

Now suppose that  $\beta$  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 from 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

$$a_1(1 - y_A - \varepsilon) + (1 - a_1)y_B > y_B \quad (\text{A4})$$

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

**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  $b_0 > \bar{b}$  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,  $\beta$  puts positive probability only on a subset of  $\{s_A, y_A\}$ . But because  $0 < s_A < y_A < z_A$  by Assumptions 1(*i*) and 2(*ii*), 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  $r_A$ :

$$1 - r_A = a_1(1 - y_A) + (1 - a_1)y_B \quad (\text{A5})$$

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

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

where the inequality follows from substituting the expression for  $\bar{a}$ , simplifying and using Assumption 2(i) and (ii). 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  $A$  from 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

$$\beta(s_A) \leq \frac{1 - b_0}{b_0} \frac{s_A}{y_A - s_A} \quad (\text{A7})$$

which follows from noting that  $b_x(s_A)$  is given by (6) in the main text. Since  $b_0 > \bar{b} > \underline{b}$ , 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 2(ii). 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

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

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 strictly inclusion holds because  $a_0 > \underline{a}$  by assumption. But  $\alpha \in (0, 1)$  implies the indifference condition

$$\begin{aligned} b_0[(1 - \beta(s_A))r_A + \beta(s_A)s_A] + (1 - b_0)s_A &= z_A \\ \iff \beta(s_A) &= 1 - \frac{z_A - s_A}{b_0(r_A - s_A)} \end{aligned} \quad (\text{A9})$$

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

**Lemma 4.** If  $b_0 < \underline{b}$  then the equilibrium set is characterized by

$$\alpha = 0, \alpha_x \text{ given by } (*) \text{ above, and } \beta(r_A) = 1.$$

*Proof.* Lemma 3(iv) implies  $\alpha = 0$ . Then Lemma 1(iii) implies  $\beta(r_A) = 1$ . Then Lemma 1(ii) implies that  $\alpha_x$  is given by (\*). Moreover, it is easy to verify that the given specifications for  $\alpha$ ,  $\beta$  and  $\alpha_x$  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.  $\square$

**Lemma 5.** If  $a_0 < \underline{a}$  and  $b_0 > \underline{b}$  then the equilibrium set is characterized by

$$\alpha = 1, \alpha_x \text{ given by } (*) \text{ above, and } \beta(r_A) = 1.$$

*Proof.* Lemma 3(iii) implies that  $\beta(r_A) = 1$ . Then by Lemma 1(ii),  $\alpha_x$  is given by (\*). Finally, it is easy to verify that country A's expected payoff from attack is  $b_0r_A + (1 - b_0)s_A$ , 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  $\alpha$ ,  $\beta$  and  $\alpha_x$  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.  $\square$

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

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

*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

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

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

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

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

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

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

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

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

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

This expression on the right hand side of (A14) is non-negative if and only if  $b_0 \geq \bar{b}$ . 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 to offering  $r_A$  is  $1 - r_A > y_B$  by Assumption 2(*i*). 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 (A12). Since we showed in Step (1) that  $\beta(y_A) = 0$ , and we know from Lemma 3(*iv*) that  $\alpha > 0$ , we need

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

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

*Step (3):* We now calculate the equilibrium values of  $\alpha$ ,  $\beta$  and  $\alpha_x$ . We have shown that  $\alpha_x(s_A) = 0$ , and  $\beta(y_A) = 0$ , which implies that  $\alpha_x$  is given by (\*) above. Since  $a_0 > \underline{a}$ , Lemma 3(iii) implies  $\beta(r_A) < 1$ . Therefore, by Lemma 2 and Lemma 3(i), we need  $\beta(r_A), \beta(s_A) > 0$ . These results imply the indifference condition (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 (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.  $\square$

**Lemma 7.** If  $b_0 > \bar{b}$  and  $\bar{a} > a_0 > \underline{a}$ , then in every equilibrium we have

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

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

*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, by 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 (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

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

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

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

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.  $\square$

**Lemma 8.** If  $a_0 > \bar{a}$  and  $b_0 > \bar{b}$  then the unique equilibrium is characterized by

$$\alpha = \frac{1-a_0}{a_0} \frac{1-r_A-y_B}{r_A-y_A}, \beta(s_A) = \frac{1-b_0}{b_0} \frac{s_A}{y_A-s_A}, \beta(y_A) = 1 - \frac{z_A-b_0y_A}{b_0(r_A-y_A)} - \frac{(1-b_0)s_A}{b_0(y_A-s_A)}$$

$$\beta(r_A) = \frac{z_A-b_0y_A}{b_0(r_A-y_A)}, \text{ and } \alpha_x(x_A) = \begin{cases} 0 & \text{if } x_A < y_A \text{ and } x_A \neq s_A \\ \frac{y_A-s_A}{1-s_A-y_B} & \text{if } x_A = s_A \\ 1 & \text{if } x_A \geq y_A \end{cases}$$

*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

$$\begin{aligned} a_0(1 - y_A - y_B - \varepsilon) &> 1 - r_A - y_B \\ \iff a_0(1 - y_A - \varepsilon) + (1 - a_0)y_B &> 1 - r_A \end{aligned} \quad (\text{A17})$$

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 (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 (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 (A10), which

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

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

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

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

Furthermore, it is easy to verify that because  $b_0 > \bar{b}$  the expressions for  $\beta(y_A)$  and  $\beta(r_A)$  given by (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  $b_x$ .  $\square$

## B. Proof of Proposition 3

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 changes the thresholds  $\underline{a}$ ,  $\bar{a}$ ,  $\bar{b}$  and  $\underline{b}$  and the equilibrium behavior characterized by  $\alpha$ ,  $\alpha_x$  and  $\beta$ . 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),  $\alpha$  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),  $\alpha$  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),  $\alpha$  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 only 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  $\alpha$  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  $b_0 > \bar{b}^*$ . In this case, we have  $\beta(s_A) = 1 - \frac{z_A - s_A}{b_0(r_A - s_A)}$  while  $\beta^*(s_A) = 1 - \frac{1 - b_0}{b_0} \frac{s_A}{1 - s_A}$ . One can verify that in the case we are analyzing, the quantity  $b_0(\beta(s_A) - \beta^*(s_A))$  is increasing in  $b_0$ , and converges to 0 as  $b_0 \rightarrow \bar{b}^*$ . Therefore, it must be that  $\beta(s_A) > \beta^*(s_A)$ .

(3) The result is straightforward in all cases except the case in which  $a_0 \in (\underline{a}^*, \bar{a}^*)$  and  $b_0 \in (\bar{b}^*, \bar{b})$  and the case in which  $a_0 > \bar{a}^*$  and  $b_0 \in (\bar{b}^*, \bar{b})$ . In the first case,  $\beta(x_A) = 0$  for all  $x_A \neq r_A, s_A$ ; so, the result follows from case 2 above. 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,  $b_0(\beta(r_A) - \beta^*(r_A))$  is increasing in  $b_0$ , and converges to 0 as  $b_0 \rightarrow \bar{b}^*$ . Therefore, it must be that  $\beta(r_A) > \beta^*(r_A)$ .

(4) Since  $\beta(y_A) = 0$  except when  $a_0 > \bar{a}$  and  $b_0 > \bar{b}$ , 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  $b_0 > \bar{b}$ . 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  $\alpha_x(s_A) - \alpha_x^*(s_A)$  is strictly decreasing in  $a_0$  on the interval  $(\bar{a}, \tilde{a})$  and is equal to 0 when  $a_0 = \tilde{a}$ . The result follows instantly.  $\square$

## References

- [1] Abreu, D. and F. Gul (2000). “Bargaining and Reputation,” *Econometrica*, 68:1:85-117.
- [2] Acharya, A. and K. Ramsay (2011). “The Calculus of the Security Dilemma,” *Quarterly Journal of Political Science*, forthcoming.
- [3] Alt, J., R. Calvert and B. Humes (1988). “Reputation and Hegemonic Stability: A Game Theoretic Analysis,” *American Political Science Review*, 82:2:445-66.
- [4] Baliga, S. and T. Sjöström (2004). “Arms Race Negotiations,” *Review of Economic Studies*, 71:2:351-69.
- [5] Banks, J. (1990). “Equilibrium Behavior in Crisis Bargaining Games,” *American Journal of Political Science*, 33:3:599-614.
- [6] Bénabou, R. and J. Tirole (2009). “Over My Dead Body: Bargaining and the Price of Dignity,” *American Economic Review*, Papers and Proceedings, 99:2:459-65.
- [7] Canes-Wrone, B., M. Herron and K. Shotts (2001). “Leadership and Pandering: A Theory of Executive Policymaking,” *American Journal of Political Science*, 45:3:532-50.
- [8] Chassang, S. and G. Padró i Miquel (2008). “Conflict and Deterrence under Strategic Risk,” *NBER Working Paper*, no. 13964.
- [9] Ely, J. and J. Välimäki (2003). “Bad Reputation,” *Quarterly Journal of Economics*, 118:3:785-814.
- [10] Fearon, J. (1995). “Rationalist Explanations for War,” *International Organization*, 49:3:379-414.
- [11] Fey, M. and K. Ramsay (2010). “Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict,” *American Journal of Political Science*, 55:1:149-69.
- [12] Fudenberg, D. and D. Levine (1989). “Reputation and Equilibrium Selection in Games with Patient Players,” *Econometrica*, 57:4:759-78.
- [13] Glaser, C. (1992). “Political Consequences of Military Strategy: Expanding and Refining the Spiral and Deterrence Models.” *World Politics*, 44:4:497-538.
- [14] Jackson, M. and M. Morelli (2009). “The Reasons for Wars - an Updated Survey,” in C. Coyne (ed.) *The Handbook on the Political Economy of War*, Elger Publishing.

- [15] Jackson, M. and M. Morelli (2007). "Political Bias and War," *American Economic Review*, 97:4:1353-73.
- [16] Jervis, R. (1978). "Cooperation under the Security Dilemma," *World Politics*, 30, 2, 167-214.
- [17] Kreps, D. and R. Wilson (1982). "Reputation and Imperfect Information," *Journal of Economic Theory*, 27:2:253-79.
- [18] Kuhn, H. (1953). "Extensive Games and the Problem of Information," in H. Kuhn and A. Tucker (eds.) *Contributions to the Theory of Games, Vol. II, Annals of Mathematics Study 28*. Princeton: Princeton University Press.
- [19] Kydd, A. (1997). "Game Theory and the Spiral Model," *World Politics*, 49, 3, 371-400.
- [20] Leventoglu, B. and A. Tarar (2008). "Does Private Information Lead to Delay or War in Crisis Bargaining?" *International Studies Quarterly*, 52:3:533-53.
- [21] Mailath, G. and L. Samuelson (2006). *Repeated Games and Reputations: Long Run Relationships*, Oxford: Oxford University Press.
- [22] Mearsheimer, J. (2001). *The Tragedy of Great Power Politics*, W.W. Norton.
- [23] Patty, J. and R. Weber (2006). "Agreeing to Fight: An Explanation of the Democratic Peace," *Politics, Philosophy and Economics*, 5:3:305-20.
- [24] Powell, R. (2002). "Bargaining Theory and International Conflict," *Annual Review of Political Science*, 5:1-30.
- [25] Powell, R. (1987). "Crisis Bargaining, Escalation and MAD," *American Political Science Review*, 81:3:717-36.
- [26] Reiter, D. (2003). "Exploring the Bargaining Model of War," *Perspectives on Politics*, 1:27-43.
- [27] Schelling, T. (1963). *The Strategy of Conflict*, New York, NY: Oxford University Press.
- [28] Schultz, K. (1999). "Do Democratic Institutions Constrain or Inform? Contrastive Two Institutional Perspectives on Democracy and War," *International Organization*, 53:2:233-66.
- [29] Slantchev, B. (2005). "Military Coercion in Interstate Crises," *American Political Science Review*, 99:4:533-547.
- [30] Waltz, K. (1959). *Man, State and War*, New York, NY: Columbia University Press.