BALANCING AGAINST THREATS WITH MORAL HAZARD

BRETT BENSON
ADAM MEIROWITZ
KRISTOPHER W. RAMSAY

PRELIMINARY

Abstract. In this paper we take a new approach to the study of alliances. Since at least 1648, an important group of alliances have involved one country pledging to the other some amount of aid in times of war. Taking a view of alliances as a form of decentralized insurance arrangements that indemnify targets against the cost of wars with potential aggressors, we develop a theory that explains why particular security agreements form and why the commitments look as they do. Our theoretical model considers both the effects of moral hazard on alliance partners and the deterrence possibilities against third parties. Our analysis explains why alliances tend to form between large countries, or between large countries and small countries, but not between small countries.

1. Introduction

In the study of international conflict, the role of alliances has always been central. Do country balance against threats? Do allies need to have an interest in other's survival in order to form agreements? Is the threat of attack sufficient to generate an alliance, or does it depend on the nature of the targets being threatened? In this paper we develop a theory of security alliances that explains the commitment that countries make to each other and the types of countries that tend to form alliances. Our theoretical analysis starts from the observation that many alliances are a form of insurance contract. Much like how car insurance pays cash to the policy holder if he is in an accident, an alliance agreement describes how much aid the ally will provide to the attacked party if there is a war. One important difference between the insurance provided by alliances and other forms of insurance is that the provider of insurance is just another country in the international system. That is, alliances can be viewed as a form of decentralized insurance. Like other kinds of insurance, alliances generate moral hazard. That is, an alliance agreement between two countries can distort a decision-maker's incentives such that they are more likely to start wars if threatened than they would be absent the alliance agreement. But unlike insurance of other kinds, alliances can also distort the behavior of countries not in the alliance. In particular, an alliance agreement between two target countries to support one another in time of war may deter a third country from initiating a conflict with either. This is not the case in the standard insurance contract where having car insurance, for example, does not change the incentives of other drivers to crash into you. The possibility of deterrence from these alliances creates incentives for partner countries to form agreements that distort their ally's behavior in a way that is costly, by generating more types that go to war, but because of the resulting change in the behavior of the challenger such distortions lead to beneficial outcomes for at least one of the alliance partners. The strength, interaction, and consequences of the distortions created by alliance agreements depend in important ways

on many aspects of the international environment. Specifically we focus on the effects of the inherent risk of a threat, the possibility of deterrence, the distribution of power among targets and challengers, and the costs and stakes of military conflict.

While a significant body of work already exist on the theory of alliances, it has mainly emphasized explaining a country's commitment to their ally and the implications of the alliance agreement for conflict. Research on commitment addresses questions regarding the reliability of leaders' promises of military assistance. Scholars have shown that a reputations for honoring today's promises benefits one's relationships with prospective allies in the future (Snyder, 1990; Smith, 1995). The reliability of commitments also can depend on domestic audiences and their interest in imposing costs on leaders when they renege and damage the national reputation (Smith, 1995).[1]

Empirically we also know that content of agreements is important because studies have shown that the nature of an agreement affects alliance reliability during a conflict. Challenging a literature that claimed only 25% of alliances agreements were, in fact, carried out, Leeds, Long and McLaughlin-Mitichell (2000) Leeds et. al. 2000 show alliances are likely to be reliable when the specific antecedent conditions of formal provisions have been activated (Leeds, 2003). Moreover, the likelihood of conflict varies dramatically depending on what promises are included in the alliance agreement (Leeds, 2003; **?**). So like much international law, it appears that most alliances are reliable most of the time, but understanding the wide variety of details of the commitments made by parties is important. That said, explanations of the content of alliance members' commitments is understudied and less well-understood. To begin to understand what drives alliance commitments, this paper examines leaders' decisions about what to promise an ally when the commitment to deliver the promise is credible.

[1]Another mechanism of explaining alliance reliability is offered by Morrow (1994), whose signaling framework demonstrates that alliances can alert non-alliance members that allies' interests are more aligned than believed, or that the alliance might be a signal of allies' costly peacetime preparation for war, which increases the credibility of allied intervention during wartime because it increases the benefits received in war relative to peace.

Content refers to many things and here it is important to be specific about what forms of content we wish to explain. Alliance provisions themselves may be written with offensive or defensive pledges (Niou and Ordeshook, 1994; Smith, 1995) or they might include specific obligations related to the tightness of alliance members' military coordination (Morrow, 1994). Another way to think about the type of commitment made in an alliance is to identify what alliance members promise to do when *casus foederis* has been triggered. Many alliances commit leaders to the complete defense of fellow alliance members. Others explicitly give alliance members discretion to determine whether to intervene and how much assistance to provide once war has occurred. Another set of alliances–those examined here–specify precise levels of military assistance that states promise to transfer to the attacked alliance member. For example, the 1656 Treaty of Defensive Alliances between Brandenburg and France stipulated that if Brandenburg was attacked, France would provide 5000 men, 1200 horses, and artillery or equal compensation to Brandenburg in exchange for Brandenburg transferring 2400 men and 600 horses to France if it was attacked.

Another distinguishing characteristic of alliance agreements is their structure. Morrow (1991) predicts that asymmetric alliances more readily form and survive. Siverson and Tennefoss (1984) show alliances are most deterrent when challengers target minor powers who have major power allies. Nevertheless, actual alliances are also formed between powerful countries. We investigate the potential for alliance between different configurations of powerful and weak states.

Must states share security interests for them to ally? In most studies of alliances some other regarding preferences emerge. So alliances are formed by countries that share a preferences for compelling concessions from target countries and form alliance agreements that contain offensive provisions (Smith, 1995; Niou and Ordeshook, 1994). Other alliances are designed to deter prospective adversary. One such alliance is a formal extended deterrence

commitment in which a protg faces a challenge from an adversary, and a third-party defender, whose own security is not at risk, specifies some amount of military assistance to transfer to the protg if it is attacked Huth (1991). Most existing formal models of alliance analyze this type of problem. To induce alliance between the third-party and protg, these models typically assume the prospective allies share preferences for the protg's security (Smith, 1995; Morrow, 1994).

In this paper, we focus on a specific type of formal agreement, which we call a security alliance. We analyze the incentives faces by two self-interested leaders who may not have any common interests, except a desire to manage an external threat, to exchange security promises. In such a world, both prospective allies face the possibility of being challenged by the same adversary, but ex ante neither knows who will face a future crisis. Our approach, therefore, departs from the model of extended deterrence alliance, focusing instead on the class of security alliances which more typically occupied the attention of early neorealist work (Walt, 1987; Christensen and Snyder, 1990; Snyder, 1984).

Snyder's (1984) account of security alliance emphasizes both the security enhancing features of an agreement to mutually threatened states as well as noting the risks of moral hazard among allies. In forming and honoring alliances, states balance their need for security and the value of their reputation against the risk of being entrapped in an undesirable war. Christensen and Snyder's (1990) chain-ganging alliances examine the similar alliance structure and also highlight the problem of moral hazard.[2]

In what follows we analyze the content of the agreements made by leaders who anticipate some risk of being attacked by a common adversary, and in response exchange securities to insure themselves against the security risk. Thus, it is possible to pin down the precise amount of military assistance leaders promise to transfer to the attacked ally. Our emphasis

[2]It is less clear in Christensen and Snyder's (1990) account why states need to enter a formal alliance. If states' survival hinges on their ability to balance against mutual threats, so much so that they would join undesirable wars alongside states with similar balancing interests, then why must those states pay the additional contracting price of formal alliance to make a promise they will already keep?

on securities exchanges is most similar to Conybeare's (1992) concept of portfolio analysis, in which alliances are viewed as investment portfolios formed for the purpose of diversifying security risks. The portfolio model, which does not specify any strategic behavior, predicts that the risk of an alliance portfolio is decreasing in the number of allies. Our approach goes beyond the portfolio model by allowing allies to bargain over securities given some exogenous risk of being attacked and adding a conflict subgame in which those securities impact war payoffs. We also introduce an additional dimension of risk by incorporating moral hazard. Security assurances encourage allies to fight in the crisis subgame because they increase the payoff to war. Therefore, decisions to promise securities to a prospective ally depend on the amount of risk created by that ally's behavior.

In what follows we show how the risk of attack and moral hazard affect leaders decisions to form alliances and how much assistance will be promised. We can also draw conclusions about what states will ally. Finally, the model leads to predictions about the effect of alliances on the probability of war.

2. Model

Consider a situation with three countries, a challenger and two potential targets. With probability π_j target j has a crisis with the challenger. Once a crisis starts the challenger decides whether or not to escalate by threatening target j. If the challenger chooses the status quo, and thus fails to escalate with a threat, the crisis ends peacefully and there is no change in the stakes controlled by the two sides. We, therefore, normalize the payoff for the status quo to 0 for the challenger and 1 for the target.

If the challenger makes a threat, on the other hand, the target country can choose to resist the threat and fight to keep the status quo, or capitulate and give in to the challenger's threat. If the target fights the dispute is settled by war. In a war the challenger wins against target j with probability p_j and pays a cost k_j. We assume that the target countries have private information regarding their costs of war in the crisis. Each target has a cost of war

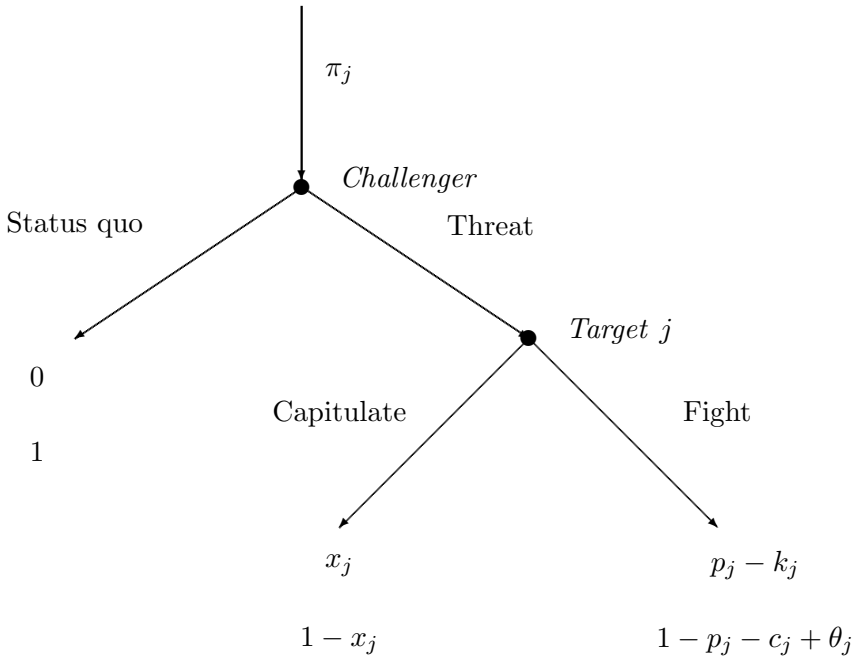

FIGURE 1. Alliance game

$c_i \in [0, \bar{c}]$. We let $F(c)$ denote the prior on this cost and assume it has a continuous density. Given a threat, the target can avoid war by capitulating, but then the target must forfeit the "stakes" of the crisis, x_j, keeping the fraction $1 - x_j$ for themselves. For simplicity we assume that the challenger's costs of fighting are known. The game is depicted in Figure 1.

Finally, to this crisis game we add an *ex ante* stage where the two potential targets can make an agreement regarding promises to come to each others aid in the case that one or the other is engaged in a war. In general, the agreement will constitute a war contingent transfer from one target to another of an amount $\theta_j \geq 0$. We can think of the *ex ante* alliance agreement as a form of decentralized insurance. Much like a spouse or parents might help each other or their children financially if they lose their job or experience an accident, we can think of these countries as making agreements that transfer resources

from one player to the other in the case of war. An alliance agreement then is a pair, $\boldsymbol{\theta} = (\theta_1, \theta_2) \in R^2_+$. These security alliances are made *ex ante* in the sense that the players do not know their costs of war at the time of agreement, though they have beliefs about the distribution of these costs.

Given an agreement, $\boldsymbol{\theta} \in \Theta^2$, we let $U_i(\boldsymbol{\theta})$ denote the expected payoff to country i from this treaty. Naturally if the parties do not agree to a treaty, then their payoffs are given by $u_i(0, 0)$.

In some situations it will be useful to distinguish between alliance agreements that are Pareto and those that are not.

Definition 1. *A treaty $\boldsymbol{\theta}$ Pareto dominates treaty $\boldsymbol{\theta'}$ if $u_i(\boldsymbol{\theta}) \geq u_i(\boldsymbol{\theta'})$ for $i = 1, 2$ with a strict inequality for at least one of the players. A treaty is Pareto efficient if no treaty Pareto dominates it. Finally, a treaty, $\boldsymbol{\theta}$ is Pareto dominant if for all other treaties, $\boldsymbol{\theta'}$ one of the following is true: $\boldsymbol{\theta}$ Pareto dominates $\boldsymbol{\theta'}$ or $u_i(\boldsymbol{\theta}) = u_i(\boldsymbol{\theta'})$ for $i = 1, 2$.*

The two countries in our model reach an agreement by bargaining over the levels of support θ_i and θ_j. We consider the situation where the bargaining protocol is the alternating–offers procedure of Rubinstein (1982) with an risk of break down (Binmore, Rubinstein and Wolinsky, 1986). In period 0, player i makes a proposal that j may accept or reject. If j accepts the game ends and the crisis game is played. If j rejects the proposal, no agreement is reached and the game continues. Continuation of the game then depends on the realization of a lottery over termination through a crisis game without treaties and the next period of bargaining. With probability z the crisis game is played and the game ends with payoffs, $U_i(0, 0)$. With probability $1 - z$, there is no crisis in this period and the bargaining phase of the game proceeds to period $t + 1$

Finally an equilibrium of our game is a subgame perfect equilibrium of the bargaining protocol where the continuation values are determined by Baysian-Nash equilibrium play

in the crisis subform. We will call a complete assessment consisting of strategy profiles and a set of beliefs for the Bayesian game a *perfect Bayesian* equilibrium.

3. Results

To analyze incentives in the alliance problem, we begin by analyzing the crisis subforms taking the alliance agreement as fixed.

Given a pair of contracts $\boldsymbol{\theta} = (\theta_1, \theta_2)$, the target's decision to go to war is well defined. In particular, target j will capitulate in equilibrium if

$$1 - x_j \geq 1 - p_j - c_j + \theta_j$$

$$c_j \geq x_j - p_j + \theta_j.$$

perfect Bayesian rationality implies that if θ_j is greater than $c_j - x_j + p_j$, then j will choose to go to war to maintain the status quo. From this condition it is clear that those target countries who anticipate some chance of war have a utility that is increasing in the commitments they extract from their ally. For the alliance partner, however, the alliance commitments create different incentives. Because alliances make wars more attractive, the targets will fight back more often, and the ally will more frequently need to transfer resources to their partner. This effect of θ_j on a country's action is analogous to *moral hazard* in insurance markets, the fact that a player is being indemnified in the case of war can make it choose to fight wars that it would otherwise avoid.

Another important aspect of the alliance problem is the way alliances influence the decisions of challengers. Obviously, it could be the case that the presence or absence of an alliance agreement between two targets has no affect on the decision of potential challengers. On the other hand, an alliance agreement may make threatening sufficiently less attractive for the challenger that it chooses to make no threat during the crisis. We will call an alliance agreement (θ_1, θ_2) *deterrent* if it the challenger would make a threat if

$\theta_j = 0$, but does not make a threat at the given this alliance agreement. Importantly, if deterrence is achieved both allies are better-off. The target in the crisis is never challenged and the ally never has to follow through on its agreement because war does not happen. How these various possibilities and incentives interact are a the center of our theory of alliance agreements.

3.1. **Large targets.** We start by considering the case where the targets can form alliance agreements that change the behavior of the challenger in some circumstances. In particular, consider the case where

$$p_j - k_j < 0 \text{ and } F(x_j - p_j)(p_j - k_j) + (1 - F(x_j - p_j))x_j > 0 \tag{1}$$

hold for both targets. Under these conditions, if the challenger believes that a threat will lead to war for sure, it will choose to keep the status quo. If, on the other hand, the challenger believes that the odds of the target fighting back without a defensive treaty are smaller, then it is willing to risk war in order for a chance at acquiring the concession. This is a situation where the challenger is potentially deterrable. Like the way the large trading countries affect world prices, we will say that a target for whom there exists an alliance agreement that deters a challenger is *large*. When condition (1) holds for both targets we say both targets are large.

In this situation there exist treaties, (θ_1, θ_2) that induce targets countries to fight regardless of their costs. In particular whenever

$$\theta_j \geq \bar{c}_j - x_j + p_j$$

all types of each target will fight if challenged and thus the challenger will keep the status quo in any realized crisis. This conclusion does not rely on the boundedness of the support of costs. In particular, since $p_j - k_j < 0$ a probability of fighting that is less than 1 is still sufficient to deter the challenger and thus, without loss of generality, we can consider the

case where $c_j \in R_+$ and still find a θ_j for which $F(x_j - p_j + \theta_j)(p_j - k_j) + (1 - F(x_j - p_j + \theta_j))x_j < 0$. Let $\underline{\theta}_i$ be the smallest amount of support that target i needs to receive from target j to deter a threat.

For the case of two large target countries and under the additional condition that targets for whom war is costless fight to maintain the status quo, we then can see that there are no equilibria where $\theta_j^* < \underline{\theta}_j$.

Lemma 1. *There is no perfect Bayesian equilibrium where* $0 < \theta_j^* < \underline{\theta}_j$.

Proof. Suppose not. That is, suppose there were some equilibrium where at period t the two targets reached an alliance agreement where, for some j, $\theta_j^* < \underline{\theta}_j$. There are two cases.

Case (1): Let $\theta_1^* < \underline{\theta}_1$ and $\theta_2^* < \underline{\theta}_2$. Then at some period t some i proposes an agreement (θ_1^*, θ_2^*) such that this proposal is accepted. As a result

$$u_j(\theta_1^*, \theta_2^*) \geq z u_j(0,0) + (1-z) W_j(t+1)$$

where $W_j(t+1)$ is j's continuation value for the game that starts after she is the veto player in period t.

Now suppose at time t country i proposes $(\underline{\theta}_i, \theta_j^*)$. First, $u_j(\underline{\theta}_1, \theta_2^*) > u_j(\theta_1^*, \theta_2^*)$ because on the path j never has to pay $\underline{\theta}_i$, but pays $\theta_1^* > 0$ with positive probability, while at the same time j's payoff to their own crisis does not change. Thus we can conclude that $(\underline{\theta}_i, \theta_j^*)$ is accepted by j at time t.

All that remains is to show that at t, i is strictly better-off proposing $(\underline{\theta}_i, \theta_j^*)$. For country i the expected utility of (θ_i^*, θ_j^*) is

$$\begin{aligned}(2)\quad \pi_i[F(x_i - p_i + \theta_i^*)(1 - p_i - \hat{c}_i(\theta_i^*) + \theta_i^*) + (1 - F(x_i - p_i + \theta_i^*))(1 - x_i)] \\ + (1-\pi)[1 - F(x_j - p_j + \theta_j^*)\theta_j^*],\end{aligned}$$

where $\hat{c}_i(\theta_i) = \mathbb{E}[c_i | c_i < x_i - p_i + \theta_i]$ denotes the expected cost of player i conditional on the cost being sufficiently low that i fights.

The expected utility of i for $(\underline{\theta}_i, \theta_j^*)$ is

$$\pi_i[1] + (1 - \pi_i)[1 - F(x_j - p_j + \theta_j^*)\theta_j^*]. \tag{3}$$

By assumption $[F(x_i - p_i + \theta_i^*)(1 - p_i - \hat{c}_i(\theta_i^*) + \theta_i^*) + (1 - F(x_i - p_i + \theta_i^*))(1 - x_i)] < 1$, and this proposal is a profitable deviation, a contradiction.

Case (2): Now suppose there is an equilibrium with $\theta_i \geq \underline{\theta}_i$ and $\theta_j < \underline{\theta}_j$. There are two sub-cases.

Sub-case (i): Suppose this agreement is reached at a time t when j is the proposer. By an argument parallel to the one in Case (1), j has a profitable deviation, a contradiction.

Sub-case (ii): Now suppose that this agreement is reached at a time t when i is the proposer and $\theta_j^* > 0$. If i increases the proposed support to country j to some $\theta_j \geq \underline{\theta}_j$, then j will never be attacked and i's expected payout to j is $0 < (1 - \pi_i)F(x_j - p_j + \theta_j^*)\theta_j^*$. This is a profitable deviation for i, a contradiction. If $\theta_j^* = 0$, but j is a proposer in some future period, j will reject this offer to get the lottery over zero and being the proposer at some future t'. This contradicts that the agreement is reached at period t.

Together these cases prove the lemma. □

If the treaties are sufficiently large, on the path of play, the challenger never advances a threat and the agreement is never activated. Following such a treaty the targets get their maximal possible payoff associated with never facing a challenge and never making any transfer of resources to the opponent. To complete our analysis, we show that agreements that are deterrent for both targets are reached without delay.

Lemma 2. *Suppose for both targets* $1 - p_i > 1 - x_i$ *and condition (1) are satisfied. Then in every perfect Bayesian equilibrium alliance agreements are reached without delay.*

Proof. Suppose not. That is, suppose there is an agreement in a perfect Bayesian equilibrium that is reached with positive probability after time $t = 0$. From Lemma 1 we know that this perfect Bayesian equilibrium agreement will be deterrent for both targets. Let j be the veto player in the first period. From period 0 the veto players expected utility is

$$\begin{aligned} &zu_i(0,0) + (1-z)([zu_i(0,0) + (1-z)[zu_i(0,0) + (1-z)[\ldots + (1-z)^s u_i(\theta_1^d, \theta_2^d)]]] \ldots \\ &= zu_i(0,0) + (1-z)u_i(0,0) + (1-z)^2 u_i(0,0) + \ldots + (1-z)^s u_i(\theta_1^d, \theta_2^d) \\ &= zu_i(0,0) \sum_{t=0}^{s-1} (1-z)^t + (1-z)^s u_i(\theta_1^d, \theta_2^d). \end{aligned}$$

To show that a deterrent agreement would be accepted in period $t = 0$, suppose it were rejected by i. Then

$$\begin{aligned} u_i(\theta_1^d, \theta_2^d) &\leq zu_i(0,0) \sum_{t=0}^{s-1} (1-z)^t + (1-z)^s u_i(\theta_1^d, \theta_2^d), \\ u_i(\theta_1^d, \theta_2^d) &\leq \frac{zu_i(0,0) \sum_{t=0}^{s-1} (1-z)^t}{1-(1-z)^s}, \\ &= zu_i(0,0) \Big[\frac{1-(1-z)^s}{z} \frac{1}{1-(1-z)^s} \Big], \\ &= zu_i(0,0) \frac{1}{z} = u_i(0,0). \end{aligned}$$

But my assumption $u_i(\theta_1^d, \theta_2^d) > u_i(0,0)$, a contradiction.

We can thus conclude that if a target i would accept the deterrent equilibrium agreement in period $s > 0$, then it would accept it in period 0. We are left to show that the proposer in period $t = 0$ is better off making the proposal today. But this is clear from an argument parallel to the one that shows the veto player is willing to accept a deterrent proposal today if it is willing to accept it in the future.

This contradiction proves the lemma. □

Using these two lemmas we can establish the following result for alliance agreements between large states.

Proposition 1. *Suppose that* $1-p_i > 1-x_i$ *for both targets and condition* (1) *are satisfied. Then in every perfect Bayesian equilibrium alliance agreements are no delay and they completely deter challenges.*

Proof. Suppose that $1 - p_i > 1 - x_i$ for both targets and condition (1) are satisfied. By Lemma 1, all equilibrium agreements completely deter and by Lemma 2 they are achieved with no delay. □

3.2. **Small targets.** Next we consider the case where the challenger cannot be deterred. This is the case where the challenger finds fighting better than the status quo. That is, we assume that $p_j - k_j > 0$ In this case, we know that for each j their expected utility as a function of values of θ_i and θ_j are given by

$$u_i(\theta_1, \theta_2) = \pi_i(F(x_i - p_i + \theta_i)(1 - p_i - \hat{c}_i(\theta_i) + \theta_i) + (1 - F(x_i - p_i + \theta_i))(1 - x_i)) + (1 - \pi_i)(1 - (F(x_j - p_j + \theta_j)\theta_j))$$

The key difference between this case and that of large targets is that the cone, of Pareto dominant treaties that could costlessly deter attacks is empty. An essential fact, however, is that the null treaty $(0, 0)$ is Pareto efficient. Every treaty that makes one player better off makes the other player worse off. This conclusion stems from the fact that any treaty other than $(0, 0)$ involves a distortion in that moral hazard induces at least one of the players to reject an offer and fight an inefficient war for some interval of costs that occurs with positive probability. The fact that the treaty compensates the fighting country only represents a redistribution of this inefficiency and proves that the $(0, 0)$ agreement is Pareto

efficient. Most importantly for our analysis the ally that is committed to making the transfer internalizes the inefficiency.

A feature of a perfect Bayesian equilibria to the alternating-offers bargaining games is that a weak participation or individual rationality constraint must be satisfied in equilibrium. If i is the proposer in period t then j will not accept an offer that does not give her at least as high a crisis game payoff as the null treaty. But the proposer would also never propose a treaty that gave her a lower payoff. Thus we see that every treaty, $\boldsymbol{\theta}$ that is reached in an equilibrium must satisfy the constrain, $u_i(\boldsymbol{\theta}) \geq u_i(0,0)$ for $i = 1, 2$.

Proposition 2. *Assume that $p_j - k_j > 0$. There is a perfect Bayesian equilibrium and in every such equilibrium the alliance agreement is $(0,0)$.*

Proof. An immediate consequence of the pair of inequalities that precede the proposition is that if a treaty, $\boldsymbol{\theta}$ is passed in any equilibrium then

$$u_1(\boldsymbol{\theta}) + u_1(\boldsymbol{\theta}) \geq u_1(0,0) + u_2(0,0).$$

But since the null treaty is Pareto efficient, this implies that no treaty which is not payoff equivalent to the null treaty can be accepted in any equilibrium. □

The contrast between the situation where there are two large states is that the only effect of an alliance is that it increases the probability of war through moral hazard, which is a social bad. In particular, there are no gains from the challenger's response to the alliance, no increase in the total welfare of the targets, and hence no value for alliances. In this case we see that security alliances born out of external threats depend crucially on the generation of new "rents" or surplus for the allies generated by changing the decision of the challenging country. In the case of two large countries this means stopping war altogether. Next we consider the case where one country is large and the other is small.

3.3. **One large and one small target.** Consider the situation where 1 is large enough so that if $\theta_L \geq \underline{\theta}_L$ initiation against 1 will be deterred, but 2 is sufficiently small such that no treaty will deter a challenger from threatening it. Specifically, consider an alliance bargaining game where in period 0 country 1 proposes an alliance agreement $(\theta_1, \theta_2) \in R^2_+$. Country 2 accepts this agreement and ends the bargaining phase or rejects the agreement.

If a proposed agreement is rejected in any period, then with probability z a crisis occurs and each target plays the crisis game with the agreement $(0, 0)$ and gets an expected payoff of $u_i(0, 0)$. With probability $1 - z$, country 2 gets to make a proposal $(\theta_1, \theta_2) \in R^2_+$, taking on the role of the proposer, and country 1 now faces a decision in this second period that is symmetric to the decision problem of country 2 in the first. It must either accept or reject the current proposal.

Our first result is that in the asymmetric case, in every perfect Bayesian equilibrium the large country ends up with a deterrent agreement.

Lemma 3. *Suppose that $1 - p_i > 1 - x_i$ holds for both target countries and for country L condition (1) holds, but not for country s. Then in every perfect Bayesian equilibrium the alliance agreement has*

$$\theta^*_L \geq \underline{\theta}_L.$$

Proof. Suppose not. That is, suppose that $1 - p_i > 1 - x_i$ holds for both target countries and for country L condition (1) holds, but not for country s, and suppose there is a perfect Bayesian equilibrium agreement (θ^*_L, θ^*_s) and $\theta^*_L < \underline{\theta}_L$. There are two cases.

Case (1): First suppose that at some time $t \geq 0$ the two targets reach an agreement of (θ^*_L, θ^*_s) and $\theta^*_L < \underline{\theta}_L$.

Sub-case (i): Suppose L is the proposer in period t. As s has accepted (θ^*_L, θ^*_s)

$$u_s(\theta^*_L, \theta^*_s) \geq z u_s(0, 0) + (1 - z)W(t + 1).$$

Evaluating s's expected utilities at this profile and $(\underline{\theta}_L, \theta_s^*)$ one sees

$$\begin{aligned} & E[u_s(\underline{\theta}_L, \theta_s^*)] - E[u_s(\theta_L^*, \theta_s^*)] \\ &= \pi_L - \pi_L(1 - F(x_L - p_l + \theta_L^*)\theta_L^*) \\ &= \pi_L F(x_L - p_L + \theta_L^*)\theta_L^* > 0 \end{aligned}$$

and s will also accept $(\underline{\theta}_L, \theta_s^*)$.

All that remains is to show that the large country is better-off proposing $(\underline{\theta}_L, \theta_s^*)$. As the large country's utility is increasing in θ_L this is a profitable deviation contradicting that (θ_L^*, θ_s^*) is a perfect Bayesian equilibrium when the agreement is accepted in period t and L is the proposer.

Sub-case (ii): Suppose that s is the proposer at period t. By a similar argument as above, if the large country accepts (θ_L^*, θ_s^*), then it will accept $(\underline{\theta}_L, \theta_s^*)$.

We are left to show that when s is the proposer at time t, s's expected utility is greater if it proposes $(\underline{\theta}_L, \theta_s^*)$. As above, we see that the difference in s's expected utility from these two alliances is

$$F(x_L - p_L + \theta_L^*)\theta_L^* > 0$$

and s is better-off with the agreement $(\underline{\theta}_L, \theta_s^*)$ because s never has to pay any transfer in this case. This contradiction shows there is no perfect Bayesian equilibrium with an agreement (θ_L^*, θ_s^*) where s is the proposer of the accepted offer at time t.

Case (2): The second case considers the situation where no agreement is ever reached and the target countries's payoffs are $u_i(0, 0)$. This means that there is some even period where L makes a proposal of $(0, 0)$ or some other $(\hat{\theta}_L, \hat{\theta}_s)$ which is rejected by s. In this period s's expected utility of rejecting is $u_s(0, 0)$.

Suppose L proposes $(\underline{\theta}_L, \epsilon)$. For s, this proposal raises its expected utility and would be accepted. This is also a strictly profitable deviation for L, contradicting that there is a $(0, 0)$ or perpetual disagreement.

This completes the proof of the lemma.

□

From this lemma we see that the gains from an alliance for the large country are always captured. The question that remains is how these gains are distributed. If we fix $\theta_L^* \geq \underline{\theta}_L$ and define

$$\bar{\theta}_s = \{\theta_s | u_L(\theta_L^*, \theta_s) = u_L(0,0)\}$$

Then the two target countries are bargaining over a pie of size $\bar{\theta}_s > 0$. Let the set of possible agreements be

$$\Theta = \{(y_1, y_2) \in R^2 | y_1 + y_2 = \bar{\theta}_s \text{ and } x_i \geq 0 \text{ for } i = L, s\}.$$

We can then think of bargaining over the terms of an alliance agreements as countries alternate in proposing elements of Θ. The alliance bargaining problem then becomes a model of alternating offers bargaining with risk of breakdown. Following the notation of Osborne and Rubinstein (1990), if z is the probability that the negotiations end after a rejection, then we will denote an alliance bargaining game with the risk of break down at a probability z as $\Gamma(z)$.

To give our first result regarding the characteristics of equilibrium alliance agreements we need some notation. Let $v_i(\theta_s, 1)$ be the amount of support given to the small country by the large country in period 0 that makes them indifferent between this particular agreement at the beginning of the game and (possibly) getting the settlement θ in period 1. That is for each player we have

$$(v_i(\theta_s, 1), 0) \sim_i (\theta_s, 1).$$

Proposition 3. *(Existence and Uniqueness) If*

$$\max\{\max_{\theta_s} |\frac{u_s'(\theta_s))}{u_s'(\hat{\theta}_s^s)}|, \max_{\theta_s} |\frac{u_L'(\theta_s))}{u_L'(\hat{\theta}_s^L)}|\} < \frac{1}{1-z} \tag{4}$$

then there is unique perfect Bayesian equilibrium to the alliance agreement game.

Proof. From Osborne and Rubinstein (1990) Theorem 3.4, we know that if an alternating offers bargaining game satisfies the conditions that

A1 Disagreement is the worst outcome.

A2 Pie is desirable.

A3 Time is valuable

A4 the preference order is continuous

A5 the preferences are stationary

A6 and there is increasing loss to delay

then the bargaining game has unique subgame perfect equilibrium. Conditions A1–A5 are obviously satisfied in the alliance game. The increasing difference condition requires,

$$\theta_s - v_i(\theta_s, 1), \tag{5}$$

where θ_s is country s's share of the surplus from the alliance agreement and v_i is the relation defined above. Let $\hat{\theta}_s^i$ be the agreement for country i that satisfies this condition.

For condition A6 to be true for the small state,

$$\frac{\partial \hat{\theta}_s^s}{\partial \theta_s} < 1.$$

Let the $v_s(\theta_s, 1)$ be denoted by $\hat{\theta}_s^s$. Then we have $\hat{\theta}_s^s$ implicitly defined by

$$u_s(\hat{\theta}_s^s) = z u_s(0) + (1 - z) u_s(\theta_s)$$

$$u_s(\hat{\theta}_s^s) - z u_s(0) - (1 - z) u_s(\theta_s) = 0$$

By the implicit function theorem we have

$$\frac{\partial \hat{\theta}_s^s}{\partial \theta_s} = (1 - z) \frac{u_s'(\theta_s)}{u_s'(\hat{\theta}_s^s)} \text{ for all } \theta_s. \tag{6}$$

A parallel argument for the large country implies we also need

$$(1-z)|\frac{u_L'(\theta_s)}{u_L'(\hat{\theta}_s^L)}| < 1 \text{ for all } \theta_s. \tag{7}$$

If z satisfies the condition of the proposition, then both of theres inequalities hold, and A6 of Osborne and Rubinstein's Theorem 3.4 is satisfied. As a result the alliance game has a unique perfect Bayesian equilibrium. □

An immediate implication of the previous proposition is that the alliance agreements can be described in the following way.

Corollary 1. *Let $\langle\langle x, t\rangle\rangle$ denote the lottery over getting a settlement x in the t^{th} period and bargaining breakdown before t. Then the unique equilibrium solution solves*

$$\langle\langle y^*(z), 0\rangle\rangle \sim_L \langle\langle x^*(z), 1\rangle\rangle \text{ and } \langle\langle x^*(z), 0\rangle\rangle \sim_s \langle\langle y^*(z), 1\rangle\rangle. \tag{8}$$

And country s accepts any proposal that $x_L \geq x_L^*(z)$ and country L accepts any proposal $y \leq y_s^*(z)$.

4. Alliances that change the probability of victory

Up to this point we have treated the alliance agreement as a cost sharing measure. In many ways this is a useful theoretical choice. In our analysis we show that even when an alliance between two potential targets does not change the payoffs to the challenger directly, the cost sharing and the resulting changes in target's incentives and actions that can change the challenger's incentives to make a threat. An alternative modeling strategy for alliances might involve the assumption that an alliance agreement $\boldsymbol{\rho} = (\rho_1, \rho_2)$ results in an increase in the probability that i *wins* a war by the quantity ρ_i. In this model then the set of possible alliances would involve $\rho \in [0, r_1] \times [0, r_2]$ with $r_i \leq 1 - p_i$. Let $c_i(r_i)$ denote the strictly increasing cost function capturing the cost to $-i$ of increasing the probability

that i wins by ρ_i. Like before, we may assume the costs are only borne if target i ends up at war with the challenger.

Much of the intuition from the private values case extends here. In particular, now, a natural notion of large countries is the case where it is possible increase the probability that i wins enough to deter initiation by the outsider. Thus the case of two large countries involves the assumption that for each i there exists a ρ_i such that $\rho_i \geq p_i - k_i$. In the case where both countries are large, there are alliances which fully deter war, but wind up costing nothing. If alliances of this form exist, and the second part of condition 1 holds, then every equilibrium must involve an alliance of this form.

The natural extension of our concept of small countries involves the assumption that $r_i < p_i - k_i$. In this case it is not possible for j to make i strong enough so that an outsider that believes i will fight rather than acquiesce prefers 0 over the lottery between x and war. Under these conditions it is not possible to deter the initiator but, in contrast to the case of private values, a treaty that changes the probability that i wins need not be inefficient. In particular, consider when ρ_i satisfies the condition $p_i - \rho_i > k_i$, so that the initiator is not deterred from attacking i. The promise ρ_i has two effects on i's payoffs. It expands the set of costs-types that will reject the offer (and thus fight) and it increases the payoff to i if it fights. Conditional on fighting, the payoff to i increases by ρ_i times the stakes of war (which are 1) and conditional on j paying $c_i(\rho_i)$ the value gained to i is exactly ρ_i. Accordingly, an alliance between two small countries can be efficient if and only if $c_i(\rho_i) \leq \rho_i$.

Thus, in the case of two small countries a non-trivial treaty is possible in equilibrium. In particular, 1 and 2 are now bargaining over treaties in which a treaty ρ results in the gain to i of $\pi_i F_i(x_i - p_i + \rho_i)\rho_i$ and it results in the cost to i of $\pi_j F_j(x_j - p_j + \rho_j)c_j(\rho_j)$. In order for i to be willing to accept (or propose) a treaty of this form in equilibrium it must provide weakly positive gains. Moreover, as condition this must be satisfied for

both players, any treaty that is accepted with positive probability in an equilibrium must simultaneously satisfy the inequalities

$$\pi_1(F_1(x_1 - p_1 + \rho_1)\rho_1 \geq \pi_2(F_2(x_2 - p_2 + \rho_2)c_2(\rho_2)$$
$$\pi_2(F_2(x_2 - p_2 + \rho_2)\rho_2 \geq \pi_1(F_1(x_1 - p_1 + \rho_1)c_1(\rho_1).$$

An immediate observation is that in the case where $c_1(\rho) = c_2(\rho) = \rho$, either neither inequality is satisfied or they are both satisfied with equality. In this case treaties may form but the yield the same payoffs as the trivial treaty, $\rho = (0, 0)$. We can then concluded that when the technology that takes a contribution ρ and turns it into an equal increase in the probability of victory for a small country, increasing the probability of victory though an agreement does not create a positive incentive for forming an alliance. The case where $c_i(\rho_i) > \rho_i$ for $i = 1, 2$, will also not support non-trivial treaties, as these treaties involve inefficiencies. We are left with the trivial case where the technology of war generates increases in the probability of winning at a higher rate then the cost of war aid, the small countries will have an incentive to increase their collective payoffs by supporting each other during fighting.

For example, consider a $c_i(\rho_i) = \alpha\rho_i$ for $i = 1, 2$ with $\alpha < 1$, and $x_1 = x_2 = x; p_1 = p_2 = p; \pi_1 = \pi_2$ with uniform distributions on the players costs. In this case the above system of inequalities becomes

$$\frac{x - p + \rho_1}{x - p + \rho_2} \geq \frac{\alpha\rho_2}{\rho_1},$$
$$\frac{\rho_2}{\alpha\rho_1} \geq \frac{x - p + \rho_1}{x - p + \rho_2}.$$

Here a set of alliances which are Pareto superior to the trivial treaty exist. For example alliances with $\rho_1 = \rho_2$ are in the interior of the set of treaties satisfying these constraints.

For our bargaining protocol some agreement in which $\rho_i > 0$ for at least one player (and both in protocols that are not too extreme) would surface. We stop short of characterizing such a model here, but make two observations. First, this possibility of a treaty between two small states can be interpreted as the natural extension of our result with at least one large state. When it is possible for the transfer between target country 1 and target country 2 to distort the behavior of an outsider treaties become supportable. But, analyzing the model where treaties affect outcome probabilities as well as costs obscures the fact that it is essential to include a technological benefit to the treaty for small countries to find treaties beneficial.

5. Conclusion

Taking a view of security alliances as a form of decentralized insurance and focusing on the commitments that allies make to one another in an environment where alliance aid leads to distortions from moral hazard we have four findings. First, when two large countries are threaten, they form alliances that look like what international relations scholars might call threat balancing. Each side commits to aid the other to a sufficient degree that the challenger is deterred from escalating a dispute. Second, we see when analyzing the case where both potential targets are small the ability of large states to manipulate the incentives of the challenger by making allies more aggressive was a key element to explaining how security commitments arise. In a world where there is no deterrence, alliances just generate more war and redistribute their costs. But since the cost of war is completely internalized by one of the two targets, these social welfare decreasing agreements cannot arise in equilibrium.

Third, when analyzing the asymmetric alliance case we see that large countries always are able to get an alliance agreement that generates private benefits through deterrence. The question of forming an alliance in this environment then turns on the bargaining between the now safe large country and the still threatened small country regarding how

much of this benefit will be returned in the form of commitments to the small country. Finally we see that when the risk of a crisis is severe, that is z is sufficiently large, there is unique equilibrium and we can make strong prediction regarding the bargaining outcome between targets. This, in part, may explain why many specific agreements are forged on the eve of a crisis and lay out in detail conditions of activation and levels and types of aid.

While we end by noting that, empirically, there are a wide variety of alliance agreements, many of which do not fall under the category we study, the basic agreement that represents an exchange of security guarantees represents an important and foundational class of alliance agreements. And, as is suggested by the empirical evidence, further understanding of the role of alliances in international politics will require a better understanding of why and what kind of agreements are written in this basic environment.

References

Binmore, Ken, Ariel Rubinstein and Asher Wolinsky. 1986. "The Nash Bargaining Solution in Economic Modelling." *The RAND Journal of Economics* 17(2):pp. 176–188.
URL: *http://www.jstor.org/stable/2555382*

Christensen, Thomas J. and Jack Snyder. 1990. "Chain Gangs and Passed Bucks: Predicting Alliance Patterns in Multipolarity." *International Organization* 44(2):pp. 137–168.
URL: *http://www.jstor.org/stable/2706792*

Conybeare, John A. C. 1992. "A Portfolio Diversification Model of Alliances: The Triple Alliance and Triple Entente, 1879-1914." *The Journal of Conflict Resolution* 36(1):pp. 53–85.
URL: *http://www.jstor.org/stable/174505*

Huth, Paul K. 1991. *Extended Deterrence and the Prevention of War*. Yale University Press.
URL: *http://books.google.com/books?id=FiveHAAACAAJ*

Leeds, Brett Ashley. 2003. "Do Alliances Deter Aggression? The Influence of Military Alliances on the Initiation of Militarized International Disputes." *American Journal of Political Science* 47(3):801–827.

Leeds, Brett Ashley, Andrew G. Long and Sara McLaughlin-Mitichell. 2000. "Reevaluating Alliance Reliability: Specific Threats, Specific Promises." *Journal of Conflict Resolution* 44(5):686–699.

Morrow, James D. 1991. "Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances." *American Journal of Political Science* 35(4):pp. 904–933.
URL: *http://www.jstor.org/stable/2111499*

Morrow, James D. 1994. "Alliances, Credibility, and Peacetime Costs." *The Journal of Conflict Resolution* 38(2):pp. 270–297.

URL: *http://www.jstor.org/stable/174296*

Niou, Emerson M. S. and Peter C. Ordeshook. 1994. "Alliances in Anarchic International Systems." *International Studies Quarterly* 38(2):167–191.

Osborne, Martin J and Ariel Rubinstein. 1990. *Bargaining and markets*. New York, NY: Academic Press.

Rubinstein, Ariel. 1982. "Perfect Equilibrium in a Bargaining Model." *Econometrica* 50(1):pp. 97–109.

URL: *http://www.jstor.org/stable/1912531*

Siverson, Randolph M. and Michael R. Tennefoss. 1984. "Power, Alliance, and the Escalation of International Conflict, 1815-1965." *The American Political Science Review* 78(4):pp. 1057–1069.

URL: *http://www.jstor.org/stable/1955807*

Smith, Alastair. 1995. "Alliance Formation and War." *International Studies Quarterly* 39(4):405–425.

Snyder, Glenn H. 1990. *Why Nations Cooperate; Circumstances and Choice in International Relations*. Ithaca, NY: Cornell University Press.

Snyder, Jack. 1984. *The ideology of the offensive: military decision making and the disasters of 1914*. Cornell Studies in Security Affairs Cornell University Press.

URL: *http://books.google.com/books?id=QvIfZaA3sq4C*

Walt, Stephen M. 1987. *The origins of alliances*. Cornell studies in security affairs Cornell University Press.

URL: *http://books.google.com/books?id=EuwgR-ogAHwC*