ISSUE LINKAGE AND INTERNATIONAL COOPERATION: AN EMPIRICAL EVALUATION

by

Paul D. Poast

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Political Science)
in The University of Michigan
2011

Doctoral Committee:

Professor William Robert Clark, Co-Chair
Professor James D. Morrow, Co-Chair
Professor Alan V. Deardorff
Professor Walter R. Mebane Jr.
Associate Professor Barbara Koremenos

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER I

Introduction

The literature on the international politics of issue linkage is old, venerable, and extensive. Wallace (1976) famously claimed that "linkage between unrelated or only loosely-related issues in order to gain increased leverage in negotiation is an ancient and accepted aspect of diplomacy" (Wallace 1976, p. 164). Sebenius (1983), drawing on several historic examples, states that "linkage is a prominent and venerable practice" (Sebenius 1983, p. 283). Recently, Tomz (2007) identified issue linkage as "a central idea from international relations theory" (Tomz 2007 p. 7).

Why is issue linkage important? Issue linkage - the simultaneous discussion of two or more issues for joint settlement - is a bargaining tactic used by states to achieve two objectives (Sebenius 1983). First, issue linkage can increase the probability of states reaching agreement.[1] Linking issues accomplishes this objective by creating benefits for a party that would otherwise find an agreement to be of little value. For example, in 2010 the Obama Administration employed issue linkage in its attempt to empty the Guantánamo Bay prison:

> "When American diplomats pressed other countries to resettle detainees, they became reluctant players in a State Department version of "Let's Make a Deal." Slovenia was told to take a prisoner if it wanted to meet with

[1] See Chapter 3 for citations

> President Obama, while the island nation of Kiribati was offered incentives worth millions of dollars to take in Chinese Muslim detainees, cables from diplomats recounted."[2]

Second, issue linkage can motivate states to remain committed to an agreement once it has been reached (Koremenos, Lipson, and Snidal 2001; Tomz 2007). For example, including a free trade provision in a greenhouse gas reduction treaty may induce all parties to not defect from their codified environmental obligations.

This positive view of issue linkage is not without its detractors. With respect to the ability of linkage offers to secure agreements, Morrow (1992) highlights how linkage offers can be interpreted as a sign of bargaining weakness, thereby undermining their effectiveness. With respect to enhancing the ability of states to commit to an agreement, scholars such as Downs, Rocke, and Barsoom (1996) would argue that states agree to add issues to a treaty only because agreement on the primary issue was likely to be reached in the first place. Moravcsik (1998) argues that when states decide to include linkage provisions in the text of a treaty, domestic opposition can relegate these provisions to nothing more than symbolic window dressing. Thus, the theoretical claim that issue linkages are beneficial for cooperation may be well known, but they also might not be true.

The truth is that, beyond some suggestive case studies and a few indirect statistical tests, there exists no direct and systematic evidence that issue linkages actually increase the probability of agreement and increase the likelihood that states will remain committed to their signed agreements. We do not know if issue linkages do in fact increase the probability of states reaching an agreement and, if so, by how much. Does issue linkage increase the probability of agreement marginally (by, for

[2] Shane, Scott and Lehren, Andrew. "Leaked Cables Offer raw Look at U.S. Diplomacy." *The New York Times.* November 28, 2010.

instance, 1 percent) or substantially (by, for instance, 50 percent)? Similarly, though past studies have explored whether states comply with treaty obligations, there is little direct and systematic empirical evidence that including linkage provisions in a treaty will bolster that treaty's credibility.

Why have no previous studies directly measured issue linkage's impact on the probability of states reaching a negotiated agreement (the formation effect of issue linkage) and adhering to that agreement (the commitment effect of issue linkage)? There are five difficulties that have prevented scholars from providing this systematic evidence. I will now summarize each.

The Problems Impeding the Empirical Investigation of Issue Linkage

The first problem is that evaluating the effect of issue linkage requires exploring what I call k-adic processes: events that involve more than two actors. Many negotiations are multilateral, but the standard unit of analysis in international relations is the state-to-state dyad. As will be shown, because the dyad is ill suited for analyzing such processes, one must develop and apply an alternative unit of analysis. I call this new unit of analysis the k-ad.

The second problem is that scholars must clearly identify that states have used issue linkage when forming an agreement. As will be discussed, this can be quite difficult, but data on military alliance treaties offers a solution. Specifically, the Alliance Treaty Obligations and Provisions (ATOP) database provides details on the provisions of all known alliance treaties formed since 1815. Some of these treaties contain provisions calling on the states to grant one another foreign aid or trade cooperation. I will argue that the inclusion of economic cooperation provisions in alliance treaties are clear instances of issue linkage.

The third problem, to borrow from Shelock Holmes, relates to 'the dogs that didn't bark': data collection efforts have focused almost exclusively on treaty negotiations that end in agreement. This criticism obviously applies to the ATOP dataset, which only codes instances in which states formed alliance treaties. However, identifying the effect these provisions have on the probability of an alliance treaty negotiation ending in agreement requires also considering instances in which such provisions are offered, but fail to close the deal. I will explain below how European diplomatic histories provide a means of creating a dataset of failed military alliance negotiations.

The fourth problem is difficulty empirically operationalizing the cooperation problems that, according to the international cooperation literature, motivate states to use issue linkage. In particular, the *enforcement problem* – the idea that the states will have an incentive to defect from the agreement once it is signed – can be especially difficult to capture empirically. I will argue below that states highly vulnerable to invasion and occupation offer a perfect way of operationalizing the existence of an enforcement problem in military alliances.

The fifth and final problem is that of missing treatment data: though the sources that I will use to identify failed alliance negotiations may accurately and fully capture all instances of negotiation, these sources may have failed to record offers of linkage. This could, in turn, impact estimates of the effect of linkage. I will show below how one can obtain bounds on the effect of issue linkage offers that account for the possibility of missing data on issue linkage offers.

Issue Linkage versus Logrolling

Though quantitative evidence on issue linkage is lacking in international relations, some may wonder if empirical insights into the effect of issue linkages can be gleaned from a related literature: logrolling in American Politics. Logrolling is defined as vote trading in order to create a legislative coalition in support of a bill. Logrolling occurs because legislators have differing intensities in preferences over proposals, meaning that if legislators voted sincerely, no proposals would pass.[3]

In some ways, the literature on logrolling mirrors that of issue linkage. First, it is quite old, with the first mention coming from Bentley (1907). Second, much of the early work on logrolling was theoretical (Buchanan and Tullock 1962, Coleman 1966, Mueller 1967, Haefele 1971, Riker and Brams 1973, Kramer 1977, and McKelvey 1976). Third, the early empirical evidence was largely qualitative and anecdotal (Mayhew 1966 and Ferejohn 1974). Fourth, though there have been attempts to quantitatively study vote trading's importance and impact, these attempts, like those in the issue linkage literature, have been unable to directly isolate the effect of logrolling.

The first quantitative study of logrolling came from Kau and Rubin (1979). They attempted to measure the influence of logrolling on Congressional votes, but they could not separate the effect of logrolling from ideology.[4] Stratmann (1992) was the first paper to identify logrolling statistically, but he looked only at vote trades among agricultural interests.

Thus, perhaps the best early quantiative study of logrolling is Stratmann (1995), who investigated the extent to which vote-trading agreements, purged of party loyalty or party pressure, helped the passage of bills that provided subsidies for special interests.[5] Stratmann points out that answering questions about vote trading and coalitions within a quantitative framework requires empirically distinguishing between party loyalty and logrolling coalitions that are organized within a party. For

[3]Stratmann 1995, p. 442.
[4]Kau and Rubin 1979, p. 381.
[5]Stratmann 1995, p. 447.

this reason, Stratmann is critical of empirical studies that include a party dummy variable in regression equations because if logrolling coalitions are organized within a party, such a dummy variable will measure both potential party loyalty or potential party discipline and membership in a logrolling coalition.[6]

Instead, drawing from data collected by Mayhew (1966), Stratmann analyzes a broad range of votes where logrolling has been reported, spanning two Congresses (1959 to 1962). Stratmann analyzes trades between groups favoring subsidies for city interests, labor interests, and farm interests. Since logrolling occurred over a series of votes, Stratmann uses a simultaneous three-equation probit model. Stratman finds, for example, that if one voted in favor of the amendment favored by the city interests (the dependent variable in one equation), that invidual was also likely to vote in favor of the amendment favored by labor interests (the dependent variable in another equation) and in favor of the amendment favored by farm interests (the dependent variable in the third equation).

Though a useful step in the direction of empirically capturing the effect of logrolling, Stratmann's study has two limitations. First, the estimated correlation between the error terms of the equations is quite large. This suggests a high and positive correlation among unobserved variables, thereby indicating that the same unobserved factors influenced the votes of the representatives. Stratmann recognizes this and admits that unobserved constituency variables, ideology, and party discipline could be driving his results. In short, it is not clear that Stratmann has isolated the effect of logrolling. Second, Stratmann is only analyzing votes where logrolling was known to have been used. How can Stratman know that the legislators' behavior was influenced by the logroll if he does not also consider instances in which a logroll was

[6] Stratmann 1995, p. 442.

not applied? Such data could be difficult to locate, but perhaps not any less difficult than identifying when the logroll was applied.

A more recent set of empirical studies is Evans (1994, 2004). Evans looks at the trading of 'pork barrel' projects for broader policy goals by applying a logit model to data on highway project votes. Specifically, Evans shows that the receipt of highway project monies increased the likelihood that a member of the House, when voting on the legislation that contained the project, supported the position of the leaders of the Public Works and Transportation Committee. However, one shortcoming of this study is that the "issues" are not fully separable. Both the primary issue (highway legislation) and the linkage issue (highway project monies) are both on the same topic: highways. In other words, Evans study shows that giving money for highway projects can encourage highways. Hence, in many ways this is showing that side-payments (i.e. money) can acquire votes, but it is not directly capturing the effect of logrolling (trading votes on different topics).

Overall, it appears that the empirical literature on logrolling in American Politics is at no more of an advanced state than that of issue-linkage in international relations. This suggests that attempts to identify the effect of issue linkage in international politics could also be of interest to scholars of American Politics.

Plan for the Dissertation

The overall goal of the three papers that follow is to empirically identify the effect of issue linkage, both on the probability of agreement formation and on the likelihood of commitment. Though measuring the effect of linkage will be directly addressed in chapters 3 and 4, both chapters require deep thinking about how to empirically evaluate multilateral events. Therefore, the chapter 2 focuses on this methodological issue. In chapter 2, I explain that exploring multilateral events requires making each event the unit of analysis. This stands in contrast to the common practice in quantitative international relations research, which is to divide the participants of multilateral events into a series of dyadic combinations. However, I also argue that using dyadic data is appropriate if the event being explored is bilateral.

After describing my approach for analyzing multilateral events, I begin discussing how I empirically measure the two effects of issue linkage: its effect on formation and its effect on commitment. The first effect is discussed in chapter 3. Chapter 3 begins by presenting a new dataset of successful and failed alliance negotiations from 1815 to 1945. While the successful military alliance negotiations come from existing data, identifying the failed negotiations required carefully reading and coding several prominent European diplomatic histories. I then code whether a negotiation participant offered to tie economic cooperation to the final alliance agreement. Finally, using matching and the techniques introduced in chapter 1, I find that, from 1860 to 1945, offering to expand the military alliance negotiations along an economic dimension increased the probability of agreement by 36 percentage points.

Chapter 4 uses buffer states (states located between two recently warring rivals) as a 'hard test' of the claim that issue linkage enhances treaty credibility. Because buffer states are especially prone to invasion and occupation, their (attempts at) alliance relations offer an ideal window through which to test the ability of issue linkage to enhance treaty credibility. More specifically, a buffer state's high vulnerability makes other states reluctant to remain committed to an alliance agreement with the buffer state. Using a variety of analysis techniques, I find that buffer states in alliances with trade provisions avoid occupation and invasion at a higher rate than buffer states in other alliance arrangements, that third parties attack buffer states in alliances with trade provisions at a lower rate than in other alliance arrangements, and that buffer states in alliances with trade provisions experience fewer opportunistic violations of the alliance terms by their alliance partners. Thus, since linkage can help buffer states to form credible commitments, then linkage should be a useful tool in nearly any context.

CHAPTER II

K-ads: The Unit of Analysis For Studying Negotiations

2.1 Introduction

Since issue linkage is a bargaining tactic, identifying its effect necessarily means treating the negotiation as the unit of analysis. This much seems obvious. What is more problematic, however, is that negotiations can involve more than two actors. Indeed, some negotiations take place between just two parties, such as the talks leading to the first strategic arms limitation treaty (SALT) (involving just the United States and the Soviet Union). However, the negotiations leading to the signing of the Treaty of Versailles, for example, involved over 30 countries. This fact is problematic because when analyzing multilateral events, quantitative international relations scholars typically divide the actors involved into a series of dyadic relations (i.e., a U.S.-France-U.K. event is converted into three events: U.S.-France, U.S.-UK, and France-U.K.).[1] This subset of observations is then added to a set of purely dyadic observations. Though this practice can dramatically increase the size of datasets, using "dy"-adic data to analyze what are "k"-adic events leads to model misspecification

[1] Bremer (1992) was highly influential in making the dyad the most prominent unit of analysis in IR, though the practice of disaggregating multilateral events into dyadic observations is too widespread to summarize. One need only pickup a statistical study in international relations over the past few decades to find an example (assuming the study did not focus solely on bilateral relations). Promient examples that apply dyadic data to multilateral wars include Bremer (1992), Russett and Oneal (1997), Peceny, Beer, and Sanchez-Terry (2002), and Reiter and Stam (2002, 2003). Studies that have applied dyadic data to the creation of multilateral trade agreements include Mansfield, Milner and Rosendorff (2002) and Mansfield and Reinhardt (2003). Lai and Reiter (2000) and Leeds et al (2002) apply dyadic data to the creation of multilateral alliances. See Bennett and Stam (2000) for an excellent discussion of the promise and pitfalls of estimating dyadic data in international relations.

and, inevitably, statistical bias.

Dyadic data are known to violate the independence assumption underpinning many statistical estimation techniques. Dyadic data commits four major violations of the independence assumption. First, the observations in the dyad-year are temporally correlated (e.g. the Russia-Germany 1938 dyad and the Russia-Germany 1939 dyad; Beck, Katz, and Tucker 1998). Second, the dyads typically share unexplained heterogeneity (Green, Kim and Yoon 2001; Beck and Katz 2001; King 2001). Third, the dyads have monadic similarity (e.g. the presence of the United States in the U.S.-France and U.S.-Brazil dyads; Ward, Siverson, and Cao 2007; Hoff 2005; Hoff and Ward 2004). Fourth, Signorino (1999) highlights the failure of scholars to adequately capture the strategic interaction between nations that is implied by dyadic data.

Though accounting for such non-independence is critical for drawing proper inferences, I am not presenting an alternative procedure for modeling such spatial, strategic, temporal, or monadic interdependencies in the data. Such features of the data will still be present and must still be modeled. Instead, this paper highlights a prior, conceptual issue arising in the context of multilateral decision making processes - namely, if the data are formed by interactions among $k > 2$ actors, then a dyadic format will not reflect this process regardless of how one models other interdependencies. For example, suppose one accounts for strategic interdependence using an estimator based upon logit quantal response equilibria. In this case, the probability of each outcome is derived by multiplying the probabilities of the actions that lead to the outcome. If the outcome is the result of actions taken by $k = 4$ actors, then considering only the actions of $k = 2$ actors (i.e., using dyadic data) will fail to capture the true probability of an outcome.[2]

This particular limitation of dyadic data is not unknown to scholars. Croco and Teo (2005), using a series of case studies, highlight the inferential bias introduced by splitting multilateral events into dyadic observations. Gibler, Rider, and Hutchison (2005), citing Weede (1980), discuss how Wallace (1976, 1979) overstates the ability of arms races to escalate into wars because he disaggregates one event of arms race induced escalation into several events, thereby inflating the number of positive cases. Signorino (1999) also identifies this problem, pointing out that dividing a k-nation event into a series of dyadic observations of size $k(k-1)/2$ greatly expands the size of the dataset, but does so without adding new information and by introducing bias. However, none of these studies, nor any previous study to which the author is aware, has sought either to identify the size of the bias introduced by evaluating k-adic events with dyadic data, nor offer a suggestion for how one should alternatively structure the data. Instead, scholars continue to divide k-actor events into a series of dyadic observations.

This paper has two goals. The first is to illustrate the bias produced when analyzing k-adic processes with dyadic data. I show, using a Monte Carlo simulation under the simplest of conditions (a cross sectional dataset in which each grouping of countries has an independent ability to form an alliance and the decision making process is non-strategic) that one cannot recover a k-adic data generating process using *dy*-adic data. One must instead evaluate the data generating process using k-adic data. In other words, one must use a dataset containing all combinations of actors (i.e., actors A, B, and C can form four multi-actor combinations: AB, AC, BC, and ABC). Of course, if the number of potential actors is even moderately large

[2]See Gent (2007) and Findlay and Teo (2006) for examples of modeling strategic interactions in multilateral events.

(perhaps N=100), this can quickly produce a dataset with observations numbering in the millions, billions, or more.[3]

Thus, the second goal of the paper is to illustrate how choice-based sampling, an approach recommended by King and Zeng (2001a; 2001b) for analyzing "rare events" data, enables one to create and analyze k-adic datasets of manageable size. Specifically, one can estimate a sample consisting of all k-ads in which the dependent variable is coded 1 (indicating, for example, if members of the k-ad formed an alliance treaty or began a conflict) and a random sample of various sized k-ads in which the dependent variable is coded 0. This solution is not without costs. For instance, some measures, such as distance, are most easily understood in a dyadic context. However, intuitive tractability and data collection simplicity do not justify continued reliance upon flawed inferences.

This paper is organized as follows. First, using Monte Carlo simulations, section 2 illustrates the bias introduced when dyadic data is used to evaluate k-adic events. Though I place this simulation (and the subsequent simulations) in the context of alliance formation, this is intended simply to give the simulation a point of reference (the main statistical points they raise could be illustrated just as easily with randomly constructed covariates devoid of any substantive motivation). Conflict onset, the formation of international trade agreements, governing party coalitions, as well as numerous other subjects could be used to contextualize the simulation. Second, section 3 uses Monte Carlo simulations to illustrate how choice-based sampling can generate a feasibly sized dataset that, when estimated, produces substantially less bias. Section 4 uses the study of alliance formation by Gibler and Wolford (2006) to illustrate how one may apply choice-based sampling to the construction of k-adic

[3] This is a problem that the governing coalition papers of Martin and Stevenson (2001) and Franklin and Mackie (1984) did not have to address as the number of potential actors in their studies were relatively small (the largest being N=10 for a few countries where up to 10 parties existed at the time of a government formation).

data. Section 5 offers a discussion of why some alternative methods, particularly spatial interdependence models and evolving network models, are not viable substitutes for using k-adic data to evaluate multilateral events. Section 6 concludes.

2.2 Illustrating the Problem with Monte Carlo Simulations

In this section, I use Monte Carlo simulations to illustrate how dyadic data cannot capture the process that produces data formed by interactions among $k > 2$ actors. Again, this is an issue of how one conceptualizes the construction of the observations and, hence, is separate from concerns of modeling spatial, temporal, strategic, or monadic interdependencies among the observations. For the sake of illustration, I place this simulation (and the subsequent simulations) in the context of alliance formation. This is intended only to give the simulation a substantive point of reference.

2.2.1 Motivating the Simulation

According to Morrow (1991), pairings of states with highly asymmetric relative physical capabilities are natural alliance partners. In essence, alliances serve as a type of "protection racket" where a small state gives foreign policy autonomy to a larger state (in the form of policy concessions or the granting of territorial access to the large state's military forces) in exchange for the large state's promise to defend it in a time of crisis. One can easily extend to multilateral agreements the Morrow story of asymmetry influencing alliance formation. The game theoretic work on N-player prisoner dilemmas (see Bianco and Bates 1990) and its extension to international cooperation (see, most recently, Stone, Slantchev, and Tamar 2008) view the presence of a large state as the key factor in creating multilateral agreements. This is because the large state can impose punishments on states that fail to meet contribution requirements.

This suggests that the capability ratio of a grouping of states is a (if not "the") major factor in determining if the states will form an alliance. Though the exact influence of the largest state's capabilities relative to the entire group's capabilities is not known, we do know that, in theory, the larger this ratio, the more likely is a multilateral agreement. For the sake of simplicity and to avoid the issue of the improper use of control variables that is rampant throughout the empirical international relations literature,[4] I will assume that the true data generating process for alliance formation can be specified as:

$$\begin{aligned} &\Pr(\text{Alliance between states A through K}) = \\ &\Phi\left(\text{cons} + \beta\frac{\max(\text{capability A},\text{capability B},\cdots,\text{capability K})}{(\text{capability A}+\text{capability B}+\cdots\text{capability K})} + \mu\right) \end{aligned} \tag{2.1}$$

where **cons** is a constant term, μ is a random element capturing the unknown and/or unobserved determinants of alliance formation, and Φ is a function taking on values strictly between zero and 1 ($0 \leq \Phi \leq 1$). β is the true parameter specifying the relationship between the value of a latent, unobserved dependent variable that determines the probability of alliance formation, y^*, and the capability ratio of states A through K.

2.2.2 Describing the Trilateral Alliance Simulation

I consider a scenario in which states can only form the most basic of multilateral alliances, *trilateral* alliances. I construct the simulation according to the following steps:

[4]For works that detail the improper use of control variables in international relations, see Kadera and Mitchell (2005), Ray (2005), Achen (2005), Clarke (2005) and Starr (2005).

STEP 1: I create a dataset of 100 observations, where each observation represents a single country. I then assign a country code (ccode) value to each country.

STEP 2: I randomly assign military "capabilities" to each of these countries. Capabilities range from 0 to 100. These capabilities are stored in the variable *cap*.

STEP 3: I reorganize the 100 countries into all possible three-country groupings. Since order is not important, these 100 countries produce 161,700 three country combinations of states (or triads). The triadic dataset includes the following variables: $triadid$, $mem1$, $mem2$, $mem3$, $cap1$, $cap2$, $cap3$. The variable $triadid$ is simply a code identifying triad i (with $i \in \{1,\ 161700\}$). The variable *mem1* is the ccode number of the first member state in triad i, *mem2* is the ccode number of the second state in triad i, and *mem3* is the ccode number of the third state in triad i. The variables *cap1*, *cap2*, and *cap3* capture the capabilities of *mem1*, *mem2*, and *mem3* respectively.

STEP 4: I compute the "capabilities ratio" of each dyad. Specifically, this is captured by the variable **cap ratio** which is calculated as:

$$\text{cap ratio} = \frac{\max(\text{cap1, cap2, cap3})}{(\text{cap1} + \text{cap2} + \text{cap3})} \tag{2.2}$$

STEP 5: I write the data generating process (DGP) of trilateral alliance formation as:

$$xb = \text{cons} + \beta \text{cap ratio} + \mu \tag{2.3}$$

where xb represents the underlying latent variables that determine alliance formation. I set cons $= -4$ and $\beta = .25$. The variable μ is a logistically distributed random error term.

STEP 6: $ALLY$, the dependent variable, is a dichotomous variable coded 1 if a triad forms an alliance, zero otherwise. To generate realizations of this dependent variable, I code $ALLY = 1$ if $xb > 0$, zero otherwise. Table 2.1 reports the values of these variables for the first 10 observations. One should notice that $ALLY = 0$ for each of these 10 observations.

Step 7: I now convert this triadic dataset into dyadic data. Thus, if a triad contains states A, B, and C, this step divides this triad into dyad A with B, dyad B with C, and dyad A with C. If $ALLY = 1$ for triad A,B,C, then this means $ALLY = 1$ for dyad A with B, $ALLY = 1$ for dyad B with C, and $ALLY = 1$ for dyad A with C. Next, I use the capabilities scores of each dyad member to compute that dyad's capability ratio.

Step 8: I take this dyadic dataset and attempt to estimate β, the parameter characterizing the relationship between **cap ratio** and $ALLY$. Since the errors are drawn from a logistic distribution, I use logit estimation. The goal is to see if the logit estimate of β, $\hat{\beta}$, is an unbiased estimate of the true β (which is equal to 0.25).

These eight steps create one realization of my dataset. Of course, this realization is determined by a single random draw of u from a logistic distribution. Because the dependent variable is computed using an error term drawn from a probability distribution, I should repeat the creation of the dependent variable via a Monte Carlo simulation. In Monte Carlo simulations, random numbers are drawn so as to model a process. The goal is to determine how random variation (or lack of knowledge or error) affects the sensitivity and reliability of the parameters characterizing the process. In this particular simulation, I wish to know how randomness impacts my ability to estimate the impact of relative capabilities on the formation of alliances. The essence of Monte Carlo simulations to iterate the process numerous times and then obtain an average value from these iterations.

Step 9: I repeat 500 times steps six through eight. This produces 500 values of $\hat{\beta}$. After each iteration, I also keep the estimated standard error around $\hat{\beta}$ (giving me 500 values of the standard error of $\hat{\beta}$).

2.2.3 Results from the Trilateral Alliance Simulation

I use the stored values of $\hat{\beta}$ and the estimated standard errors to compute three common criteria for evaluating estimator performance in Monte Carlo simulations: Bias, Root Mean Squared Error, and Overconfidence. Bias is the difference between the average value of the coefficient estimate and the true coefficient value. Root Mean Squared Error is calculated in three steps: (1) computing the squared difference between each iteration's coefficient estimate and the true coefficient value; (2) summing up these values and dividing the total by the number of iterations; and (3) taking the square root of this average value. Overconfidence is the standard deviation of the coefficient estimates divided by the average reported coefficient standard error. This is a measure of standard error accuracy. For all three measures, the smaller the value, the more accurate the estimator.

Column 1 of Table 2.2 shows how estimating the triadic dataset with triadic data produces, as one would expect, relatively unbiased coefficient estimates. However, column 2 shows quite convincingly that the $\hat{\beta}$ produced using the dyadic dataset does not accurately estimate β. These results suggest that the existing approach of dividing multilateral alliances into a series of dyadic observations produces biased estimates of the true parameter. This is unsurprising, as one should not reasonably expect a *dyadic* measure of capability asymmetry to be equivalent to the *triadic* measure of capability asymmetry employed in this simulation's actual DGP. Unfortunately, this is exactly the technique employed by scholars of international relations (to divide multilateral events, in this case the formation of alliances, into a series of bilateral observations).

2.2.4 Indentifying the Source of the Bias

What is the exact cause of the bias in the multilateral simulation? Setting aside problems of non-independence of observations (which, as already highlighted, is present in dyadic data as well), dividing a k-adic event into dyadic observations leads to a classic case of measurement error in $\mathbf{X}$, the vector of independent variable values. Recall that the capability ratio in the true DGP is

$$\text{cap ratio} = \frac{\max(\text{cap1, cap2, cap3})}{(\text{cap1} + \text{cap2} + \text{cap3})} \tag{2.4}$$

while the independent variable in the estimated model is the dyadic capability ratio

$$\text{cap ratio} = \frac{\max(\text{cap1, cap2})}{(\text{cap1} + \text{cap2})} \tag{2.5}$$

Suppose that the third member of the alliance is never the largest member. Therefore, the dyadic data will always have the correct numerator value (either *cap1* or *cap2*). However, even in this ideal scenario, the estimated independent variable is systematically higher because the denominator is missing one term, *cap3*. As a result, the observation $\mathbf{X}_i$ for $i = 1, \cdots, n$ is actually $\mathbf{X}_i = \mathbf{W}_i + \mathbf{U}_i$, where $U_1, ..., U_n$ are uniformly distributed (because a uniform distribution was used to generate the capability scores of each state) with $E[\mathbf{U}] \geq 0$.[5] Thus, for either the probit or logit

[5]It should be noted that $U_1, ..., U_n$ are not independent as some observations contain the same third country.

model, we obtain (ignoring the constant term)

$$\begin{aligned}
\Pr(Y_{T_i} = 1) &= \Pr(Y^*_{T_i} \geq 0) \\
&= \Pr(\beta \mathbf{X}_i + \epsilon_i \geq 0) \\
&= \Pr(\beta(\mathbf{W}_i + \mathbf{U}_i) + \epsilon_i \geq 0) \\
&= \Pr(\epsilon_i \geq -\beta(\mathbf{W}_i + \mathbf{U}_i)) \\
&= 1 - F(-\beta(\mathbf{W}_i + \mathbf{U}_i)) \\
&= F(\beta(\mathbf{W}_i + \mathbf{U}_i))
\end{aligned}$$

where the last line is possible if F is symmetric (which is the case for both the logistic and normal distributions). Of course, the parameter will be estimated via Maximum Likelihood, where each outcome of Y_{T_i} follows a Bernoulli density function, $f(Y_{T_i}) = p_i^{Y_{T_i}}(1-p_i)^{1-Y_{T_i}}$. In other words, each Y_{T_i} takes on either a value of 0 or 1 with probability $f(0) = (1-p_i)$ and $f(1) = p_i$. Hence, $p_i = \Pr(Y_{T_i} = 1) = F(\beta(\mathbf{W}_i + \mathbf{U}_i))$. Thus, the likelihood function is $L = f(Y_{T_1}, Y_{T_2}, \cdots, Y_{T_n})$. Even if, for $i \neq j$, each Y_{T_i} were independent of each Y_{T_j} (which is not the case since multiple dyads contain the same country) so that the log likelihood is $\ln L = \sum_{i=1}^{n} Y_{T_i} \ln p_i + (1 - Y_{T_i}) \ln(1 - p_i)$, it would still be the case that $p_i = F(\beta(\mathbf{W}_i + \mathbf{U}_i))$. Consequently,

$$\ln L = \sum_{i=1}^{n} Y_{T_i} \ln F(\beta(\mathbf{W}_i + \mathbf{U}_i)) + (1 - Y_{T_i}) \ln(1 - F(\beta(\mathbf{W}_i + \mathbf{U}_i))) \qquad (2.6)$$

One will note that the presence of $\mathbf{U}_i$, where $E[\mathbf{U}_i] \geq 0$ will inflate the value of $\mathbf{X}_i$. Because the values of Y_{T_i} are fixed, the larger value of $\mathbf{X}$ must necessarily reduce $\hat{\beta}$, the estimate of the true β. Thus, $\hat{\beta} \neq \beta$.

2.3 Modeling *K*-adic Data Using Choice-Based Sampling

Scholars have relied upon dyadic data to analyze international events because it provides intuitive tractability, is computational convenient, and simplifies the collection of data. Moreover, scholars have made great strides in devising estimation adjustments that account for temporal, spatial, monadic, and strategic violations of the independence assumption present in many dyadic datasets. However, none of these adjustments, nor any estimation correction, can account for the bias produced by evaluating k-adic events with dyadic data.

The simulation in the previous section show that estimating the formation of *trilateral* alliances with *triadic* data will produce unbiased estimates of the parameter on **cap ratio**. Therefore, the solution seems obvious: estimate a dataset with all possible k-ads. Fortunately, by all "possible" k-ads, I do not mean to suggest that if there are n countries then one needs to include all k-ads of size n or less. Instead, I will show that if the k-adic event of interest contains, at most, $k < n$ countries, one need only estimate a dataset with all combinations of states up to size k. Unfortunately, creating a dataset of all combinations of size $k < n$ has a major downside: it still dramatically expands the dataset's observations. For a system of 100 countries, just as in the above simulations, a dyad only dataset contains 4,950 observations, but a triad only dataset contains 161,700 observations. If one were to consider a dataset of four country alliances, the dataset size would explode to 3,921,225 observations. Consequently, it would be infeasible or impossible to estimate a dataset capable of explaining the creation of an alliance the size of NATO, which was formed by 12 countries (1.0504210511067e+15 observations)!

2.3.1 Choice-based Sample of Triadic Data

Choice-based sampling on the dependent variable (see King and Zeng 2001a, 2001b) may offer a means of creating a computationally manageable dataset appropriate for estimating k-adic data. Because so few triad observations, relative to the total number of triads, contain the formation of a military alliance, one is left with a classic “rare events” dataset (binary dependent variable characterized by dozens to thousands of times fewer events [coded with a positive value] than non-events [coded with a 0]). When presented with data of this type, King and Zeng (2001a, 2001b) recommend sampling on the dependent variable as it avoids the issues commonly associated with rare events data such as underestimating the probability of an event. The sampling method entails constructing a dataset containing all observations for which the dependent variable is coded with a positive value, along with a random sample of observations for which the dependent variable is coded 0. According to King and Zeng (2001a, p. 702), it is acceptable to collect anywhere from two to five times more 0’s than positive values, though one should attempt to collect as many zero values as is computationally feasible. Thus, in the simulations that follow, I collect ten times more 0’s than ones. Even then, this creates a dataset between two and three thousand observations, which is highly manageable from a computational perspective.

When drawing the sample, it is important to stratify by k-ad. This means if a dataset has 100 dyads where Y=1 and 50 triads where Y=1, one should attempt to draw 1000 dyads where Y=0 and 500 triads where Y=0. Why should one sample in this manner rather than simply draw a nonstratified random sample from the full population of possible outcomes? The reason is that higher order k-ads quickly overwhelm lower order k-ads with respect to quantity. For instance, if a dataset contains all five-ads, quad-ads, tri-ads, and dyads of 100 actors, then the full population contains 75,287,520 five-ads and only 4,950 dyads. A non-stratified random sample of the Y=0 k-ads would contain virtually no Y=0 dyads, even if dyads comprise the majority of Y=1 observations!

Having obtained a stratified choice-based random sample of k-ads where the dependent variable equals zero, this sample can be combined with the k-ads where the dependent variable equals 1. This combined dataset can then be estimated using a *rare events logit model*, which, by and large, is a logit model that applies a post-estimation correction to the constant term (called prior correction) to account for the fact that sampling on the dependent variable has artificially inflated the prominence of observations where the dependent variable equals 1. Since the dataset on which the model is estimated is a stratified sample, one must weight the observations from each strata by the inverse probability of being drawn from the sample. For example, if there are 4,950 total dyads, and Y=1 for 100 of these dyads, then each Y=0 dyad has a $\frac{1}{4{,}850}$ probability of being drawn. When estimating the model, each Y=0 observation in the sample should be multiplied by $\frac{1}{\frac{1}{4{,}850}}$.

After applying choice-based sampling, I have a cross-sectional dataset of approximately 55,000 triad observations (where approximately 5,000 of which *Alliance Formation* is coded 1). The third column of Table 2.2 presents the results obtained from estimating a choice-based sample of the triadic dataset using a King and Zeng (2001a, 2001b) rare-events logit model. Comparing this parameter estimate to that obtained from estimating the full triadic dataset, one can see that the parameter estimates are nearly identical, with the Overconfidence measure suggesting that the rare-events estimate produces slightly more variance (which is expected, given that it contains fewer observations).

2.3.2 Accounting for All K-ads with Choice-Based Sampling: Proof of Concept

Though this solution works for the triadic dataset, what about a dataset in which the largest k-lateral alliance contains 4 countries or a dataset containing k-lateral alliances of multiple sizes? The latter is of particular importance since this is the shape of actual datasets in international relations. For example, the

Alliance Treaty Obligations and Provisions dataset contains 648 military alliance treaties formed between 1815 and 2005. Of these, 536 are bilateral alliances, 47 are trilateral alliances, 23 are quadrilateral alliances, 11 alliances have five members, and 38 have 6 or more members (with the largest alliance containing 50 members).[6] As mentioned above, estimating a dataset with all possible combinations of all possible alliance sizes is computationally infeasible. For instance, a dataset with all possible dyadic, triadic, and quadratic combinations of 100 countries contains $4,950 + 161,700 + 3,921,225 = 4,087,775$ observations! However, one can still sample on the dependent variable in order to obtain parameter estimates for such data. To show this is the case, this subsection provides a "proof of concept" focusing on a simulated dataset containing bilateral and trilateral alliances.

Bilateral-Trilateral Simulation

For this simulation, I place the states into all possible dyadic and triadic combinations. This generates a combined dataset of 166,550 observations (where an observation is any grouping of states, dyadic or triadic). The variable **cap ratio** is the ratio of the capabilities of a k-ad's largest member over that k-ad's total capabilities. Next, I set β_1, the parameter on **cap ratio**, to 25 and the constant term is set to -25. With these parameter values, a typical simulation produces approximately equal numbers of bilateral and trilateral alliances (typically 120 to 130 each). As in the above simulations, the DGP also includes a logistically distributed error term.

Table 2.3 reports the results from 500 Monte Carlo simulations of this DGP. One can see from column one that applying logit estimation to the full triad-dyad dataset produces, on average, parameter estimates close to the true parameter value. The second column reports the average parameter estimates from 500 Monte Carlo

[6] See Leeds (2002) for further details on the ATOP dataset.

simulations where the data is converted back into dyadic data. As in the above simulation with triadic data, this produces large bias in the parameter estimate.

In the third column, I present the results from a "quick fix" one might be tempted to apply when faced with a multilateral event: dropping the multilateral events. Given that dyads, by definition, capture bilateral relations, some readers may decide a simpler solution for obtaining unbiased estimates lies in simply excluding multilateral events from the data. Therefore, I rerun the simulation with the triads removed from the dataset prior to estimation.[7] This does produce dramatically less bias in the estimates and, therefore, is not an unreasonable approach. However, if one wishes to model multilateral observations, dropping the multilateral cases is obviously not an option. Moreover, as the next simulation will show, the bias produced by this approach increases with the number and variety of k-ads in the dataset.

The fourth column reports the results from estimation with choice-based sampling on the dependent variable. The bias is substantially reduced compared to converting the data into dyadic data. Therefore, choice-based sampling appears to offer a viable and computationally feasible means of obtaining relatively unbiased parameter estimates of actual k-adic data.

A FIVE-adic Simulation

To more fully illustrate the ability of choice-based sampling to create a feasibly sized k-adic dataset that reduces estimation bias, I consider a final simulation in which the maximum size of a k-ad is five countries. I focus on a FIVE-adic dataset because, as the above description of the ATOP dataset illustrates, there are very few k-adic alliances with more than five participants.

For this simulation, I place the countries into all possible combinations of 2, 3, 4,

[7] This approach is adopted in some studies, such as Kremmer (1998), with the explicit desire to avoid placing multilateral events into a dyadic data (Remmer, 1998: 35).

and 5 countries. This generates a combined dataset of 79,375,395 observations (4,950 + 161,700 + 3,921,225 + 75,287,520). The variable **cap ratio** is defined the same as in the earlier simulations and the DGP is the same as in the above simulations, except **cap ratio** is computed with the capabilities of two, three, four, or five states, depending on whether an observation is a dyad, triad, "4"-ad, or "5"-ad.

Next, I set β_1, the parameter on **cap ratio**, to 26 and the constant term is set to -27. With these parameter values, a typical simulation produces about 190 bilateral alliances, 180 trilateral alliances, 110 quadrilateral alliances, and 70 five member alliances. Table 2.4 reports the results from estimating this dataset by applying choice-based sampling to the full dataset (column 1), dividing the data into dyadic combinations (column 2), or dropping the k-adic observations (column 3). One can immediately see that the bias produced by analyzing the 5-adic DGP with dyadic data is dramatically more pronounced than in the previous simulation. Moreover, estimation with the choice-based sample outperforms both estimation with a dataset in which all k-adic alliances are split into their dyadic combinations and estimation with a dataset in which the k-adic observations are simply dropped.

2.4 Application: Alliance Formation in International Relations

I will now illustrate how one can apply choice-based sampling to actual data that follows a k-adic DGP. I will do so using the study of alliance formation by Gibler and Wolford (2006), who draw on the dyad-year research design of Lai and Reiter (2000). This is one of the only studies that conducts multi-variate estimation of alliance formation. The dependent variable *alliance formation* is coded 1 the year two states become alliance members, zero otherwise. Because this study is particularly interested in the relationship between regime type and alliance formation, Gibler and Wolford (2006) mention how dyadic data could overstate the role of democracy on alliance formation. In particular, Gibler and Wolford (2006: 139) highlight how the bulk of democratic dyads that formed alliances are contained in an incredibly small number of alliances. For example, NATO accounts for more than 55 percent of the jointly democratic allied dyad-years. Nevertheless, Gibler and Wolford (2006), in order to match as closely as possible previous research designs, test their model using all dyad-year data drawn from all alliances.

I am not setting out to show that the results of Gibler and Wolford (2006) fail to hold. Instead, I am using their study because, as with the above simulated examples, alliance formation is a useful example of a k-adic DGP that has typically been tested using dyadic data. Fully replicating and critiquing their study, or any study that splits into dyads a k-adic process, will require scholars to develop new measurements of covariates that were previously coded only in the dyadic context. For example, *geographic distance* is easy to conceptualize for a dyad, but what does it mean in a k-adic dataset? Is it the maximum distance between any two of k members or is it the average distance between the k members? Similarly, what is a joint democracy k-ad? If a k-ad contains five states, is it a joint democracy k-ad only when all five states are democracies? If so, does that not treat a k-ad where 4 of the 5 states are democracies as equivalent to a k-ad where 1 of 5 states are democracies? Alternatively, perhaps one could construct a continuous measure of joint democracy such as the proportion of states in the k-ad that are democracies. My goal is not to rectify such measurement issues, as answers will depend on the particular research question. Instead, by illustrating how one can properly construct and test a dataset for an event that is inherently k-adic, I will propose and apply reasonable codings for such variables.

2.4.1 Real Alliance Formation and Capabilities Data

Before more fully applying the Gibler and Wolford (2006) model of alliance formation, I begin with a simple model that closely follows the above simulations. Specifically, I test a single covariate model where the dependent variable is the formation of a Correlates of War military alliance (similar to Gibler and Wolford 2006) and the independent variable is the capability ratio (the capabilities of the largest state over the sum of the k-ad's capabilities) where capabilities is measured using the Correlates of War composite index of national capabilities (CINC) score. The dependent variable *ally* is coded 1 the year an alliance forms, zero otherwise. Because this replication focuses on the decision to form a new alliance, I consider states who join an alliance after the year of its initial formation as having not joined the alliance (the decision to join an existing alliance is a worthy research question, but is treated here as distinct from the decision to create a new alliance).

The column (1) of Table 2.5 reports the results from applying logit estimation with clustered standard errors to a dyadic dataset of alliance formation. The column (2) of Table 2.5 reports the results when using a rare events logit to estimate a choice-based dataset which directly measures all k-ads that formed alliances. It is important to make two notes regarding the results in column (3). First, I use eight times more *ally*=0 observations than *ally*=1 observations. Second, the estimation does not include k-ads of size 6 or greater. This is for two reasons. First, the dataset only contains four k-ads of such size. Hence, if alliance formation is a "rare event", then the formation of alliances with six or more members is an "unusual" event. Second, the set of possible *ally*=0 k-ads of 6 or more members is simply enormous. For example, with 196 countries (the number of countries in the Gibler and Wolford (2006) dataset), all combinations of 6 are equal to 72,887,293,024.

Since there is only a single six member alliance, I would draw only ten 6-ads where *ally*=0. This is problematic because when using a stratified choice-based sample, one must weight each observation by the inverse probability of that observation being drawn from its stratum. In the case of 6-ads where *ally*=0 (where the *ally*=0 6-ad stratum contains $72,887,293,024 - 1$ observations), this produces a probability of $\frac{10}{72,887,293,023} = 0.0000000014$ or an inverse probability weight of 7,288,729,302. Placing such a massive weight on a single observation renders the standard errors of the point estimates uninformative. Since some datasets will have more than 196 actors and some will have less, the decision of what constitutes an "unusual" versus simply a "rare" k-adic event must be left to the analyst.

Table 5 shows that splitting all k-ads into dyadic relations (column 1) leads to a negative and statistically significant value on the coefficient for capability ratio. In contrast, the coefficient is positive when one accounts for all k-adic combinations.

2.4.2 Reconsidering Gibler and Wolford (2006)

Gibler and Wolford (2006), drawing upon Lai and Reiter (2000), model alliance formation as a function of several variables. For the purposes of my illustrative application, I will only use a subset of the variables they consider: *Common Threat*, *Geographic Distance*, and *Joint Democracy*. I will include with these variables the *capability ratio* of each k-ad. These variables are chosen because they are (1) consistently found to be important determinants of alliance formation and (2) are examples of variables problematic to code in k-adic data.

In dyadic data, *Common Threat* is a dichotomous variable coded 1 if both states participated in a Militarized Interstate Dispute against the same third state sometime in the previous 10 years, 0 otherwise. Coding this variable in the k-adic context creates similar difficulties to coding *joint democracy*: if a k-ad contains five states, does it not face a joint threat if only 4 of the 5 states participated in a MID against the same third state? Given that the variable *Common Threat* is intended to capture the idea that a group of states will have a strong incentive to form an alliance when *all* members of that group face the same threat, I will adopt such a coding rule: *Common Threat* is coded 1 if each state has participated in a MID against the same third state sometime in the previous 10 years, 0 otherwise.

In dyadic data, *Geographic Distance* gives the square root of the capitol to capitol distance, unless states are contiguous, in which case distance is set to 0. The potential complications with coding *Geographic Distance* with k-adic data were discussed above. I will code *Geographic Distance* in k-adic data by applying the "weakest link" principle of Oneal and Russett (1997).[8] This means I will represent the geographic distance of the entire k-ad using the geographic distance of the most distant pair of states.

In dyadic data, *Joint Democracy* is a dichotomous variable coded 1 when both members of the dyad are democracies, 0 otherwise. As mentioned above, this coding rule is a bit problematic when applied to k-adic data. Should one only consider a FIVE-ad where four of five states are democracies, to be equivalent to a Five-ad where only one of the five states is a democracy? I suggested above that perhaps one could use a continuous measure of democracy, such as the proportion of states in the k-ad that are democracies. Therefore, I code *Joint Democracy* using two approaches: as the proportion of states in a k-ad that are democracies and as a dichotomous variable coded 1 when all members of the k-ad are democracies, 0 otherwise. This will allow me to compare how the results are changed by using an alternative coding

[8]Oneal and Russett (1997) use as a measure for the entire dyad the minimum state level value for the dyad. For example, they measure a dyad's overall level of trade integration by using the lower of the two state levels of trade integration (if state A has trade integration of 40 percent and state B has trade integration of 30 percent, then the trade integration for the dyad is 30 percent).

rule. Descriptive statistics for these variables are reported in Table 2.6, along with a comparison to the typical dyadic values of these variables.

The results are reported in Table 2.7. Comparing column (3) to column (4), one can see that using a continuous or dichotomous measure for *Joint Democracy* does not drastically alter the coefficient estimates. Comparing column (1) of Table 2.7 (splitting all k-ads into dyadic relations) to column (3) of Table 2.7 (estimation with choice-based sampling of k-ads) reveals two major changes in the results. First, the sign on the variable for *capability ratio* flips from negative to positive. While the model estimated with dyadic data identifies *capability ratio* as having a significant and negative effect on the probability of alliance formation, the model estimated with k-adic data finds that the effect is positive and insignificant. Second, the coefficients on the remaining variables are dramatically larger in the k-adic model. To illustrate the substantive impact of these larger coefficients, consider a change in the relative risk associated with going from having no common threat (*common threat*=0 in both the dyadic and k-adic models) to having a common threat (*common threat*=1 in both the dyadic and k-adic models).[9] Estimating this model with dyadic data shows that such a change increases the risk of forming an alliance by 5.02 times. However, estimating this model with k-adic data shows that such a change increases the risk of forming an alliance by 420 times.[10]

Upon seeing these results, some scholars may wonder if a simpler approach for modeling a k-adic process without bias would be to simply incorporate into dyadic data a dummy variable that accounts for the k-adic concept. For instance, consider again the Belgium-Turkey example that opened the paper. Given that both

[9] Bennett and Stam (2007) suggest using the risk ratio to substantively evaluate logit coefficients as the rare occurrence of many international events render their predicted probabilities to be exceedingly small (Bennett and Stam, 2007: 67-69).

[10] The probabilities and relative risk ratios are computed using prior correction by applying the *relogitq* command in STATA. Replication do files are available upon request.

joined NATO because of the presence of the United States in the alliance, could one not simply add a variable for "Alliance Formation with the US" or even "Alliance Formation with a Super Power"?

Depending on the research question, such a reasonable "quick fix" may be appropriate (i.e., if the scholar is studying the influence of the United States in the formation of alliances). However, it is important to note that not all multilateral alliances include a major power. Additionally, a dummy variable does not capture the reason *why* the presence of a major power leads to the formation of an alliance. Is it because the major power poses a threat, offers security, or creates the "correct" balance in the capability ratio? This is not made clear by the simple inclusion of a dummy variable.

2.5 Alternative Approaches and Their Limitations

Though the inclusion of a dummy variable will not address the misconceptualization of multilateral events as fitting a dyadic data generating process, scholars may still wish to model a k-adic process using dyadic data. Therefore, it is worth discussing some approaches that attempt to retain dyadic data and why these approaches, though quite useful in other contexts, are not yet suitable for modeling k-adic processes: bilinear mixed-effects Hierarchical models, spatial interdependence regression models, and evolving network models.

2.5.1 Bilinear Mixed-Effects Hierarchical Model

Ward, Siverson, and Cao (2007) and Hoff and Ward (2004) use the Bilinear Mixed-Effects model developed by Hoff (2005) to address monadic dependency in dyadic data. In essence, this model enables scholars to overcome a problem that is the mirror image of the issue I raise: standard approaches to analyzing non-directed dyadic data (i.e., movement from state i to state j is considered the same as from j to i) hold that the dependence of observations having a common sender and the dependence of observations having a common receiver are both zero. This is a problem because, it seems unreasonable to assume that all dyads containing the United States are independent from one another. The bilinear mixed-effects model can account for this country specific dependency by explicitly incorporating both dyadic and monadic (country specific) characteristics into the regression model.

Formally, suppose there is a binary outcome, $y_{i,j}$, which is either 0 or 1, indicating the presence or absence of a "link" from i to j.[11] Suppose we are interested only in estimating the linear relationships between responses $y_{i,j}$ and a vector of variables $\mathbf{x}_{i,j}$, which could include characteristics of unit i, characteristics of unit j, or characteristics specific to the pair. Thus, we can consider the regression model

$$y_{i,j} = \beta'\mathbf{x}_{i,j} + \epsilon_{i,j} \tag{2.7}$$

The generalized least squares estimate of $\hat{\beta}$ and its covariance matrix depend on the joint distribution of the $\epsilon_{i,j}$'s only through their covariance. Next, two key assumptions are made.

Key Assumption 1: Invariance of $\epsilon_{i,j}$ It is commonly assumed in regression problems that the regressors $\mathbf{x}_{i,j}$ contain enough information so that the distribution of the errors is invariant under any combination/arrangement of i and j. This is known as "weak row-and-column exchangeability" of an array.

Key Assumption 2: $\epsilon_{i,j}$ is Gaussian with mean 0 For undirected dyadic data (in which $y_{i,j} = y_{j,i}$), the first assumption implies that $\epsilon_{i,j}$ is equal in distribution to

[11] The formal discussion is adopted from Hoff, 2005, pp. 286 - 287

$f(u, \alpha_i, \alpha_j, \gamma_{i,j})$, where $u, \alpha_i, \alpha_j, \gamma_{i,j}$ are independent random variables and f is a function to be specified. When combined with the second assumption, we can now express $\epsilon_{i,j}$ as

$$\epsilon_{i,j} = \alpha_i + \alpha_j + \gamma_{i,j} \tag{2.8}$$

where (α_i, α_j) is distributed multivariate normal with mean zero and variance $\Sigma_{\alpha_i,\alpha_j}$ and $(\gamma_{i,j}, \gamma_{j,i})$ is distributed multivariate normal with mean zero and variance $\Sigma_{\gamma_{i,j},\gamma_{j,i}}$. Because $\alpha_i, \alpha_j, \gamma_{i,j}$ are independent random variables, then

$$\Sigma_{\alpha_i,\alpha_j} = \begin{pmatrix} \sigma^2_{\alpha_i} & 0 \\ 0 & \sigma^2_{\alpha_j} \end{pmatrix} \tag{2.9}$$

and

$$\Sigma_{\gamma_i,\gamma_j} = \begin{pmatrix} \sigma^2_{\gamma_{i,j}} & 0 \\ 0 & \sigma^2_{\gamma_{i,j}} \end{pmatrix} \tag{2.10}$$

This means the covariance structure of the errors (and thus the observations) is

$$\begin{aligned} E(\epsilon^2_{i,j}) &= \sigma^2_{\alpha_i} + \sigma^2_{\alpha_j} + \gamma^2_{i,j} \\ E(\epsilon_{i,j}, \epsilon_{j,i}) &= 0 \\ E(\epsilon_{i,j}, \epsilon_{i,k}) &= \sigma^2_{\alpha_i} \\ E(\epsilon_{i,j}, \epsilon_{k,j}) &= \sigma^2_{\alpha_j} \\ E(\epsilon_{i,j}, \epsilon_{k,i}) &= 0 \\ E(\epsilon_{i,j}, \epsilon_{k,l}) &= 0 \end{aligned} \tag{2.11}$$

so that $\sigma^2_{\alpha_i}$ represents the dependence of observations having a common sender and $\sigma^2_{\alpha_j}$ represents the dependence of observations having a common receiver. Standard approaches to analyzing dyadic data assume that both of these values are zero, but this is theoretically unlikely (i.e., it seems unreasonable to assume that all dyads containing the United States are independent from one another). Thus, the Ward *et al.* (2007) approach provides a better way of analyzing *dyadic* data that accounts for the characteristics of the individual members of the dyad. This is useful when the dyad is the appropriate unit of observation and could be extended to account for individual level characteristics in k-adic data, but it does not allow one to avoid using the k-ad as the unit of observation.

2.5.2 Spatial Interdependence Regression Model

The spatial interdependence specification of Franzese and Hays (2007a, 2007b, 2007c, and 2008) is applied whenever an outcome in country i is influenced by the connection country i has with country j. For example, spatial interdependence regression models have been widely applied in political science to the study of interstate capital tax rate competition, more pejoratively referred to as the "race to the bottom." Specifically, a common practice in statistical models of international capital tax competition is to control for the influence that tax rates of countries neighboring state i have on the tax rate of state i. Unfortunately, Franzese and Hays (2006) point out how previous studies on capital tax competition, such as Hays (2003) and Basinger and Hallerberg (2004), improperly specify such spatial interdependence. For instance, Hays (2003) fails to weight the importance of the tax rate of one country relative to another.

Formally, if N is the number of countries and T is the number of time periods, then the connection between country i and country j is typically captured by $\mathbf{W}$, an $NT \times NT$ block-diagonal matrix where element w_{ij} reflects the degree of connection from i to j according to some metric (such as the similarity or complimentarity between i's and j's economies or their trade bundles). For instance, a standard $\mathbf{W}$

matrix is expressed as

$$\left[\begin{array}{c:c:c}
\begin{matrix} w_{11_1} & w_{12_1} & w_{13_1} & \cdots & w_{1n_1} \\ w_{21_1} & w_{22_1} & w_{23_1} & \cdots & w_{2n_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{n1_1} & w_{n2_1} & w_{n3_1} & \cdots & w_{nn_1} \end{matrix} & 0 & 0 \\
\hdashline
0 & \ddots & 0 \\
\hdashline
0 & 0 & \begin{matrix} w_{11_T} & w_{12_T} & w_{13_T} & \cdots & w_{1n_T} \\ w_{21_T} & w_{22_T} & w_{23_T} & \cdots & w_{2n_T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{n1_T} & w_{n2_T} & w_{n3_T} & \cdots & w_{nn_T} \end{matrix}
\end{array}\right] \tag{2.12}$$

Suppose "connection" in this instance is a shared border. Therefore, if country i and country j share a border then $w_{ij} = 1$. Otherwise $w_{ij} = 0$. Hence, this block-diagonal matrix can be rewritten as

$$\left[\begin{array}{c:c:c}
\begin{matrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 1 & \cdots 0 & \end{matrix} & 0 & 0 \\
\hdashline
0 & \ddots & 0 \\
\hdashline
0 & 0 & \begin{matrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 1 & \cdots 0 & \end{matrix}
\end{array}\right] \tag{2.13}$$

where the diagonal in each matrix is zero (as it contains "self-referencing entries;

e.g. w_{11} contains the impact of country 1 on country 1). Having constructed this spatial-weighting matrix, the impact on policy y of country i by country j is captured with

$$\mathbf{y} = \rho \cdot \mathbf{W}\mathbf{y} \tag{2.14}$$

where $\mathbf{y}$ is an $NT \times 1$ vector of outcome observations stacked by time (i.e. time 1, country 1 to N, then time 2, country 1 to N, through time T). Combined, $\mathbf{Wy}$ reduces to a vector, where the parameter ρ captures the impact the spatially-weighted outcome of countries $-i$ has on the outcome of country i.

Returning to the example of capital tax rate competition, the spatial interdependence variable, $\mathbf{Wy}$, can be written as

$$\mathbf{Wy} = \begin{bmatrix} w_{1,1\ldots17}1965 & w_{1,1\ldots17}1966 & \cdots & w_{1,1\ldots17}1995 & w_{1,1\ldots17}1996 \\ w_{2,1\ldots17}1965 & w_{2,1\ldots17}1966 & \cdots & w_{2,1\ldots17}1995 & w_{2,1\ldots17}1996 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ w_{16,1\ldots17}1965 & w_{16,1\ldots17}1966 & \cdots & w_{16,1\ldots17}1995 & w_{16,1\ldots17}1996 \\ w_{17,1\ldots17}1965 & w_{17,1\ldots17}1966 & \cdots & w_{17,1\ldots17}1995 & w_{17,1\ldots17}1996 \end{bmatrix} \tag{2.15}$$

where $w_{1,1\ldots17}1965 = (0 \times 1965 \text{ tax rate}_1) + (w_{1,2} \times 1965 \text{ tax rate}_2) + \ldots + (w_{1,17} \times 1965 \text{ tax rate}_{17})$ and the weights, $w_{i,j}$ are normalized. Franzese and Hays (2007b) calculate the spatial weight using a standardized *binary contiguity-weight matrix* which begins by coding $w_{ij} = 1$ for countries i and j that share a border and $w_{ij} = 0$ for countries that do not border. Next, the resulting spatial-weighting matrix is then row standardized by replacing the ones in each country's row in the weight matrix with $1/N$, where N is the number of countries the country borders. This procedure normalizes the sums across rows of cell entries to 1 and creates a non-uniform weighting matrix. In other words, Franzese and Hays (2007b) emphasize that the weighting matrix captures two pieces of intuition. First, that country $i's$ importance in determining country $j's$ capital-tax rate may not be the same as country $j's$ importance in determining $i's$ tax rate. Second, that in the substantive issue of tax competition, investors allow tax rate differentials to influence their investment decisions, *ceteris paribus* (only if two economies are in close proximity will tax rates play a decisive role as multinational corporations seek to use the country as an "export platform").

Given the emphasis spatial interdependence regression models place on joint determination of policy outcomes, one might suspect that this approach offers a way to properly estimate the creation of multilateral alliances. Specifically, one could conceptualize membership in the same alliance as analogous to "neighboring" states in the capital taxation context. Returning to the notation from the previous section, $\mathbf{y}$ is an $NT \times 1$ vector of capability observations stacked by country and the *binary contiguity-weight matrix* is computed by coding $w_{ij} = 1$ for countries i and j that are members of the same *multilateral* alliance, $w_{ij} = 0$ for countries that are not part of the same multilateral alliance.

However, one must keep in mind that the weighting matrix captures the ability of state i to influence state j on policy y. In other words, it captures whether or not i and j have a connection, which is precisely what one needs to estimate! This means that whereas the spatial regression model treats $\mathbf{W}$ as a prespecified right hand side variable, $\mathbf{W}$ is actually the element one needs to estimate as a left hand side variable. It is for this reason that spatial interdependence regression will be inappropriate for modeling k-adic data.

2.5.3 Evolving Network Analysis

Another approach directly models the interdependence of states as a network and then statistically estimates the network data. This is the approach of Warren (2009). His primary finding is that states prefer to ally with states who share similar patterns of alliance ties. Unfortunately, Warren only asserts that the existing methods create a bias; he neither illustrates this bias, nor shows that his approach of statistically estimating alliance data does, in actuality, reduce the bias.

Longitudinal-network (or "evolving networks") models attempt to simultaneously model the connections and the effect of connections or of others' actions via the connections on node characteristics (such as their behaviors).[12] However, it is unlikely that such models will properly model a k-adic DGP. To understand why this is the case, one need only briefly consider the setup of co-evolving networks models.

Formally, let N actors be connected according to an observed, binary endogenous, and time-variant connectivity matrix, $\mathbf{x}$, with elements $\mathbf{x_{ij}(t)}$, representing the connection between actor i and j at time t (which is analogous to the weighting matrix, $\mathbf{W}$, of Franzese and Hays). Let $\mathbf{z}$, be a vector of N observed, binary behaviors at time t (analogous to $\mathbf{y(t)}$ in Franzese and Hays). Actors have opportunities to make changes in their network connections, switching on or off one time or doing nothing. When the opportunity to change network connections arrives for some i, this actor chooses to change the status on one of his/her $N-1$ connections, turning it on or off, or leaving them all unchanged. The actor makes this choice by comparing the values of some objective function specified by

$$f_i^{net}(x, x', z) + \epsilon_i^{net}(x, x', z) \tag{2.16}$$

[12] The discussion that follows is drawn from Franzese, Hays, and Kachi (2009) and Steglich, Snijders, and Pearson (2007).

where f^{net} is a deterministic objective function that can be interpreted as a measure of the actor's satisfaction with the result of the network decision, and ϵ^{net}, is a random disturbance term representing unexplained change that is assumed extreme-value distributed. This, coupled with the additional assumption that the data are IIA, allows the objective function to take on multinomial logit shape of categorical choice (where each category is a relation with another actor).

This specification illustrates two reasons why the evolving networks approach will not properly model k-adic data. First, the notation $\mathbf{x_{ij}(t)}$ illustrates that the presence of a connection is *dyadically* measured. More concretely, even though this approach can identify the presence of a connection between any two of n countries and even determine if these links serve to "close" a triangular relationship, it cannot distinguish between a triangular relationship that is closed due to the presence of a single trilateral alliance and a triangular relationship that is closed due to the presence of interlinking bilateral alliances.

Second, the underlying IIA assumption means these models treat each node's decision regarding which ties to form as independent of every other nodes' decisions. Thus, we do not directly model, for the specific edges between specific i, j, and k, that the probability of i and j being connected is a function of the probability that j and k are connected.

Third, current methods for statistically estimating network data, such as the *SIENA* software package, can estimate the presence of ties and the similarity in covariate values between no more than two states. For example, in the above simulations, all countries have a country specific explanatory variable (i.e., the level of capabilities). The values of this explanatory variable can be entered into a network-analytic program, which then estimates a summary statistic (such as capability "similarity" between two states). This summary statistic is some function of the edges and/or nodes – which is to say, it is some function of the 1's or 0's that indicate a connected or a non-connected pair of nodes and/or of characteristics of those nodes. For instance, *SIENA* uses the following formula to compute "similarity":

$$1 - \left(\frac{|v_i - v_j|}{r_V}\right) \tag{2.17}$$

where v_i is the capability score of state i, v_j is the capability score of state j, and r_V is the difference between the highest and lowest capability scores *in the dataset*. Hence, because this formula only measures the similarity in capabilities between two states, *SIENA*, in essence, only estimates how this dyadic statistic impacts the probability of two states forming an alliance. Consequently, network analytic models cannot circumvent the essentially dyadic nature of the information in the data as recorded and used (e.g., they can not distinguish from $i - j - k$ connected in three binary treaties from $i - j - k$ connected in one trilateral agreement).[13]

2.6 Conclusion

Negotiations often involve more than two actors, but international relations scholars often divide the actors in multilateral events into a series of dyadic relations. However, through a series of simulations, I show that one cannot use *dy*-adic data to recover what is a k-adic data generating process. Instead, one must analyze k-adic events using k-adic data. Doing otherwise will result in flawed inferences.

Having developed the concept of k-adic data, I now have an appropriate unit of analysis for conducting analysis on international negotiations. The next chapter will

[13]Two other methods for estimating network data include exponential random graph models (ERGM) (see Robins and Morris (2007) for a primer on these models) and neural network models, as applied by Beck, King, and Zeng (2000). However, both are greatly limited in their ability to capture k-adic processes as they require a dyadic based measure of connectivity between nodes.

make use of k-adic data to answer the first of my two substantive questions: does issue linkage increase the probability of states reaching agreement.

Tables

Table 2.1: Sample of Complete Triadic Dataset

Dyad	Mem1	Mem2	Mem3	Cap1	Cap2	Cap 3	Cap Ratio	ALLY
102	1	2	3	70.01	62.24	26.23	0.44	0
102	1	2	4	70.01	62.24	16.30	0.47	0
102	1	2	5	70.01	62.24	51.48	0.38	0
102	1	2	6	70.01	62.24	85.39	0.39	0
102	1	2	7	70.01	62.24	35.41	0.42	0
102	1	2	8	70.01	62.24	24.88	0.45	0
102	1	2	9	70.01	62.24	24.29	0.45	0
102	1	2	10	70.01	62.24	34.11	0.42	0
102	1	2	11	70.01	62.24	66.35	0.35	0
102	1	2	12	70.01	62.24	5.38	0.51	0

Table 2.2: Trilateral Alliance Simulation Results (True $\beta_1 = 0.25$)

	Triadic DGP estimated with Triadic Data	Triadic DGP estimated with Dyadic Data	Triadic DGP estimated with Choice-Based Sample
Average $\hat{\beta}_1$	0.251	0.46	0.251
Bias	0.001	0.21	0.001
Root Mean Squared Error	0.13	0.66	0.16
Over Confidence	0.17	1.09	0.24

Table 2.3: Bilateral-Trilateral Alliance Simulation Results (True $\beta_1 = 25$)

	Triadic-Dyadic DGP estimated with Triadic-Dyadic Data	Triadic-Dyadic DGP estimated with Dyadic Data	Triadic-Dyadic DGP estimated with Trilateral Alliances Removed	Triadic-Dyadic DGP estimated with Choice-Based Sample
Average $\hat{\beta}_1$	25.03	22.54	25.18	25.21
Bias	0.03	-2.46	0.18	0.21
Root Mean Squared Error	1.39	4.05	2.17	2.07
Over Confidence	2.35	8.27	3.17	3.32

Table 2.4: FIVE-adic Simulation Results (True $\beta_1 = 26$)

	FIVE-adic DGP estimated with Choice-Based Sample	FIVE-adic DGP estimated with Dyadic Data	FIVE-adic DGP estimated with Non-Dyads Removed
Average $\hat{\beta}_1$	28.3	13.95	34.31
Bias	2.3	-12.05	8.31
Root Mean Squared Error	0.01	1.34	7.73
Over Confidence	0.05	2.53	18.72

Table 2.5: *K*-ad Year Alliance Formation Regressed on *Capability Ratio*

DATASET	(1) Dyadic	(2) Remove K-adic Alliance	(3) K-adic Choice-Based Sample
Capability Ratio	-1.79*** (0.157)	-1.08** (0.39)	7.21** (3.43)
Constant	-3.69*** (0.132)	-6.13*** (0.33)	-49.81*** (2.69)
N:	570,390	570,390	215
Estimation Technique	Logit with standard errors clustered on the dyad	Logit with Standard errors clustered on the dyad	Rare-events logit

Standard errors reported in parentheses (non-clustered standard errors produce similar results for models 1 and 2)

Table 2.6: Descriptive Statistics

	Observations	Mean	Std. Dev.	Min	Max
K-adic Data					
Distance	287	66.51	27.87	0	109.29
Joint Threat	299	0.04	0.16	0	1
Joint Democracy (continuous)	295	0.29	0.33	0	1
Joint Democracy (dummy)	299	0.12	0.32	0	1
Capability Ratio	296	0.73	0.21	0.18	0.99
Dyadic Data					
Distance	515753	63.57	24.38	0	111.33
Joint Threat	516914	0.05	0.22	0	1
Joint Democracy	411476	0.10	0.30	0	1
Capability Ratio	570390	0.83	0.15	0.5	0.99

Table 2.7: *K*-ad Year Alliance Formation with Gibler and Wolford (2006) Data

DATASET	(1) Dyadic	(2) Remove K-adic Alliance	(3) K-adic Choice-Based Sample	(4) K-adic Choice-Based Sample
Distance	-0.042***	-0.055***	-0.10***	-0.11***
	(0.001)	(0.004)	(0.02)	(0.03)
Common Threat	1.615***	1.341***	6.02***	5.42***
	(0.079)	(0.169)	(0.95)	(0.97)
Joint Democracy (continuous)	-0.406***	-0.439	-2.78**	
	(0.136)	(0.293)	(1.16)	
Joint Democracy (dichotomous)				-2.56*
				(1.38)
Capability Ratio	-0.576**	0.358	1.33	0.63
	(0.250)	(0.639)	(3.49)	(4.21)
Constant	-3.46***	-5.591***	-23.65**	-22.68
	(0.216)	(0.56)	(3.69)	(4.45)
N:	411,476	411,476	202	203
Estimation Technique	Logit with standard errors clustered on the dyad	Logit with Standard errors clustered on the dyad	Rare-events logit	Rare-events logit

Standard errors reported in parentheses (non-clustered standard errors produce similar results for models 1 and 2)

CHAPTER III

Does Issue Linkage Increase the Probability of Reaching Agreement?

3.1 Introduction

Having introduced the k-ad as a unit of analysis capable of accomodating multilateral events, I can now explore the first substantive question posed in the introduction: do issue linkages increase the probability reaching a negotiated agreement? This chapter will offer the first direct and systematic evidence that issue linkage can help "seal the deal" on a negotiated agreement.[1]

Examples of the existing qualitative evidence include the US-USSR arms control talks in McGinnis (1986), the numerous studies from security and international political economy featured in the 1985 special "Cooperation Under Anarchy" issue of *World Politics*, and, more recently, Davis (2009)'s account of the Anglo-Japanese alliance treaty talks between 1902 and 1923. With respect to large-n tests, Davis (2004)'s study of agricultural trade negotiations is the first (and perhaps only) study that attempts to see if issue linkage can clinch a negotiated agreement. Using agricultural commodity negotiations between the US and Japan or between the US and EU from 1970 to 1999, Davis finds that higher levels of linkage are associated with higher

[1] In contrast, there is extensive large-n analysis on the use of economic sanctions/bribes to achieve short term policy concessions (see Dixon 1989, Martin 1993, Drezner 2000 and 2003, Hufbauer et. al. 2007, Stam and McGillvray 2004, Kuziemko and Werker, 2006, Thacker, 1999, Stone, 2008.

levels of negotiated agricultural liberalization. However, Davis' *linkage* variable actually codes the institutional setting of the negotiation (since increased institutionalization is positively correlated with increased opportunities to link across trade products), not the direct presence of a linkage offer.[2] Consequently, though Davis shows that issue linkage *and* institutionalization appear to increase the probability of negotiated agreements, one cannot conclude from the study that the positive association between the *linkage* variable and agricultural liberalization is due exclusively and directly to linkage.

A primary reason no previous studies directly measure issue linkage's impact on the probability of states reaching a negotiated agreement is that data collection efforts have focused almost exclusively on treaty negotiations that end in agreement. For example, the Alliance Treaty Obligations and Provisions (ATOP) database provides details on alliance treaties, some of which contain economic linkage provisions. However, identifying the effect these provisions have on the probability of an alliance treaty negotiation ending in agreement requires also considering instances in which such provisions are offered, but failed to close the deal.

Therefore, drawing from diplomatic histories covering European relations prior to 1945, I create a dataset of failed military alliance treaty negotiations involving at least one European state from 1815 to 1945. I focus on European states prior to 1945 as these are the states and time period for which the diplomatic historical record is most complete. Combining this data with existing data on alliance treaty negotiations that ended in agreement, I test if offering to expand military alliance negotiations along an economic dimension (specifically trade) increases the probability of agreement. Military alliances offer a useful window through which to test the claims of the

[2] For example, the highest score for the *linkage* variable, 4, codes if the negotiations took place during the Uruguay round of the General Agreement on Trade and Tariff negotiations, while the lowest score for *linkage*, 1, codes if the negotiations are simply bilateral agricultural talks.

issue linkage literature, largely because the explicit inclusion of a trade cooperation provision in a military alliance treaty is an obvious form of issue linkage. Also, since alliances address something as essential as the ability of states to counter external threats, the state-by-state variation in the salience of alliance formation should be much less than in negotiations over other treaties.

Applying matching techniques, I find that the effect of linkage can be quite substantial: trade linkage increased the probability of agreement by 36 percentage points during the 1860 to 1945 time period. Using a series of sensitivity analysis tests, I find that this results is quite insensitive to omitted variable bias and the possibility of missing linkage offers.

The remainder of this chapter is organized as follows. Section 2 describes issue linkages and the argument that they increase the probability of states reaching agreement. The section also presents dissenting views and ends by arguing that these contrarian views persist due to a lack of systematic quantitative evidence to the contrary. Section 3 describes the two primary hurdles preventing scholars from quantitatively studying the effect of issue linkage: a difficulty in identifying when unrelated issues and a lack of data on failed negotiations. Section 4 describes how data on military alliance negotiations can overcome both hurdles. Section 5 describes my empirical research design, while section 6 presents the results from my analysis and subjects these results to a variety of sensitivity tests. Section 7 concludes.

3.2 Issue Linkage in Theory

Time and again, scholars highlight issue linkage as a critical diplomatic tool. Sebenius, drawing on several historic examples, states that “linkage is a prominent and venerable practice” (Sebenius 1983, p. 283), while Tomz claims that linking issues is "a central idea from international relations theory" (Tomz 2007 p. 7). What is linkage, why is it so important, when will states use it, and who, if anyone, disagrees with the prevailing view that issue linkage is useful for negotiation tool? This section reviews the extant literature's answers to these questions.

3.2.1 What is issue linkage?

Issue linkage is a bargaining strategy.[3] Specifically, Sebenius (1983) defines issue linkage as the simultaneous discussion of two or more issues for joint settlement. Similarly, Haas (1990, p. 76) defines issue linkage as bargaining that involves more than one issue. For example, states could have salmon catch quotas negotiated in connection with the nutritional needs of consumers or have a nuclear weapons test ban negotiated along with limits on strategic weapons.[4]

When issues are linked explicitly by having both addressed in the final treaty text, this is known as expanding the "scope" of a treaty (Koremenos, Lispon, and Snidal 2001, p. 770). Though the practice of including all issues in the same treaty text is quite common (and most anecdotal examples drawn from the literature focus on these instances), it can be the case that, from time to time, linkage is implicit. This means the states agree to link agreement on separate issues, but choose to sign individual agreements for each issue. As will be discussed in the research design section, empirically evaluating the impact of linkage is most feasibly done by focusing on instances of explicit linkage, rather than implicit linkage.

[3] This is different from the use of the word "linkage" to describe complex interdependence of nations (Keohane and Nye 1977), national-international interconnectedness (Wilkenfeld 1973), or across-system phenomena (Rosenau 1969).

[4] Examples from Haas 1990, p. 76.

3.2.2 Why is issue linkage important?

Within in the international cooperation literature, issue linkages are part of a more general concept called *side-payments*. A side-payment occurs when policymakers use either direct monetary payments (e.g. bribes) or material concessions on other issues (i.e. issue linkages) to encourage concessions on a given issue (Friman 1993, p. 388 and Tollison and Willett 1979, p. 426).[5] Thus, the prevailing view is that issue linkages can help states reach an otherwise unattainable level of cooperation.

To understand how issue linkages increase the probability of states reaching agreement, consider the two types of issue linkage identified by Haas (1980): tactical and substantive (Haas 1980, p.371).[6] Tactical linkage occurs when the issues being linked are in no intellectually coherent way related to one another (Haas 1980, p. 373). This is arguably the most common conceptualization of linkage. As Tollison and Willett state, "most of the highly publicized cases of proposed issue linkages appear to have been motivated by attempts of individual countries or groups of countries to extend their dominant bargaining or veto power in one particular issue area into other areas" (Tollison and Willett 1979, p. 425). As Wallace famously claimed "linkage between unrelated or only loosely-related issues in order to gain increased leverage in negotiation is an ancient and accepted aspect of diplomacy" (Wallace 1976, p. 164).

If used to provide a positive inducement, linking unrelated issues diminishes conflict during negotiations (Aggarwal 1998, p. 16). However, Wagner (1988) argues that even when issue linkage is used in a blatant power play, thereby increasing

[5]It is worth noting that Tollison and Willett view issue linkage as superior to direct monetary payments because direct monetary payments are extremely unlikely to be politically feasible (Tollison and Willett 1979, p. 426). Unfortunately, they do not go on to detail the nature or source of these infeasibilities.

[6]Haas identifies a third type of linkage: fragmented linkage, which is employed when there is uncertainty over the distribution of benefits from creating a coalition with the goal of acquiring joint gains. Much like tactical linkage, intellectual coherence between the issues is secondary to the objective of maintaining the coalition and the goal of fragmented linkage is to increase the probability that states will reach or remain in agreement.

conflict during the negotiation, the linkage is still directed toward creating a cooperative arrangement that, absent linkage, would not be possible. This is why Haas states that tactical linkage "is a cheap way to increase pay-offs because it expands the agenda of possible benefits to be derived. Since the issues are not inherently connected, the sacrifice of a peripheral demand poses no problem as long as what is really wanted is accomplished" (Wagner 1988, p. 479)

In contrast to tactical linkages, substantive linkages result from knowledge making evident the intellectual coherence of issues. For example, engineers, scientists, economists, and civil servants in Latin America began to think of "technology transfer" as a single issue area covering a variety of issues such as foreign capital inflows, patent acquisition, and constructing factories. Thus, rather than each issue being an end in itself, they were considered a collective means toward a more complicated end (Haas 1980, p. 374). Substantive linkages originate when new scientific and technical information make evident to the negotiation participants (perhaps through the persuasion of another participant) the coherence of previously disparate issues (Haas 1980; Aggrawal 1998, p. 16).

Though making substantive linkages distinct from tactical linkage can be useful for understanding the process by which linkage is employed, such a distinction is unnecessary for understanding the end goal of linkage. Tactical linkages (whether as an inducement or as a powerplay) and substantive linkages are both directed toward helping states reach an otherwise unattainable level of cooperation. This point is echoed by Oye (1992). For Oye, whether linkage is used to generate mutual gains (which Oye calls "exchange" linkage), to coerce (which Oye calls "extortion" linkage), or to draw attention to connections between functionally related issues (which Oye calls "explanation" linkage), they all boil down to a single objective: to increase the probability that states reach a level of cooperation that would otherwise be unattainable; "[If] the linker prefers that the linkee play Y...[then] extortion, exchange, and explanation are all undertaken to predispose the linkee toward choice Y" (Oye 1992, pp. 38-43).

3.2.3 When do states choose to use issue linkage?

When will states require the use of issue linkage to reach a cooperative outcome? According to the literature, it is when states need to overcome the cooperation problems that can appear during negotiations, namely distribution problems, enforcement problems, and a large number of actors (Koremenos, Lipson, and Snidal 2001).

Distribution problems arise when actors have different preferences over alternative possible agreements. For instance, the benefits of an issue could accrue primarily to a few actors, while the costs fall disproportionately on others. In this instance, adding another issue to the negotiations can redistribute the benefits. This, in turn, allows all participants to experience some gain. The ability of linkage to overcome a distribution problem is echoed by others in the literature. Sebenius holds that linkage is important because "what appears incontestably a bare minimum on one particular issue by itself may in fact be flexible when this issue is considered together with favorable settlements on other questions" (Sebenius 1983, p. 298). Tollison and Willet point out that when the distribution of benefits from agreement is highly skewed toward one (or a few) countries, "linkage of issues with offsetting distributional patterns can help promote agreements which otherwise might go unconsumated because of distributional effects" (Tollison and Willett 1979, p. 427). For instance, during the Nuclear Nonproliferation Treaty talks, the five nuclear powers offered the transfer of peaceful nuclear technology to smaller states in order to induce the small states to forgo nuclear weapons (Koremenos, Lipson, and Snidal 2001, p. 770).

Even in crisis bargaining situations, Morgan highlights how "an agreement leading to the peaceful resolution of an international crisis often becomes possible when an issue, not originally in contention, is brought into the bargaining for linkage purposes" (Morgan 1990, p. 311). This is similar to the view of linkage by Morrow; "a linkage deal requires two issues that the sides believe are of different importance. Each side receives concessions on the issue it believes is of greater relative importance. ...[If] done properly, both sides prefer the linkage deal to going to war over the initial issue" (Morrow 1992, p. 155).

Enforcement problems arise when one state believes that a negotiating partner is susceptible to reneging on the agreement. In this instance, an additional issue could incentivize all parties to remain committed to the treaty. This is the theoretical argument found in Stein (1980), where combining payoffs across games compels states to maintain the optimal, rather than sub-optimal, single-shot equilibrium.[7] Using linkage as an enforcement mechanism has been widely discussed by economists in the context of including environmental and labor standards in free trade agreements (Limão 2005, 2007). For example, the United States-Jordan Free Trade Agreement, the United States-Singapore Free Trade Agreement and the Dominican Republic-Central American Free Trade Agreement all include labor standards provisions, including dispute settlement and sanctioning mechanisms for failure to adequately comply with these standards.[8]

The third problem issue linkage can overcome is having a large number of participants in a negotiation. More actors in a negotiation means more preferences and, hence, the possibility these preferences will diverge. In this case reaching agree-

[7] Other work looking at linkages across games include Bernheim and Whinston 1986, McGinnis 1986, and Lohmann 1997.

[8] Article 6 of the US-Jordan Agreement, Article 17 of the US-Singapore Agreement, and Article 16 of DR-CAFTA. Text of each treaty available at http://www.ustr.gov/trade-agreements/free-trade-agreements/ (downloaded on 1-11-2011).

ment on a single issue could prove difficult. However, as with solving the distribution problem, adding issues provides more opportunities for the actors to experience gain.

With all three cooperation problems (distribution problem, enforcement problem, large number of actors), linking issues can generate additional benefits that can either satisfy a reluctant party or give parties an incentive to remain committed to the final agreement. Hence, it is by overcoming these cooperation problems that linkage allows states to reach a level of cooperation that would not otherwise be possible.

3.2.4 Who disagrees?

Whether to redistribute benefits or incentivize commitment, the claim found in much of the literature is that expanding negotiations along a new issue dimension can secure and maintain an otherwise unattainable level of cooperation. This view, however, is not unanimously held. Some notable scholars question the ability of linkages to secure otherwise unattainable cooperative arrangements.[9]

Moravcsik (1998) argues that the major constraint on linkage strategies lies in their domestic distributional implications. Though linkages may create benefits, they also create domestic losers who, if highly concentrated, tend to generate more political pressure than winners. This means linkage will be domestically viable only where adjustment costs are moderate and the potential for linkage is limited. Moravcsik expects to see issue linkage employed only when the costs are imposed on relatively diffuse, unorganized, or unrepresented groups (such as taxpayers or consumers). Since such circumstances are rare, this limits the ability of negotiators to effectively employ linkages; "On this logic, the potential for linkage is far more limited than the

[9]Grieco (1988) also questions the ability of issue linkage to secure cooperative outcomes. However, Grieco's claim focuses when the two states that have difficulty reaching agreement on one issue and decide to link to an issue where the benefits still favors the same actor over the other. This is not the scenario envisioned by the proponents of issue linkage: that the linked issue offers relatively more benefits to the state that benefited relatively less on the original issue.

potential for concessions within issue-areas" (Moravcsik 1998, p. 65)[10] As a result, Moravcsik argues that, when linkage does occur, it will likely take the form of "symbolic concessions", rather than substantively meaningful trade-offs (Moravcsik 1998, p. 65).

Morrow (1992), focusing on crisis bargaining situations, highlights reasons to expect linkage offers to be ineffective and infrequent. Even if one assumes that the promise pertaining to the linkage issue is itself enforceable, states could fail to use linkage because of its signaling properties.[11] According to Morrow, offers of linkage will signal a party's resolve during a crisis. If the offer does not communicate a willingness to fight, the receiver may refuse linkage in hope of gaining a better bargain. Thus, even when a linkage deal could make both sides better off by avoiding war, the sides may abstain from offering linkage because the linkage offer could be interpreted as a sign of bargaining weakness.

There are further reasons to suspect that states will be reluctant to employ linkage even in non-crisis bargaining situations. implementing linkage is not costless. First, frivolous or extraneous use of linkage can create "brittle" agreements, whereby failure in one area can 'unravel' an entire agreement (McGinnis 1986; Koremenos et al 2001). Second, the linkage provision could prove politically unpopular with domestic audiences. Morgan provides the example of Austria and Italy prior to the Seven Weeks War. Despite being desperate for funds, Austria refused cash payment from Italy and Prussia in exchange for Venice. Franz Joseph believed that accepting the Italian offer would inflict a serious blow to the prestige of the empire and be highly immoral (for "selling" people and their homes to foreigners) (Morgan 1990, p. 328). This is in addition to the 'domestic losers' constraint highlighted above by Moravcsik.

[10]This critique of linkage could be countered by by Putnam's concept of *synergistic linkages*, where linkages could just as easily be used to generate domestic support as to create domestic opposition (Putnam 1988, p. 446).

[11]See also Eichengreen and Frieden 1993 on the need to ensure the credibility of the linkage offer.

Third, linking unrelated issues can be costly simply because it requires bureaucrats to ensure that the state is in compliance with the new provision. Thus, linkage may not work because, put simply, it might prove too costly to implement.

3.2.5 Addressing the debate: The need for empirical evidence

Much of the literature holds that linkage increases the probability of agreement. This view does, however, face some prominent challengers. These scholars claim that linkage will be infrequent, merely symbolic, and/or completely ineffective. Though numerous studies have provided anecdotal evidence of linkage's ability to increase the probability of states reaching agreement, the prevailing view of issue linkage's beneficial impact has not been subjected to systematic quantitative analysis. Until this has been done, one will not know if the prominent view is on sound empirical footing or if the contrarian view has a solid empirical case. Therefore, the next section will explain the two primary reasons why previous work has not been able to systematically test the effect of linkage on international negotiations.

3.3 Why Have We Not Measured the Effect of Linkage?

There are two major reasons that scholars have been unable to systematically measure the effect of issue linkage on the probability of states reaching a negotiated agreement. The two reasons are (1) a difficulty in identifying when unrelated issues are linked and (2) a lack of data on failed negotiations.

3.3.1 Inability to identify linkage between unrelated issues

When looking at a given treaty, it can be difficult to tell if the issues could have been addressed in separate negotiations. This is important, because if it is not the case that the issues could have been addressed in separate treaties, then it is unlikely that the issues were linked for the purpose of achieving a cooperative outcome. Koremenos, Lipson, and Snidal make this point quite clearly:

> "One difficulty in analyzing scope is that the issues themselves are not clearly defined. Does trade in all commodities constitute an issue? Or should we distinguish agricultural goods from manufactures? ... *The problem is simplified when negotiations are expanded to cover items that could clearly be dealt with separately or were not previously linked*" (Koremenos, Lipson, and Snidal 2001, p. 771. Emphasis added).

3.3.2 No data on failed negotiations

The second and perhaps primary reason no previous studies have directly measured issue linkage's impact on the probability of states reaching a negotiated agreement is that data collection efforts have focused almost exclusively on successful negotiations. There are several examples of such data collection efforts. The Trade Agreement Dataset lists all trade agreements involving reciprocal concessions on tariffs or tariff-equivalents (such as import quotas) over the 1815 to 1914 time period (Pahre 2008). The United Nations Conference on Trade and Development (UNTAD) lists all bilateral investment treaties (BIT) signed since 1959 (the year the first BIT was signed between Germany and the Dominican Republican). The Continent of International Law (COIL) database which draws from the United Nations Treaty Series (UNTS), records detailed information on the various provisions in a variety of treaties (Koremenos 2005). The Alliance Treaty Obligations and Provisions (ATOP) dataset provides a comprehensive list of all known military alliance treaties formed between 1815 and the present (Leeds, Ritter, Mitchell, and Long 2002). The Multilateral Agreement and Treaty Record Set (MARTS) includes 6976 multilateral treaties signed between 1595 and 1995 (Denemark and Hoffman 2008). Finally, The World Treaty Index seeks to provide a comprehensive dataset of treaties formed during the 20th century (Pearson 2001).

These treaty collection efforts are immensely useful. However, to identify the effect linkage provisions have on the probability of states reaching a negotiated agreement, one must also consider instances in which such provisions are offered, but fail to close the deal. This requires data on negotiations that end in nonagreement.

3.4 A Solution: Military Alliance Treaty Data

Observational data on military alliance treaties provide an ideal window through which to measure the effect of issue linkage. This is because military alliances address the two hurdles that prevent direct and systematic investigation of issue linkage's effect: a difficulty in identifying when unrelated issues are linked and a lack of data on failed negotiations.

3.4.1 Alliances and the linking of unrelated issues

Military alliance treaties are formal agreements to (1) come to the defense of another state, (2) cooperate with another state in attacking a third state, or (3) abstain from attacking another state. There already exists, via the Alliance Treaties Obligations and Provisions (ATOP) database, a large amount of data on military alliance agreements. Most importantly, because ATOP codes the various provisions of each military alliance treaty from 1815 to the present, we know that some alliance treaties contain explicit economic cooperation provisions.[12] These provisions call for either the granting of foreign aid or for reducing trade restrictions between the

[12] Some research has already made use of this data on economic provisions in alliances, but for the purpose of evaluating linkage (see Long and Leeds 2006). Powers (2004, 2006) and Powers and Goertz (2006), do not use ATOP data, but do study how many regional economic institutions frequently incorporate explicit security cooperation provisions.

parties (such as the granting of Most Favored Nation status).[13]

Consider a few examples of these economic cooperation provisions.[14] Article 5 of the 1971 alliance between the Soviet Union and Egypt (ATOPID 3670) states, "...The parties shall expand trade and shipping between the two states on the basis of the principles of mutual advantage and most favored nation treatment." Article 9 of the 1946 mutual defense pact between the United Kingdom and Jordan (ATOPID 3040) proclaims that "Neither High Contracting Party will extend to the nationals or commerce of the other treatment less favorable in any respect than that which he accords to the nationals and commerce of the most favoured foreign country."

Such provisions, especially those regarding trade, are of interest because, though arguments exist that purport the benefits of economic and security linkages (Gowa and Mansfield 1993), states quite frequently (if not primarily) negotiate alliance agreements and trade agreements separately from one other. For example, the United States and Canada are members of an alliance (the North Atlantic Treaty Organization) that was negotiated separately from and long before its current trade treaty (the North American Free Trade Agreement). Since there is no inherent reason that the two issues *must* be linked to one another, the explicit tying of economic cooperation to a military alliance is an obvious form of issue linkage. Stated differently, alliances with trade provisions are obvious instances in which the negotiations were "expanded to cover items that could clearly be dealt with separately or were not previously linked" (Koremenos, Lipson, and Snidal 2001, p. 771). Of course, one

[13]These economic provisions are captured by ATOP's ECAID variable. ECAID codes instances in which a military alliance treaty includes a provision granting economic concessions to one or both sides of the agreement. It is an ordinal variable taking on a value between 0 and 3: 0 if no such provision is found in the treaty; 1 if general or nonspecific obligations for economic aid are found in the treaty; 2 if one or more members promise economic support for postwar recovery; and 3 if one or more members commits to trade concessions, including the granting of MFN status. The ECAID variable is described in Leeds, Ashley. ATOP Codebook. 2005. p.31, available at http://atop.rice.edu/.

[14]Each example below is taken from the answer to question 55 of the ATOP code sheets for the respective treaties. Question 55 reads, "55. Does the treaty include provisions for economic aid or other enticements (include trade concessions, post war recovery, etc.)? (Yes, No) If Yes, describe these provisions." These code sheets are available at http://atop.rice.edu/.

might argue that these provisions shown above are so general that they exemplify the "shallow linkage" that Moravcsik claims is the only plausible form of linkage. However, this would suggests that if I find that these relatively general provisions can indeed increase the probability of agreement, then one should expect the same to be true of those instances in which states are able to write more specific provisions.

3.4.2 Alliances and failed negotiations

The ATOP dataset is an outstanding source of information on alliance agreements, but there exists no comparable data on alliance negotiations that failed to end in agreement. As a result, when previous studies have used ATOP data to study alliance formation, the dependent variable has been coded in the following fashion: those groups of states (typically state-to-state dyads) that formed alliances and those groups of states that did not (Leeds et al. 2002; Gibler and Sarkees 2004; and Gibler and Wolford 2006). This is problematic, as the latter group conflates those dyads that actually began alliance negotiations but failed to reach an agreement and those dyads that never even attempted negotiations. In order to measure the effect that offering to include economic linkages has on the probability of a group of states forming an alliance, I must know which instances witnessed a failed attempt to form an alliance agreement and, in particular, which of these failed negotiations witnessed the offer of an economic linkage provision.

The sources

I require a source of information from which I can identify failed negotiations. A logical starting point is foreign ministry archives or collections of foreign diplomatic documents such as the *British Foreign and State Papers.* However, this amounts to looking for a needle in a haystack and is costly both in terms of money (for travel) and time. For instance, if one were to focus only on British foreign documents, failed attempts could be identified (assuming the ministry wished to keep documents of the failure), but after extensive time spent reading these documents, one would only have coded the failed negotiations of a single country.

Therefore, an alternative approach is to draw upon the decades of archival research already conducted by historians. This can be done by using published diplomatic histories. Other, highly prominent and widely used international relations datasets were created through similar sources. For instance, diplomatic histories were used to identify cases of alliance formation by Leeds and her co-authors when constructing the ATOP dataset and by Singer and Small (1966) when constructing the original Correlates of War listing of military alliances. Another example includes the *strategic rivals* dataset of Colaresi, Rasler, and Thompson (2007) They draw upon diplomatic and political histories of individual state's foreign policy activities to determine when and with whom decision-makers thought they were in rivalry relationships.

I use a number of diplomatic historical sources, such as the following prominent histories: *European Alliances and Alignments* by William Langer, *A Diplomatic History of Europe Since the Congress of Vienna* by Rene Albrecht-Carrie; *The Transformation of European Politics , 1763 to 1848* by Paul Schroeder; *The Struggle for Mastery in Europe, 1848 to 1918* by John Taylor; and *The Lights That Failed: European International History, 1919 to 1933* by Zara Steiner. My selection of secondary sources is European centered, but this is reasonable given the composition of the ATOP dataset. Specifically, nearly seventy six percent of the alliances formed between 1815 and 1945 have *only* European powers. Thus, perhaps unsurprisingly, European countries were the most involved in negotiating military alliances during this time period. Additionally, I follow Leeds (2002) by focusing on the pre-1945 time period when creating the first version of this dataset. Another advantage of concentrating only on the earlier pre-1945 time period is that the diplomatic historic record is more complete for this time period.

The coding rule

Reading diplomatic histories in order to create a dataset of failed alliance negotiations requires a coding rule. Since I am interested in identifying attempts by states to form an actual alliance treaty (as opposed to attempts to create an informal coalition) I develop the following coding rule: there must exist evidence of a meeting (correspondence of letters or physical meeting) at the diplomatic level (between ambassadors, heads of state, or foreign ministers) where a proposal of a formal (i.e. written) alliance (mutual defense pact, offensive pact, neutrality pact, a military consultative agreement, or a non-aggression pact) is made and then evidence of a rejection/refusal (one side must decline forming the alliance).

When applying this coding rule, it is important to keep in mind that, for example, declining to form an alliance may not take the form of a simple “no” response. These are, after all, diplomats (and, hence, their response could be quite diplomatic)! Moreover, it is worth emphasizing that evidence of a meeting need not entail the two diplomats or heads of state being physically present in the same location. Instead, a “meeting” could entail an exchange of letters. Inevitably, using diplomatic histories as source material will involve making some subjective judgments. Interpreting diplomatic histories lacks the strict objectivity associated with, for example, counting treaty texts. However, as Colaresi, Rasler, and Thompson state, “No phenomenon is so clearcut that counting it does not require some level of interpretation...The point remains that measurement choices rarely boil down to interpreting the raw information versus allowing the facts to speak for themselves. Some interpretation of the raw information is inevitable" (Colaresi, Rasler, and Thompson 2007, p. 29).

Applying the coding rule

I read through these histories looking for instances that meet the criteria of my coding rule.[15] To illustrate how I applied the coding rule, consider the following account from Taylor of a failed attempt at forming an alliance:

> "[Russian Chancellor] Gorchakov said to [French President] Thiers: 'We shall occupy ourselves later with uniting France to Russia', and [Russian Tsar] Alexander II added: 'I should much like to gain an alliance like that of France, an alliance of peace, and not of war and conquest.' These words, uttered on 29 September 1870, defined the Franco-Russian alliance as it was achieved twenty years later; they were of no use to Thiers in the circumstances of the moment. He returned to Paris empty-handed; and the French had to try to reverse the Prussian victories by their own efforts" (Taylor 1954, pp. 214-215).

This excerpt from a larger passage shows that the Russian Chancellor and the French President met, that there was discussion of creating an alliance between the two nations ('I should much like to gain an alliance like that of France, an alliance of peace, and not of war and conquest'), and that this attempt failed (He returned to Paris empty-handed). Another example comes from Schroeder:

> "Russian policy was not hostile to Britain, nor was it opposed to all reform of the Ottoman Empire...In 1836 [Russian diplomat and foreign minister] Nesselrode began seeking an entente with Britain, for the sake of general peace and Russia's economic development. His feelers were ignored

[15] To demonstrate the plausibility of my coding of negotiation failures, I conduct several inter-coder reliability checks. The details of these checks, as well as a guide for recreating the entire dataset, are available upon request.

> at London. Instead, from 1834 to 1838 [British foreign secretary] Palmerston considered various ideas for shoring up the Ottoman Empire against Russia" (Schroeder 1994, p. 735).

Again, there is evidence that the Russian diplomat broached the idea of an alliance to the British diplomat (His *feelers*...), but that this idea was rejected (...were ignored at London.). In addition to identifying failed alliance negotiations, I also code which failed negotiations witnessed the offer of economic linkage. In order to match as closely as possible the coding of economic linkage offers found in the ATOP dataset, I code economic linkage offers as any identifiable offer of trade cooperation or foreign aid. For example, in 1816 Spain sought British assistance in suppressing rebellions within its colonies. However, the British made any assistance conditional on Spain opening its colonies to trade (Schroeder, 1994, p. 630). Spain rejected this demand. Another example involves the following account of Prussia and England in 1850,

> "[Prussian Ambassador] Radowitz was sent to London, more to console him than with any serious purpose. He was empowered to offer the British government reductions in the Zollverein tariff in exchange for an alliance...Radowitz had no success" (Taylor 1954, p. 41).

Applying this coding rule, I identify 127 failed alliance negotiations from 1815 to 1945 involving at least one European state. For each negotiation, I code the following information: year of negotiation; states involved in the negotiation; type of alliance being negotiated (mutual defense pact, offensive pact, neutrality pact, a military consultative agreement, or a non-aggression pact); and whether or not the negotiation witnessed an offer of economic linkage (trade or aid).

3.5 Research Design

3.5.1 Data

Combining the data I collected on failed alliance negotiations with the relevant ATOP alliances (alliances formed between 1815 and 1945 involving at least one European state), I have a complete dataset of 308 alliance negotiations involving at least 1 European power from 1815 to 1945. 181 of these negotiations were successful, 127 were unsuccessful, and 20 had an economic linkage offer (12 with trade offers and 8 with aid offers). The rarity of linkage offers in my dataset is consistent with the claims made by the critics of linkage (such as Morrow 1992) that linkage offers will be more difficult and, consequently, less frequent than scholars originally conjectured. This is particularly unsurprising given that the stakes in alliance negotiations, though not quite as high as in the crisis bargaining situations analyzed by Morrow, are higher than other, non-crisis bargaining situations (as they still involve measures to protect and ensure the survival of the state).

To give the reader a sense of the data, Table 3.1 reports the countries that conducted at least 10 negotiations during the 1815 to 1945 time period, along with the "agreement rate" of these countries (i.e. the number of negotiations that resulted in an alliance, divided by the total number of negotiations in which that country participated). What is notable is that though the major European military powers during this time period (Russia, Germany, France, Austria, and Britain) are at the top of the list, their agreement rates range only between 41 and 54 percent. In contrast, the agreement rates of several minor military states (such as Serbia, Poland, Greece, and Bulgaria) are substantially higher.

3.5.2 Unit of observation

The unit of observation is the alliance negotiation. Since several of the negotiations (both successful and unsuccessful) include more than two states, I follow the recommendation provided in chapter 2 and do not divide the negotiations with more than two members into dyadic observations. This gives me an initial cross sectional dataset of 308 negotiations over the entire 1815 to 1945 time period.

In the analysis that follows, I will only consider those negotiations that took place from 1860 to 1945. This is for both practical and substantive reasons. Practically, data is limited for several covariates for much of the early 19th century. Rather than apply imputation methods to fill in these values, I prefer to focus on the time period for which I have confidence in the data.

Substantively, 1860 is a reasonable starting year since I am concentrating on offers of trade cooperation. 1860 marks when free-trade arrangements in general and the most-favored nation principle in particular became an accepted tool by European diplomacy (Held et al 1999 p. 155 and Bairoch 1989). It is widely recognized by political economists that the signing of the Cobden-Chevalier treaty between Britain and France in 1860, which was the first major free trade agreement between European powers, prompted the growing acceptance of open trade policies throughout Europe (Pahre 2008, Rogowski 1989, and Frieden 2006). In his extensive study of European trade relations prior to 1913, Pahre empirically illustrates the subsequent explosion of international trade treaty initiations after 1860 (Pahre 2008, p. 319).[16]

---

[16] Whether the Cobden-Chevalier treaty was the result of widespread commercial trade becoming technologically feasible (as improvements in the steam ship, rail, and telegraph cable technology lowered the costs of international trade) or due to the British adoption of a unilateral free trade policy in 1849 (with the elimination of the Corn Laws and Navigation Acts) illustrating the benefits of open trade (see, in particular, Schonhardt-Bailey 1996) has been discussed by many scholars and is beyond the scope of this article. For a full discussion, see Eichengreen 1996, Kenwood and Lougheed 1999, Frieden 2006, and Pahre 2008, Bordo, Eichengreen, and Irwin 1999, and Oatley 2004.

3.5.3 Outcome variable

The outcome or dependent variable, *agreement*, is coded 1 if a negotiation ends in agreement, zero if it does not end in agreement. It is important to note that I am only interested in whether agreement is reached. This variable does not indicate if the negotiations were "successful" in the sense that they resulted in a treaty that can be considered just, fair, or equitable. Also, it does not indicate in any way whether states will remain committed to the treaty for a long period of time. Instead, it is only focused on capturing whether the parties formed the treaty.

3.5.4 Key independent variable

The key independent variable is *trade linkage*. This is a binary variable coded 1 if there is an offer to expand the negotiation along an economic dimension (specifically trade), zero if there is not. I am focusing on trade because, between aid and trade, this issue is most clearly separable from an alliance. It can often be the case that the aid provision calls for money that is to explicitly be spent on strengthening the military of the other party. In this case, one could not claim that the aid dimension of the negotiation was separable from the alliance dimension of the negotiation. It is important to note that I am only interested in identifying if the offer to include a linkage provision increases the probability of agreement and, if so, the size of this effect. Answering this question does not require identifying which actor proposed the economic linkage provision and which state benefited most from the economic cooperation provision.[17]

[17] Moreover, for many instances, the state proposing the economic linkage and the state most interested in receiving the economic concession are not clearly distinct. For instance, it is often the case that state A proposes to include a trade cooperation provision because state B told state A that doing so was necessary to seal the deal. In other words, it was state B that initially "proposed" the linkage, even though state A is the one giving aid to state B.

3.5.5 Contextual covariates

One could conduct an initial test of the effect of trade linkage on the probability of agreement by simply comparing the agreement rate of those negotiations that witnessed trade linkage offers to the agreement rate of those negotiations that did not witness linkage offers. However, there could very well exist, on average, highly systematic differences between negotiations with linkage offers and negotiations without linkage offers. Thus, empirically measuring the effect of trade linkage offers on the probability of agreement requires accounting for these systematic differences so as to ensure that, put simply, I am comparing comparable negotiations. This means identifying and operationalizing those factors that are correlated with the presence of linkage and correlated with negotiation agreement.

Military Capabilities of Parties: As one would expect, the core security literature on alliance formation views relative and/or combined capabilities as key to determining whether a group will form an alliance (Morgenthau 1973, Waltz 1979, Walt 1987). However, military capabilities are also correlated with the offer of economic linkages. First, capabilities can capture the distribution problem. Debates can arise over who will contribute what and how much to the alliance. Second, capabilities can indicate if the states face an enforcement problem. According to Morrow (1991), states with large amounts of physical military capabilities (a major power) have high levels of security and autonomy, while states with small amounts of physical military capabilities (a minor power) have only a high level of autonomy. Alliances between states with asymmetric capabilities will enable both states to achieve a more even mixture of autonomy and security because each member brings a high level of different abilities to the alliance. In contrast, an alliance between states with symmetric capabilities, where both states are more likely to have a high level of the same resource (say security) and a low level of the other, generates very little utility surplus. Thus asymmetric alliances are less likely to break in a given period than symmetric alliances because changes in the weaker power's capabilities will not greatly alter the nature of the autonomy for security tradeoff. This means symmetry in capabilities leads to defection, which, in turn, implies an enforcement problem.

As disucssed in chapter 1, when using data that has a k-adic structure (i.e. data in which each observation represents the characteristics of two or more actors), there are several ways to capture the combined/relative capabilities of a group: the ratio of the largest k-ad member to all members in the k-ad; the ratio of the largest to smallest k-ad member; or the size of the smallest k-ad member (following a "weak link" argument).[18] In the results that follow, I will match according to the third of these three metrics. Thus, to measure the military capabilities of the k-ad, I use the number of military personnel held by the member with the smallest number of military personnel as reported by the Correlates of War project.

Presence of a Buffer State: Buffer states are states located between two warring rivals. Poland is a classic example of a buffer state as it was located between Austria, Prussia, and Russia during the 1700's, between Russia and Germany during the 1920's, and between the Soviet Union and Western Europe during the Cold War. According to Fazal (2004 and 2007), buffer states are especially prone to violent *state death*, which Fazal defines as "the formal loss of foreign policy control *to another state*" via military invasion (Fazal 2007, p. 17). The rivals on either side of the buffer state fear that its opponent will conquer the buffer state, thereby gaining a strategic advantage. Though maintaining the sovereignty of the buffer state is ideal for both rivals (as it creates a barrier between the rivals that decreases the

[18] For more on the 'weakest link' rational, see Oneal and Russett 1997.

probability of war), both rivals know the other has an incentive to invade the buffer and gain the strategic advantage. This commitment problem leads inevitably to the demise of the buffer state.

With respect to alliances, Fazal argues that buffer states face a *catch-22* situation: the very factors that compel them to seek alliances also prevent them from forming alliances. According to Fazal, "States - especially threatened states - must balance to survive. But threatened states are unlikely to be able to balance precisely because they *are* threatened" (Fazal 2007, p. 230). Thus, *buffer states* can operationalize the existence of cooperation problems that could prevent agreement during a military alliance negotiation. In particular, if at least one of the states involved in the negotiation is a buffer state, this buffer state could be concerned that the other states will renege from their alliance obligations. Therefore, I create a binary variable called *buffer* coded 1 if, in year t, negotiation i contains at least 1 buffer state as identified by Fazal, zero otherwise.[19]

Geographic Distance Between Parties: Accounting for the distance between the negotiation partners is important for three reasons. First, alliances with neighboring states are easier to sustain than alliances with geographically distant states. Specifically, neighboring states will find it easier to provide military support (as it will not prove as logistically demanding to move forces) and will have an incentive to maintain good relations (by not defecting from the alliance). This is because geographic contiguity reduces the actual transaction costs of moving resources and because contiguity removes the need to move forces through the territory of another state. Second, it is widely recognized in the theoretical and empirical alliance literature that distance plays a key role in determining alliance partners (Walt 1987).

[19] List of buffer states found in Fazal 2007. I thank Tanisha Fazal for providing her data upon request.

Third, distance is a key determinant, not just of alliance formation, but also trade agreement formation. Therefore, it is important to include distance as a variable in order to control for the possibility that the states in the negotiation were already likely to form an alliance and, therefore, simply wanted to place both agreements in the same treaty to reduce transction costs. For these three reasons, I create a variable, *contiguous*, coded 1 if all the states in the negotiation are geographically proximate, zero otherwise.[20] The contiguity of the states is determined using the geographic distance data computed by the *EUGene* software (Bennett and Stam 2004. p. 17).

Number of Parties in the Negotiation: Besides influencing whether or not the negotiations witness a linkage offer, it is also recognized that the number of states involved in a negotiation can alter the dynamics of a negotiation (Sebenius 1983). Therefore, I match on the number of states involved in the negotiation by including a variable indicating if 2 (the minimum number of states involved in a negotiation in my dataset), 3, 4, or 5 (the maximum number of states involved in a negotiation in my dataset) states are involved in the negotiation.

Negotiating during a crisis period: Consider Figure 3.1. It shows the number of negotiations by year from 1815 to 1945. Note the spikes in diplomatic activity around the time of the Crimean War (1853 to 1856), Franco-Prussian War (1870 to 1871), the start of World War I (1914), and the start of World War II (1939). These spikes reveal the tendency of states to seek out alliance partners during the lead-up to major international crises and confrontations. The upcoming crisis likely influenced both the probability of negotiations ending in agreement (since states are more desperate to form an alliance) and the probability of negotiations witnessing a

[20] I obtain similar results if I instead use the maximum distance between any two states in the k-ad.

linkage offer (since, again, states will be more desperate to form an alliance). In fact, Langer states that "the great coalitions of modern history were almost always made just before the outbreak of war or during the course of the conflict itself" (Langer 1966, p. 5). To account for these crisis periods, the variable *peak* equals 1 if the negotiation took place during or one year prior (i.e. in the lead up) to the "peak" years in Figure 3.1 (1815, 1833, 1848, 1854, 1866, 1914, and 1940), zero otherwise.

First Time Negotiation: By "first time negotiation", I mean the first time a group of states appeared in my dataset as being involved in an alliance negotiation that either ended in agreement or non-aggreement. It is quite reasonable to suspect that the first time a particular group of states negotiate an alliance will have very different dynamics than if this group of states have been involved in past alliance negotiations (whether or not those past negotiations ended in agreement). For instance, there will be more uncertainty about the intentions and motivations of states when they are involved in their first ever alliance negotiations as a group. Additionally, if a group of states tried to form an alliance in the past but failed, one could reasonably supsect that such a group will have different incentives for reaching an agreement compared to a group of states that had never previously attempted to form an alliance.

The variable *prior* is coded 1 if a group of states are engaged in their first negotiation as a group, zero otherwise. For example, if a group of three states have never engaged in negotiations as a triad, then this is considered a first negotiation (even if two of the members had engaged in a prior negotiation).

Offensive/Defensive Alliance Negotiation: The variable *offensive/defensive* is coded 1 if a negotiation is focused on the formation of an offensive or defensive alliance, zero otherwise. Because offensive and defensive alliances require actionable obligations, many scholars consider the circumstances leading to their creation to be distinct from consultation, neutrality, and non-aggression pacts (Gibler and Sarkees 2004; Long, Nordstrom, and Baek 2007). Stricter obligations can also make it more difficult to reach agreement and make it more likely that a state will renege on a commitment.

Democractic Participants: Some of the critics of issue linkage, particularly Moravscik, point out that domestic political considerations could constrain the ability of states to use issue linkage or to find offers of linkage politically acceptable (Moravscik 1998). Therefore, it would seem important to somehow capture the extent to which states are subject to domestic political pressure. Though my reading of the historical record of alliance negotiations does not suggest that domestic political pressure played a substantial role in the formation of alliances during much of the pre-1945 time period (as leaders sought to form alliances based on considerations of power politics and due to the fact that many countries during this time period were not liberal democracies), I nevertheless attempt to account for the possibility of domestic political influence. Since one can reasonably assume that democracies are more vulnerable to such domestic political pressures than autocratic regimes, it seems operationalizing the potential influence of domestic political pressure could be accomplished by measuring the presence of democratic states in the negotiation.

There are several approaches for capturing the presence of democratic states in a negotiation. One approach is to create a joint democracy variable, coded 1 if all the participants in the negotiation are democracies, zero otherwise. However, a major drawback of this coding is it treats a negotiation with 4 democracies and 1 non-democracy as equivalent to a negotiation with 1 democracy and 5 non-democracies. Therefore, I instead create the variable *proportion democracies.* This variable captures the percentage of states in a negotiation that are democracies, where a state is considered a democracy if it has a score of 6 or higher on the 21 point Polity IV scale (Marshall and Jaggers 2002).

3.6 Empirical Analysis

3.6.1 Why matching

I have a binary key independent variable (offer of trade linkage or no offer of trade linkage), a binary outcome (alliance agreement reached or no alliance agreement reached), and a variety of covariates that account for contextual factors. Scholars would typically analyze such data using a structural model such as a logit or probit where contextual covariates are included as control variables. However, my data has three features that can make applying structural models problematic.

First, the existence of linkage offers is rare in my sample. This could lead to perfect separation of the data (e.g. there are no linkage offers when negotiations end in non-agreement). Second, I have a small number of overall cases. This means identification in a logit or probit will be driven primarily by the parametric structure imposed on the data, not by the data itself. Moreover, given the large number of control variables, the power of my statistical tests will be dramatically reduced. Third, linkage offers are not randomly assigned. Groups of states negotiating an alliance choose to incorporate a linkage provision. I want to account for this selection bias by, stated simply, ensuring that I am comparing comparable observations.

For these reasons, I turn to matching. Matching, as developed and described by Cochran (1953) and in a series of articles by Rubin (2006), is a pre-analysis procedure that uses minimal structural or parametric assumptions to separate treatment effects from shared background characteristics (Morgan and Winship 2007 and Ho et al. 2007). It consists of pairing each subject in a treatment group with a subject in a control group that has similar (though perhaps not identical) values for a series of covariates. For example, suppose the covariates are gender and age. Then a perfect match for a male that is 37 years old in the treatment group would be a 37 year old male in the control group.[21] The goal of this process is to minimize (if not eliminate) all systematic differences between the treated and control groups other than exposure to the treatment. In short, matching ensures that I am comparing relatively similar negotiations.

Though I can not manipulate the presence of a linkage offer in a negotiation, I assume that trade linkage can be conceptualized as a treatment *if* one has matched on the proper covariates. In other words, I am assuming that if one has two groups of states that are similar on a variety of characteristics (the distribution of military power within each group, the presence of a buffer state within each group, etc.), then the strategic interactions and nature of the negotiations within these two groups will be similar, if not identical (i.e. the presence of negotiation strategies such as ultimatums and brinksmanship will be similar between the two groups). When this is the case, the only observable difference between the two groups will be the presence of a linkage offer.

I implement the nearest neighbor matching algorithm developed by Adadie and Imbens (2002, 2006), which matches the treated and control subjects that are the closest match (rather than requiring an exact match).[22] The determination of a match is made in three steps. First, one identifies those variables on which matches should be based. Call these variables X. Second, one determines the probability of each observation receiving the treatment by regressing the treatment on X.[23]

[21] Example drawn from Rubin 2006, p. 12

[22] The matching algorithm can be run using the **match** command written by Abadie, Drukker, Herr, and Imbens 2003 for STATA.

[23] This regression is conducted using a logit model. I am not concerned with using logit to compute the propensity score because I am not trying to identify the statistical significance of a particular variable on the probability of

This probability is called the propensity score. Third, the propensity score is then used to match observations to one another. The Abadie and Imbens procedure also allows cases to be used as a match more than once. Compared to matching without replacement, matching with replacement generally lowers the bias but increases the variance (Abadie and Imbens 2006 and Abadie, Drukker, Herr, and Imbens, 2003). Once matches are made, one can estimate the average effect of the treatment, or *average treatment effect* (ATE). This is computed by using a simple difference of means t-test between the treated observations and control observations in each pair.[24]

This section presents the average effect of trade linkage offers on the probability of an alliance negotiation ending in agreement. Prior to showing these results, I also provide statistics that evaluate the **balance** (i.e. the extent to which the control and treated groups are similar) achieved via matching. Though evaluating balance prior to computing the treatment effect is useful, one cannot be certain he or she has fully minimized unobserved heterogeneity between treated and control groups in observational studies. Therefore, I end the section with sensitivity analysis on the estimated treatment effects.

3.6.2 Evaluating balance

Since there are no standard practices for evaluating balance in observational tests (Imai, King, and Stuart 2008; Sekhon and Diamond 2008), I present p-values from several tests: t-tests (which assess means and variances relative to a t-distribution), Wilcoxon rank-sum tests (which assess the difference in medians across the groups), and Kimogorov-Smirnov (KS) tests (which assess the similarity in all moments of

being in the treatment group (in which case, concerns about limited power would apply).

[24]The matching procedure written by Abadie, Drukker, Herr, and Imbens 2003 computes the *average treatment effect* by taking the difference in outcome between the treatment observation and control observation for each matched pair and then reporting the coefficient on the constant from a constant only least squares regression (where the difference in outcome is the dependent variable). This produces nearly identical results to the simple difference of means test.

the distributions). In a matched sample, p-values do not refer to a formal hypothesis test, but high p-values suggest that the samples are similar to one another – which is desirable when conducting matching (Imai, King, and Stuard 2008, p. 497-498).[25] In particular, the more a p-value is above 0.10 (the highest standard critical value for t-tests), the more confidence the researcher has that balanced is achieved.

The K-S test is an especially strict test since it is sensitive to the location and shape of the cumulative distribution functions of the populations (Lyall 2010, p. 182). Therefore, I will give primacy to its results. Table 3.2 shows that balance is not achieved for several variables, most notably the *offensive/defensive alliance* variable. Since it is argued in the alliance literature that offensive and defensive alliances are much different than neutrality pacts and non-aggression pacts (due to offensive and defensive pacts calling for actionable security obligations), it seems reasonable to restrict the sample to just negotiations in which an offensive or defensive alliance is being negotiated (i.e. set *offensive/defensive*=1). After making this restriction, I re-run the matching algorithm on the restricted sample. Table 3.2 shows that p-values from the K-S tests are now all well above the 0.10 confidence level, suggesting excellent balance on these covariates.

To make this discussion of balance concrete, consider an actual match made by the matching algorithm. In this particular match, the "treated" negotiation is between Austria-Hungary and Serbia in 1881, while the "control" negotiation is between Prussia and Italy in 1862. As it turns out, these are very comparable negotiations. They were negotiations to form a defense pact between rather asymmetric military powers (Austria-Hungary and Prussia being the major powers. Serbia and Italy being

[25] Imai, King, and Stuart 2008 highlight that the use of hypothesis tests (to, for example, determine if the difference in means is statistically different from zero at some threshold) is unnecessary and inappropriate for evaluating balance because balance is entirely an in-sample property and involves no reference to populations or superpopulations. Therefore, there is no statistical threshold below which the level of imbalance is always acceptable.

the minor powers). These were the first negotiations ever between these countries. In fact, the only real difference between the negotiations is that Austria offer to include a provision to reduce duties on foodstuffs from Serbia, while no such offered was made in the case of Prussia and Italy. As it turns out, the Prussian-Italian negotiation ended without an agreement, while the Austrian-Hungary-Serbia negotiation ended in agreement.

3.6.3 Estimated effect of trade linkage

After minimizing, to the extent possible, the differences across negotiations that witnessed linkage and negotiations that did not witness linkage, what causal effect does an offer of trade linkage have on the probability of agreement? Table 3.3 reports the average treatment effect (ATE) for each sample after the observations are matched using the above variables.

Looking at the results in the ATE column, one can see that the effect of trade linkage is strongly positive. This is true for both the restricted and unrestricted samples (note that the difference in the treatment effect estimates between the restricted and unrestricted samples indicates the extent to which the lack of balance in the unrestricted samples biases the results). Specifically, it appears that an offer of trade linkage increases the probability of agreement by 28 percentage points in the unrestricted sample, with 0.95 confidence intervals showing that the effect is statistically significant. For the restricted sample, an offer of trade linkage is estimated to increase the probability of agreement by 36 percentage points, with the effect once again statistically significant. The difference between the two results shows that including the negotiations over non-aggression and neutrality pacts mitigates the effect of linkage. Overall, both results suggest that issue linkage does increase the probability of agreement and that the magnitude of the effect is quite substantial.

3.6.4 Sensitivity analysis

How robust is the estimated positive effect of trade linkage offers? This section presents results from two forms of sensitivity analysis to show that the identified positive results are quite robust.[26]

Result driven by omitted confounder?

It is plausible that the effect from an unobserved confounder could alter the substantive meaning of the estimated positive ATE. Conducting such analysis is especially important since matching is based on an unconfoundedness assumption, which states that the researcher should observe all variables simultaneously influencing the participation decision and outcome variables. Because this is a strong identifying assumption, one should check the sensitivity of the estimated results with respect to deviations from this identifying assumption. Rosenbaum bounds are a standard procedure for conducting this form of sensitivity analysis. Specifically, letting P_i be the probability that individual i receives the treatment and letting $P_i/(1-P_i)$ be the odds that individual i receives the treatment, then Rosenbaum (2002) shows that the bounds on the odds ratio of $\frac{P_i/(1-P_i)}{P_j/(1-P_j)}$ is given by

$$\frac{1}{e^{\gamma}} \leq \frac{P_i/(1-P_i)}{P_j/(1-P_j)} \leq e^{\gamma} \tag{3.1}$$

It is common to let $e^{\gamma} = \Gamma$ (Keele 2009, p. 8). Equation (1) says that both matched individuals i and j have the same probability of participating in the treatment group if $\Gamma = 1$. However, if, for example, $\Gamma = 2$, then individuals who appear to be similar (in terms of covariate X) could differ in their odds of receiving the treatment by as much as a factor of 2. Thus, if the two individuals are identical on matched covariates, then the estimated treatment effect will still hold even if some

[26] Results from a variety of other robustness checks are available in an on-line appendix.

unobserved covariate makes one individual *twice* as likely as the other individual to receive the treatment. In this sense, Γ is a measure of the degree of departure from a study that is free from hidden bias (Becker and Caliendo 2007).

Table 3.4 reports the results from estimating the Rosenbaum bounds. Of particular interest is the upper bound p-value. This captures the extent to which my estimated ATE is over-reporting the positive impact of trade linkage for the post-1880 time period. If the p-value is significant at a particular level of Γ, then this means the effect would still hold even if an unobserved covariate were to increase the odds of witnessing an economic linkage offer by Γ times. According to Keele (2009) typical studies using observational data have a Γ value between 1 and 2. Table 3.4 shows that even if the odds of a group of states witnessing an economic linkage offer were 4 times higher because of the effect of an unobserved covariate, my inference regarding the effect of economic linkage still holds at the 0.1 confidence level. This is well above the standard range for observational studies. In other words, the effect of an unobserved cofounder would have to be quite large in order to alter my results. Since I matched on the major factors that influence alliance formation and linkage offers, as identified by the alliance literature and the historic diplomatic record, it seems unlikely that an unobserved cofounder would have such a large effect.

Result contingent on missing treatment data?

Though the sources from which I identify failed alliance negotiations may have accurately and fully captured all instances of negotiation, it is very plausible to suspect that the historians generating these sources overlooked offers of economic linkage (since, to these historians, making note of such offers may have been of minor importance) or, perhaps even more likely, such offers were never mentioned in the actual diplomatic record. In other words, my data collection approach may under report the prevalence of economic linkage offers in failed alliance negotiations.[27] This could, in turn, impact estimates of the magnitude of the average treatment effect.

Under reporting the prevalence of economic issue linkage in failed negotiations means issue linkage is more closely associated with failures than what is reported in the data. Molinari (2010) offers an approach for estimating treatment effects in observational studies with missing treatment data. In the basic setup (details of which are in a supplemental packet that is available upon request) the sharp bounds on the classic treatment effect, $\Pr[y(1) = 1|d = 0] - \Pr[y(1) = 1|d = 0]$, are

$$-1 \leq \Pr[y(1) = 1|d = 0] - \Pr[y(0) = 1|d = 0] \leq 1 \tag{3.2}$$

which are not informative. In order to have stronger identifying power and create narrower bounds on the CTE, Molinari makes assumptions on the treatment response and on the treatment selection rule. Following Manski (1997), Molinari suggests assuming monotonicity in the response functions as a means of improving the identification of the CTE bounds. In particular, one can make the weak MTR (Monotonicity Treatment Response) assumption - the treatment has no negative/positive effect - and the weak MTS (Monotonicity Treatment Selection) assumption - if one divides the population into two groups according to the received treatment, then average outcome of the group without the treatment is less/more than the average outcome of the group with the treatment. The assumptions can be either positive or negative. It is important to note that 0 is the sharp upper bound on the treatment effect when weakly negative monotonicity is imposed and is the sharp lower bound on the treatment effect when weakly positive monotonicity is imposed.

[27] I assume that the presence of economic linkage offers in the successful negotiations is accurately reflected in the data. It is possible that a state may have refused an offer of a linkage provision, but since I am assuming that such linkages are proposed to seal a deal that would not otherwise be possible, such a scenario is not considered. Additionally, it may be the case that, for some of the negotiations that ended in agreement, a linkage offer was made and accepted, but this was not recorded in the final agreement. If this is the case, then I am understating the association of linkage offers with successful negotiations. This should bias against finding a positive effect for linkage offers. Therefore, this makes the fact that I have found a positive association even more convincing.

Given that matching identified a positive effect I will assume that, rather than flipping the sign from positive to negative, the presence of missing treatment data may reduce the size of the positive effect in these time periods. Thus, the key is to determine the extent to which the effect is reduced. Computing the bounds under weakly positive monotonicity produces

$$0 \leq \Pr[y(1) = 1] - \Pr[y(0) = 1] \leq 0.25$$

This indicates that, under weakly positive monotonicity, offering trade linkage can increase the probability of agreement by up to 25 percent percentage points. Thus, the effect can still be quite substantial. Though this robustness check suggests that the estimated average treatment effects slightly overstate the effect of trade linkage (36 percentage points, compared to 25 percentage points), recall that by "missing treatment" I suspect that I am under reporting the prevalence of economic linkage offers in failed negotiations. Hence, one should expect that allowing more trade linkage offers on the failed negotiations should reasonably result in producing a lower estimated effect.

3.7 Conclusion

Over the years, much ink has been spilt describing how issue linkages facilitate the conclusion of negotiated agreements. Theoretical work and numerous case studies have shown that it is a key mechanism by which states can secure negotiated agreements. However, these theoretical claims have not been subjected to direct and systematic analysis. Using data on alliance negotiations involving European states, I find that offering to include trade linkage in an alliance treaty increases the probability of agreement by approximately 36 percentage points from 1860 to 1945.

It appears that trade linkage offers can substantially increase the probability of agreement. However, this is only half the story. The other substantive claim regarding issue linkage is that it increases the probability that states will remain *committed* to the negotiated agreement. I test this claim in the next chapter.

Figure 3.1: Number of Military Alliance Negotiations, 1815 to 1945

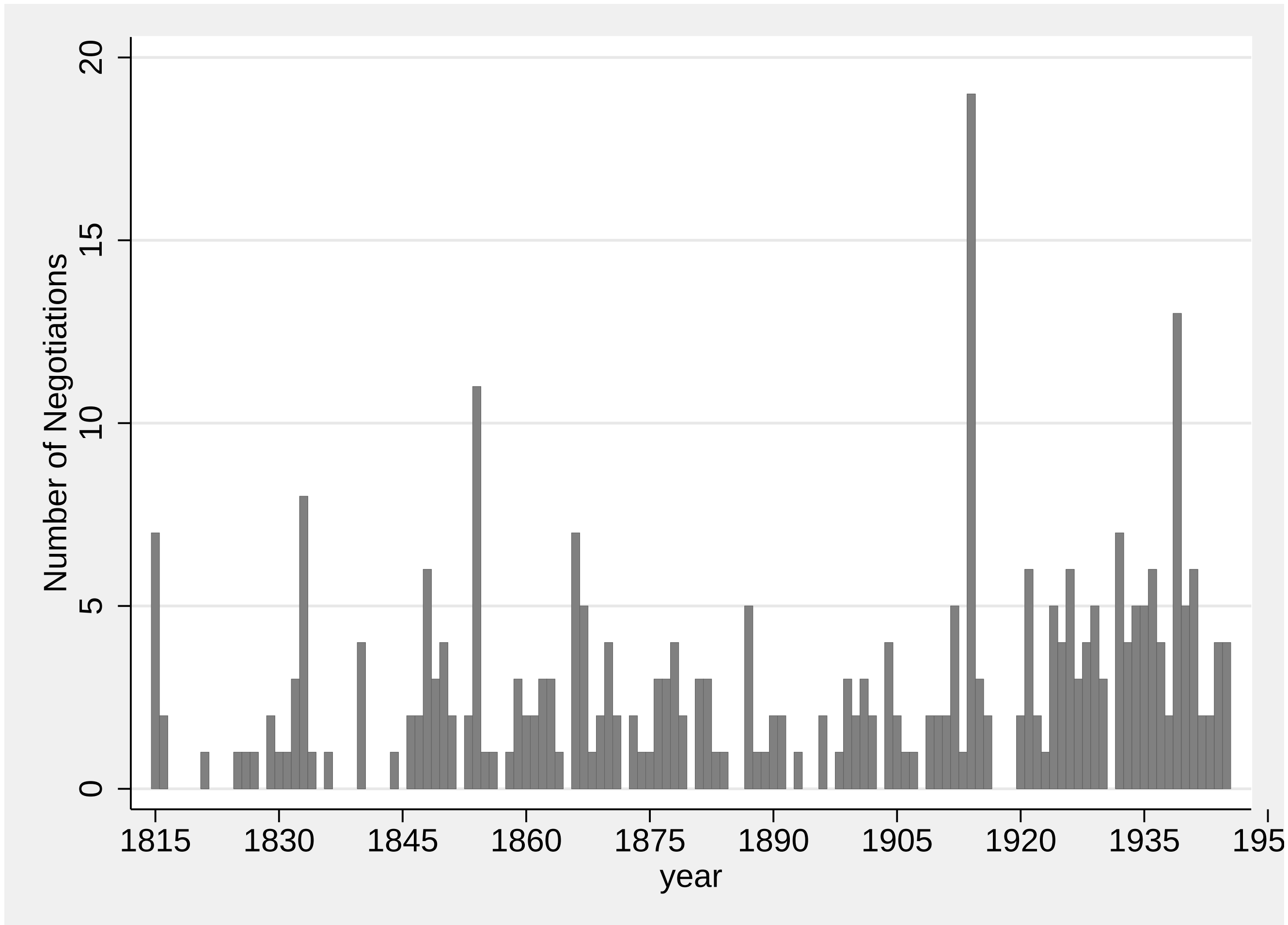


Table 3.1: Alliance Negotiations by Country (minimum 10), 1815 to 1945

Country Name	Number of Negotiations	Success Rate
Russia	112	0.54
Germany	102	0.41
Britain	88	0.45
France	87	0.51
Austria	69	0.52
Italy	45	0.75
Turkey	28	0.71
Serbia/Yugoslavia	20	0.85
Romania	19	0.74
Spain	13	0.77
Bulgaria	13	0.69
Poland	11	0.72
Greece	10	0.90
Japan	10	0.80

Table 3.2: Balance Summary Statistics p-values

	Difference in Means	Wilcoxon Rank Sum	Kimogorov-Smirnov
1860 to 1945			
Military Personnel	0.01	0.01	1.00
Contiguity	0.00	0.00	0.02
Number of States	0.06	0.11	0.87
Peak Year	0.44	0.43	1.00
Prior Negotiation	0.01	0.01	0.06
Buffer	0.02	0.02	0.21
Offensive/Defensive	0.00	0.00	0.00
Joint Democracy	0.91	0.47	0.73
1860 to 1945, offense or defense only			
Military Personnel	0.09	0.08	1.00
Contiguity	0.09	0.09	0.49
Number of States	0.06	0.28	0.99
Peak Year	0.78	0.78	1.00
Prior Negotiation	0.63	0.63	1.00
Buffer	0.04	0.04	0.25
Joint Democracy	0.78	0.95	1.00

Table 3.3: Effect of Trade Linkage on Alliance Formation, 1860 to 1945

Sample	Effect of Trade Linkage	Lower Bound .95 CI	Upper Bound .95 CI	Observations	N
Unrestricted	0.28	0.07	0.49	235	
Restricted*	0.36	0.11	0.61	140	

Match On: Military Size, Buffer, Contiguity, Crisis Period,
Joint Democracy, Offensive & Defensive alliances, prior negotiation
* Only Negotiations on Offensive or Defensive Alliances

Table 3.4: Rosenbaum Bounds Sensitivity Analysis

Gamma	Upper Bound p-value	Lower Bound p-value
1	4.4e-10	4.4e-10
1.1	5.0e-09	3.1e-11
1.2	3.7e-08	2.1e-12
1.3	2.0e-07	1.5e-13
1.4	8.70E-07	1.00E-14
1.5	3.10E-06	6.70E-16
1.6	9.10E-06	0
1.7	0.000024	0
1.8	0.000056	0
1.9	0.000118	0
2	0.000233	0
2.1	0.000428	0
2.2	0.000741	0
2.3	0.00122	0
2.4	0.001922	0
2.5	0.002909	0
2.6	0.004255	0
2.7	0.006032	0
2.8	0.008319	0
2.9	0.011193	0
3	0.014728	0
3.1	0.018994	0
3.2	0.024053	0
3.3	0.029961	0
3.4	0.03676	0
3.5	0.044487	0
3.6	0.053162	0
3.7	0.062799	0
3.8	0.073397	0
3.9	0.084946	0
4	0.097425	0
4.1	0.110806	0

CHAPTER IV

Does Issue Linkage Improve Treaty Credibility?

4.1 Introduction

The previous chapter offered the first systematic evidence that issue linkage offers can substantially increase the probability of states reaching a negotiated agreement. This chapter explores the second substantive question posed in the introduction: do linkage provisions actually enhance compliance with treaty obligations?

As mentioned in the introductory chapter, several studies have explored whether states comply with treaty obligations (Simmons 2000, Von Stein 2005). However, there is little direct and systematic empirical evidence that including linkage provisions in a treaty will bolster that treaty's credibility, eventhough this is a major theoretical claim of the issue linkage literature.[1] Only Leeds and Savun (2007), with respect to the inclusion of non-military cooperation provisions in alliance treaties, provide evidence that linkage is associated with a reduced risk of a treaty's terms being violated. However, the role of linkage was not the focus of their study and, as a result, Leeds and Savun do not test implications that could verify the causal impact of linkage on credibility. Similarly, Long and Leeds (2006) theorize that economic linkages in military alliance treaties can be used to clinch and maintain agreements, but they do not empirically test whether economic cooperation provisions fulfill either

[1]See also Simmons 2010 for a review of the empirical compliance literature.

of these roles.[2]

Therefore, building from the alliance data used by Long and Leeds (2006) and Leeds and Savun (2007), this chapter provides a precise test of the ability of linkage provisions to enhance the credibility of alliance commitments. Specifically, I use the alliance relations of *buffer states* as a 'hard case' for the ability of linkage provisions (specifically provisions calling on the states to engage in trade cooperation) to create credible alliance treaty commitments.[3] A buffer is a state located between two states that recently engaged one another in militarized conflict or view one another as hostile strategic rivals. A classic example of a buffer state was Poland during the 1700s, the 1920s and 1930s, and the Cold War. The alliance relations of buffer states create a hard case for treaty compliance because they are especially prone to invasion and occupation, thereby making other states reluctant to remain committed to an alliance agreement with the buffer state. According to Fazal (2007), in her extensive study on the propensity of buffer states to experience invasion and occupation, "states - especially threatened states - must balance to survive. But threatened states are unlikely to be able to balance precisely because they *are* threatened."[4] Thus, if linkage can enhance the credibility of commitments to buffer states, then linkage should be a useful tool for enhancing treaty credibility in nearly any context.

After reviewing how, in general, issue linkage can solve enforcement problems that reduce the credibility of international commitments (section 2), I detail how trade provisions in military alliance treaties can operationalize the use of issue linkage and how buffer states can operationalize the presence of an enforcement problem as it

[2]Powers (2004, 2006) and Powers and Goertz (2006) also explore the linkage of trade provisions and security arrangements, but in the reverse direction. Specifically, they observe that many regional economic institutions, particularly those exclusive to developing countries, have begun to more frequently incorporate explicit security cooperation provisions.

[3]The idea of a 'hard case' as expressed by Waltz, is to identify situations in which parties have strong reasons to behave contrary to the predictions of one's theory (Waltz 1979, p. 123).

[4]Fazal 2007, 230

relates to military alliances (section 3). Next, in section 4, I present four testable hypotheses regarding the alliance relations of buffer states that, if supported, will show that linkage provisions do enhance the credibility of alliance treaties. Testing these hypotheses in section 5 generates four key results. First, I find that buffer states in alliances with trade provisions experience fewer opportunistic violations of alliance terms by their alliance partners. Second, drawing from the theoretical expectations expressed by Morrow (1994), Smith (1995), and Fearon (1997) that credible alliances should more effectively deter attacks than non-credible alliances, I find that buffer states in alliances with trade provisions avoid occupation and invasion at a higher rate than buffer states in other alliance arrangements. Relatedly, I find that buffer states in alliances with trade provisions are attacked at a lower rate than buffer states in other alliance arrangements. Finally, I find that the likelihood a buffer state is going to be attacked in the first place has no statistically discernible impact on which buffer states are able to form an alliance with a trade provision.

Given these findings, section 6 explores the mechanism by which trade linkage provisions enhance alliance credibility. Specifically, the case of the Franco-Polish alliance of 1921 illustrates how the trade provisions are indeed the result of buffer states seeking to improve the credibility of the alliance. In this particular case, Poland used the promise of continued access to coal to motivate France to form a strong alliance. Section 7 concludes.

4.2 Issue Linkage and Credibility

4.2.1 Cooperation problems

The difficulties that states face when attempting to form an agreement are labeled *cooperation problems.* Koremenos, Lipson, and Snidal (2001), building from Oye (1979), identify the two most important cooperation problems (particularly when studying issue linkage) as the *distribution* and *enforcement* problems.[5] I will now describe each in turn.

The distribution problem (also called a bargaining problem) arises when actors have different preferences over alternative possible agreements. For instance, the benefits of an issue could accrue primarily to a few actors, while the costs fall disproportionately onto others. Theoretically, this is illustrated by the classic battle of the sexes game. This game contains two pure strategy equilibria: one equilibrium producing an outcome that favors one actor, another equilibrium with an outcome favoring the other. For example, state B and state A may agree to create a missile defense system, but disagree on the location of the interceptor missile launcher: states A and B may both want the launcher placed on their respective territories.

The enforcement problem arises when one state believes the other state is susceptible to reneging on the agreement. This will occur when actors find (current) unilateral noncooperation so enticing that they sacrifice long-term cooperation.[6] The enforcement problem is illustrated by the standard prisoner's dilemma game, where both actors know that the other has an incentive to defect from mutual cooperation, despite the fact that the resulting outcome from mutual cooperation is beneficial for society as a whole. For example, state A and state B may agree to develop a rapid reaction defense force and may agree on the number of troops to contribute to such a force. However, once the agreement is signed, state B may decide to free-ride on the contributions of state A by under contributing towards the force.

Fearon (1998) highlights how the two problems are linked and, therefore, both are capable of complicating the ability of states to reach an agreement. For instance, if

[5]Technically, Koremenos, Lipson, and Snidal identify three, but the first they identify - a large number of actors with divergent preferences - is nearly identical to the second - solving a distributional problem - as both deal with, using the framework of Sebinius 1983, expanding the winset of the negotiations.

[6]Koremenos, Lipson, and Snidal 2001, p. 776.

the two states take steps that will rectify the enforcement problem (thereby suggesting that the agreement will be longer lasting), then the states will bargain harder for a favorable allocation of the treaty's benefits. However, when studying treaty credibility, only the enforcement problem is relevant (if one side has an incentive to defect from the agreement, then how can actors claim that the agreement is credible?). Thus any study seeking to isolate the ability of particular treaty provisions to improve the credibility of an alliance must be sure to clearly isolate the enforcement problem. The research design section (section 3) will explain how buffer states can operationalize the existence of an enforcement problem as it relates to the creation of military alliances.

4.2.2 How issue linkage can help

When faced with an enforcement problem, states have a variety of tools that can enhance the credibility of the agreement. These tools range from restricting membership in the treaty, to increasing the centralization of the treaty by creating dispute settlement institutions or joint command structures. One of the tools most frequently highlighted by scholars is to expand the scope of the treaty by utilizing *issue linkage*, which Sebenius (1983) defines as the simultaneous discussion of two or more issues for joint settlement.[7] According to this literature, issue linkage is a key instrument that states can use to reach and maintain agreement. The basic idea is that if two sides cannot reach agreement when negotiating on one issue, adding a second issue increases the probability of agreement. For instance, if the states face a distribution problem, expanding the treaty negotiations along another dimension can redistribute the benefits, thereby allowing all participants to experience gain. Returning to the example of two states agreeing where to place a missile defense

[7] The theoretical literature on issue linkage includes, but is not limited to, those works cited in footnotes 2 and 3.

launcher, if the launcher is placed on state A's territory, a provision could be placed in the treaty calling on state A to compensate state B with military aid.

Alternatively, and most relevant for this paper, if the states face an enforcement problem, including a second issue in the treaty could incentivize all parties to remain committed to the final agreement. This is because the state with an incentive to defect on the primary issue will not wish to lose the stream of benefits generated by the linked issue. In the above example of states contributing troops to a rapid reaction force, tying a second issue - such as the offer of preferential trade access - to the maintenance of troop contributions could incentivize a state to remain committed to the overall agreement.

A classic example of using linkage to resolve a commitment problem is the 1987 Montreal Protocol on Substances that Deplete the Ozone Layer. Article 4 calls on all signatories to ban imports of controlled substances, products composed of controlled substances, and products made using controlled substances from nonsignatories and a ban on exports of the same products from signatories to nonsignatories.[8] The chief U.S. negotiator to the negotiations explained that

> "the objective of such restrictions was to stimulate as many nations as possible to participate in the protocol by preventing nonparticipating countries from enjoying competitive advantages and by discouraging the movement of CFC production facilities to such countries. These provisions were critical, since they constituted the only enforcement mechanism in the protocol."[9]

In short, one should expect linkage provisions to enhance the credibility of an agreement for states facing an enforcement problem. Of course, this expectation is

[8] See text of the Montreal Protocol (pp. 20 - 24), available at www.unep.org/OZONE/pdfs/Montreal-Protocol2000.pdf

[9] Benedick, 1991, p. 91.

very broad. Moreover, it contains two concepts that could be difficult to operationalize over a large number of cases: *the presence of issue linkage* and *the presence of an enforcement problem.* In fact, with respect to the presence of linkage, Koremenos, Lipson, and Snidal (2001) state that

> "One difficulty in analyzing [issue linkage] is that the issues themselves are not clearly defined. Does trade in all commodities constitute an issue? Or should we distinguish agricultural goods from manufactures? Although there is no general answer to this difficult task of assessing issue scope, focused empirical research can reveal the extent to which actors narrow or broaden the range of matters being addressed. *The problem is simplified when negotiations are expanded to cover items that could clearly be dealt with separately or were not previously linked.*"[10]

In the research design section (section 3), I will explain how trade provisions in alliance treaties can overcome the difficulty of identifying when issues are separable.

4.3 Research Design

Given the above discussion, empirically evaluating the ability of issue linkage to enhance the credibility of agreements requires operationalizing the enforcement problem and operationalizing when issue linkage has been employed. This section will begin by explaining how the inclusion of trade provisions in alliance treaties operationalizes the use of linkage in treaties. The section will then explain how the alliance relations of buffer states provide a means of operationalizing the enforcement problem.

[10]Koremenos, Lipson, and Snidal 2001, p. 771. Emphasis added.

4.3.1 Operationalizing issue linkage: Trade provisions in alliance treaties

When looking at a given treaty, it can be difficult to tell if the issues could have been addressed in separate negotiations. This is important, because if it is not the case that the issues could have been addressed in separate treaties, then it is unlikely that the issues were linked for the purpose of enforcing an agreement.

Data on military alliance treaties can overcome this difficulty. Specifically, thanks to the Alliance Treaties Obligations and Provisions (ATOP) database there exists a large amount of data on military alliance agreements. Most importantly, because ATOP codes the various provisions of each known military alliance treaty from 1815 to the present, we know that some alliance treaties contain explicit economic cooperation provisions. These provisions call for either the granting of foreign aid or for reducing trade restrictions between the parties (such as the granting of Most Favored Nation status). Specifically, fifty six (out of 648) military alliance treaties formed since 1815 include trade cooperation provisions. Trade provisions in alliance treaties are captured by the ECAID variable of ATOP. It is an ordinal variable taking on a value between 0 and 3: 0 if no such provision is found in the treaty; 1 if general or nonspecific obligations for economic aid are found in the treaty; 2 if one or more members promise economic support for postwar recovery; and 3 if one or more members commits to trade concessions, including the granting of MFN status.[11]

Alliances with economic cooperation provisions are not necessarily common, but they are also not especially rare. To give the reader a sense of the prevalence of economic provisions in military alliance treaties, consider Figure 4.1. Figure 4.1 shows, for various time periods, the ratio of alliances that contain an economic cooperation provision over all alliances created in that time period. Notice that, overall, alliances

[11] See the ATOP Codebook available at http://atop.rice.edu/.

with economic cooperation provisions comprise approximately nine percent of all recorded alliance treaties since 1815. Some decades, such as the 1940s, 1950s, and 1970s witnessed relatively widespread use of these provisions, while other decades, such as the 1930s, saw none.

Consider a few examples of these economic cooperation provisions.[12] Article 5 of the 1971 alliance between the Soviet Union and Egypt (ATOPID 3670) states, "...The parties shall expand trade and shipping between the two states on the basis of the principles of mutual advantage and most favored nation treatment." Article 9 of the 1946 mutual defense pact between the United Kingdom and Jordan (ATOPID 3040) proclaims that "Neither High Contracting Party will extend to the nationals or commerce of the other treatment less favorable in any respect than that which he accords to the nationals and commerce of the most favoured foreign country." Finally, article 3 in the 1953 alliance between Britain and Libya (ATOPID 3235) states, "In return for facilities provided by His Majesty The King of Libya for British armed forces in Libya on conditions to be agreed upon, Her Britannic Majesty will provide financial assistance to His Majesty The King of Libya, on terms to be agreed upon as aforesaid." Such provisions, especially those regarding trade, are of interest because, states quite frequently (if not primarily) negotiate alliance agreements and trade agreements separately from one other.

4.3.2 Operationalizing the enforcement problem: Buffer states

Buffer states are states located between two recently warring rivals or two states with hostile relations. Poland is a classic example of a buffer state as it was located between Austria, Prussia, and Russia during the 1700's, between Russia and Ger-

[12]Each example below is taken from the answer to question 55 of the ATOP code sheets for the respective treaties. Question 55 reads, "55. Does the treaty include provisions for economic aid or other enticements (include trade concessions, post war recovery, etc.)? (Yes, No) If Yes, describe these provisions." These code sheets are available at http://atop.rice.edu/.

many during the 1920's, and between the Soviet Union and Western Europe during the Cold War. Fazal (2007) codes a state as a buffer when it is geographically located between two states that are engaged in an interstate rivalry, where the identification of a rivalries is drawn primarily from Goertz and Diehl (1992) and Bennett (1997) (along with the imperial rivalry dyads of the UK-Russia and UK-France). Both Goertz and Diehl and Bennett use a Militarized Interstate Dispute (MID) density approach for identifying rivals. A MID is a historical case of conflict in which one state (a sender) threatens, displays, initiates military force short of war, or initiates war against another state (a target). According to Diehl and Goertz (Bennett), this approach means two states are rivals if they have engaged one another in at least 6 (5) MIDS against one another in the previous 20 (5) years. Data on buffer state identification was provided by Fazal upon request and a complete list of Fazal's buffer states is found on table B.1 of Fazal (2007).

Buffer states and foreign invasion

What makes buffer states particularly interesting is that, according to Fazal, buffer states are especially prone to violent *state death*, which Fazal defines as "the formal loss of foreign policy control *to another state*" via military invasion.[13] From 1816 to 1992, Fazal identifies 50 cases of violent foreign occupation and, using survival analysis, Fazal finds strong empirical evidence that buffer states experience violent foreign conquest at a rate disproportionately higher than other states.[14]

Why do buffer states experience foreign invasion and occupation at a higher rate than other states? Fazal argues that the rivals on either side of the buffer state fear that its opponent will conquer the buffer state, thereby gaining a strategic advantage. Though maintaining the sovereignty of the buffer state is ideal for both rivals (as

[13] Fazal 2007, 17

[14] A complete list of these state deaths are found on pages 21 - 23 of Fazal (2007).

it creates a barrier between the rivals that decreases the probability of war), both rivals know the other has an incentive to invade the buffer and gain the strategic advantage. This commitment problem leads inevitably to the demise of the buffer state. However, Fazal also identifies a sharp decline in violent state death (invasion and occupation) since 1945. Of the 50 instances of violent state death identified by Fazal from 1816 to 1992, just two occurred after 1945: South Vietnam in 1975 and Kuwait in 1990.[15] According to Fazal, this decline in violent state death is explained by the diffusion of a norm against non-conquest after World War II, as embodied in Article 2.4 of the United Nations Charter; "All members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state."[16] An alternative explanation is the cessation of wars between the great powers, which is attributed primarily to the creation and continued existence of nuclear weapons.[17] The decline of major power war explanation is especially plausible given that great powers are those states most capable of "killing" other states. The post-1945 decline in major power war has been noted by numerous scholars, though its permanency is a source of continued debate. Scholars such as Mueller (2004) argue that great power war has become inconceivable, while Mearsheimer (2001) argues that great power war is simply in a historic lull.

It is beyond the scope of this paper to determine if the decline in violent invasion and occupation is due to a rising non-conquest norm, a cessation of great power war, a combination of the two, or some third factor. Instead, the virtual end of violent foreign invasion and occupation of buffer states after 1945 means any study that

[15] A complete list of these state deaths are found in Fazal 2007, pp. 21 - 23.

[16] Quoted from Fazal 2007, 170.

[17] For scholars supportive of the view that nuclear weapons explain the decline in great power war, see among others, Jervis 1988 and Gaddis 1992 and 1999. For a counter argument, see Mueller 2004.

considers the need of buffer states to deter foreign military invasion and occupation should focus primarily on the pre-1945 time period (when the threat of invasion appears most salient).

Buffer states and alliances

That buffer states are especially prone to foreign invasion and occupation means buffer states have a great incentive to form military alliances. However, buffer states have difficulty inducing states to form alliances with them. For example, Persia attempted on a number of occasions to form an alliance with Britain. However, Persia was rebuffed each time, as the British were unwilling to commit themselves to its defense.[18] Another example is Hungary in 1849. After declaring independence from the Hapsburg Empire, Hungary faced the threat of invasion by Russia. Hungary appealed for assistance from Britain, but Britain rejected the offer.[19] This was due, in large part, to Russia already having troops stationed and mobilized in neighboring Transylvania. After the British refusal, Hungary was invaded and occupied by Russia. For these reasons, Fazal argues that buffer states face a *catch-22* situation: "States - especially threatened states - must balance to survive. But threatened states are unlikely to be able to balance precisely because they *are* threatened."[20] The result is that "the politics of location and timing hamstring state's strategies to *cheat death*."[21]

Given the reluctance of other states to form alliances with buffer states, one should expect that those buffer states fortunate enough to secure alliance partners will face difficulties ensuring that their partners remain committed to the alliance. For example, Poland (the classic buffer state) formed an alliance with Prussia in 1790, but, in

[18] Fazal 2007, 129.
[19] Albrecht-Carrie 1958, p. 73
[20] Fazal 2007, 230
[21] Fazal, 2007, p. 149. Emphasis added.

response to an appeal to protect Poland from Russian advances, the ambassador to Frederick William of Prussia notified Poland's King Stanislas in January 1792, "My master does not consider himself bound by the treaty of 1790 to defend his army the hereditary monarchy, as established by the Constitution of May 3, 1791"[22]

In addition to this example, consider Table 4.1. Table 4.1 is created using Fazal's identification of a buffer state (see above) and the *Term* variable of the ATOP dataset. The *Term* variable categorizes the way an alliance ends. For alliances that remain in effect as of December 31, 2003, and for alliances that are censored due to the loss of independence of one or more alliance members in action unrelated to the alliance, *Term*=0. If the alliance ends when its provisions are fulfilled (either because its specified term has been completed or the goals of the alliance have been accomplished), *Term*=1. If the alliance ends due to violation of provisions by one or more members, including willful abrogation before the scheduled termination date, *Term*=2. If the alliance ends when some or all of the members negotiate a new relationship, *Term*=3.[23] Table 4.1 shows two groups of alliances: those in which at least one member is a buffer state and those in which no members are buffer states. It then reports the percentage of alliances in each group in which *Term*=2 (i.e., one of the members committed a willful violation of the treaty's terms). One should immediately notice that alliances with buffer states experience violations of the alliance treaty at a substantially higher rate compared to alliances without buffer states (52 percent compared to 41 percent). Moreover, a two-sided t-test shows that this rate of violation has a p-value of 0.07. This simple test suggests that alliances with buffer states do indeed face a significantly higher risk of alliance treaty violation.

This inability of buffer states to form credible alliance commitments makes them

[22]Quoted from Fazal 2007, 115

[23]Description of the *Term* variable is drawn from the ATOP codebook, p. 18.

ideal for studying the ability of linkage provisions to enhance alliance credibility: if issue linkage can enhance the credibility of alliance commitments for these particularly vulnerable states, then linkage should work for any state.

4.4 Specific Testable Hypotheses

If linkage creates incentives for states to remain committed to an alliance agreement, then what empirical regularities should one expect to observe? Having described how the alliance relations of buffer states can operationalize the enforcement problem and how trade provisions in alliance treaties can operationalize the use of issue linkage, it is now possible to identify and test more specific hypotheses regarding the influence of issue linkage on treaty credibility. In this section, I will describe four such implications: (1) buffer states should experience reduced willful violations of treaty terms when in treaties with trade provisions, (2) buffer states should experience enhanced deterrence in treaties with trade provisions, (3) buffer states should experience lower rates of invasion when in alliances with trade provisions, and (4) alliances with trade provisions are not formed by buffer states that were less likely to be attacked and invaded in the first place.

First, it should be the case that if trade provisions enhance the credibility of the alliances of which buffer states are party, it should be the case that the presence of a linkage provision makes leaders less likely to willfully break international agreements in violation of their negotiated terms. Stated differently, buffer states in alliances with trade provisions should experience fewer violations of alliance agreement provisions by alliance partners compared to buffer states in alternative alliance arrangements. This can be stated as hypothesis 1

Hypothesis 1: *Buffer states in alliances with trade provisions should experience alliance treaty provision violations at a lower rate than buffer states in alliances without trade provisions or buffer states in no alliances.*

Second, a primary function of an alliance is to deter attacks. In the broader literature on conventional and nuclear deterrence, the credibility of a promised response to attack (either a promise by the target of the attack or a promise of assistance by a third party defender) is fundamental to determining if deterrence will succeed.[24] According to Smith (1995), the more that third parties perceive an alliance as credible, the more this should increase the deterrence capability of the alliance. Morrow (1994) and Fearon (1997) argue that the perceived credibility of an alliance can be increased if states engage in costly peace-time actions that demonstrate their commitment to one another and their willingness to defend one another. Such costs are typically conceived as prewar military coordination, which can vary from engaging in joint planning and capability coordination, to basing troops on an ally's territory. Such coordination is credible because it is a costly signal (e.g., coordination between allies will limit the flexibility of each ally in the event of a war) of a state's willingness to support its allies.

Trade provisions in alliance treaties can also produce costly peace time signals and, therefore, deter attacks. Since the trade provision is placed in the alliance agreement, the legal basis for the alliance can not be nullified if a party defects from the trade agreement, but the legal basis of the trade agreement is nullified if one of the states defects from the alliance. In other words, the basis for trade, the trade agreement, is made explicitly contingent on adherence to the alliance agreement (and not the other way around). Hence, if the states want to maintain the trade relationship,

[24]For the classic works dealing with the issue of credibility in deterrence, see Shelling 1960; Maxwell 1968; Jervis 1970; and George and Smoke 1974. For an alternative view, see Mearsheimer 1983.

they will have to uphold the alliance commitment. In this way, the trade provisions serve as a strong form of what Fearon (1997) calls "hand tying": signals that create audience costs that the leader of a state will suffer due to the reaction of domestic political audiences to a perceived failure in the management of foreign policy. Thus, trade provisions should add a layer of credibility to the alliance which will, in turn, deter threats. This leads to hypothesis 2:

Hypothesis 2: *Buffer states in alliances with trade provisions should experience fewer attacks than buffer states in alliances without trade provisions or buffer states in no alliances.*

Third, though testing hypothesis 2 is useful, the primary goal of forming an alliance by a buffer state is to not just reduce the rate of attack, but to make sure that, for those attacks that do take place, the attacks do not result in state death. Thus, it should also be the case that the trade provisions enhance the credibility of the alliance to such an extent that the surrounding rivals are less likely to invade and occupy the buffer state. This is perhaps the most difficult test of the ability of the linkage provision to enhance the credibility of the agreement. Stated as a hypothesis:

Hypothesis 3: *Buffer states in alliances with trade provisions should experience foreign invasion and occupation at a lower rate than buffer states in alliances without trade provisions or buffer states in no alliances.*

Fourth, as famously argued by Downs, Rocke, and Barsoom (1996), states may only sign agreements they think are going to work. Thus, perhaps the fact that a buffer state is able to form an alliance with a trade provision is because this buffer state, compared to other buffer states, was unlikely to be attacked in the first place. This could possibly also explain any findings of enhanced deterrence (hypothesis

2) or survival (hypothesis 3). Therefore, if the trade provisions do have an actual effect on enhancing the credibility of the alliance, then it should be the case that the likelihood of buffer state survival should not have a different impact on the probability of a buffer state forming an alliance with a trade provision compared to forming an alliance without a trade provision. Stated as a hypothesis

Hypothesis 4: *When entering the alliance, the likelihood of attack and invasion should not be lower for buffer states that form alliances with trade provisions compared to buffer states that form alliances without trade provisions.*

The next section will present the procedures for testing each of these hypotheses and the results from those tests.

4.5 Empirical Analysis

4.5.1 Test One: Buffer States and Opportunistic Violations

Recall the first hypothesis: buffer states in alliances with trade provisions should experience alliance treaty provision violations at a lower rate than buffer states in alliances without trade provisions or buffer states in no alliances.

Consider Table 4.2. It again uses the *Term* variable of the ATOP dataset to capture if the alliance ends due to violation of provisions by one or more members, including willful abrogation before the scheduled termination date. It also uses the ATOP *ECAID* variable to identify when an alliance contains a trade cooperation provision. In this table, the unit of observation is the buffer-state-year prior to 1945. Thus, table 4.2 reports the percentage of buffer-state-years that witnessed treaty violations in two alliance arrangements: for alliances that contained no trade cooperation provisions and alliances that contained trade cooperation provisions.

Table 2 shows quite clearly that no year in which a buffer state was a member of an alliance with a trade provision witnessed a willful violation of the alliance's terms. In contrast, 17 percent of the years in which a buffer state was a member of an alliance with no trade provisions witnessed a willful violation of the alliance's terms. The difference in these percentages is statistically significant at above the 0.99 confidence level of a simple two-sided t-test. Moreover, due to the lack of violations for buffer states in alliances with trade provisions, there is no need (nor is it possible) to conduct regression based survival analysis using, for example, a Cox proportional hazard model.

Therefore, the simple comparison of percentages in Table 4.2 offers rather strong support for the claim that the trade linkage provisions enhanced the credibility of the alliances. Relatedly, consider Table 4.3. Fewer willful violations should be associated with longer lasting alliances (assuming that alliances with trade provisions are not, on average, written with an exceptionally short duration clauses). Table 4.3 shows that, over the pre-1945 time period, the average amount of time a buffer state was in an alliance with a trade provision is 21 years, which is substantially higher than the 13 years for buffer states in alliances without trade provisions. Together the evidence provided by Tables 4.2 and 4.3 appear to support hypothesis 1: buffer states in alliances with trade provisions experience fewer willful violations of the treaty terms.

4.5.2 Test Two: Buffer States and Attacks

Recall hypothesis 2: buffer states in alliances with trade provisions should experience fewer attacks than buffer states in alliances without trade provisions or buffer states in no alliances. Testing if buffer states in trade alliances do indeed experience fewer attacks than buffer states in alternative alliance arrangements requires operationalizing attacks on buffer states. To operationalize when a state is a buffer state, I again use the identification of buffer states found in Fazal (2007) as described above. I operationalize attacks using MID data for the 1815 to 1945 time period (Ghosen, Palmer, and Bremer 2004).

It is important to clarify when the onset of a MID does and does not test my argument. Suppose a buffer state, having formed an alliance, becomes emboldened and, consequently, initiates a MID against a target state. If that target state subsequently initiates a MID against that buffer state, this is not a disconfirming case. Instead, what I really wish to determine is if trade-alliances deter *unprovoked* MIDs. Therefore, for all state-years in which a buffer state is part of an alliance with a trade provision, part of an alliance with no trade provision, or part of no alliance, I will count the number of state-years that witnessed the onset of an MID against a buffer state in which the buffer state did *not* initiate an MID against the sender in the previous 5 years.[25] In other words, I code the number of buffer-state-years that witnessed an unprovoked MID and then use this to compute three measures: the percentage of buffer state-years in an alliance with a trade provision that witnessed an unprovoked attack; the percentage of buffer state-years in an alliance with no trade provision that witnessed an unprovoked attack; the percentage of buffer state-years in no alliance that witnessed an unprovoked attack. Of course, alliances in the second category could include scope expansion provisions on issues other than trade. Though I do not directly account for this possibility, not doing so will bias my results towards the null of finding no difference in the rates of attack between trade alliances and nontrade alliances.

The findings from this test are summarized in Figure 4.2. Though alliances with trade provisions do not perfectly deter unprovoked MIDs (36 percent of buffer state-

[25]The following results are robust to lower and higher number of years.

years in alliances with trade provisions witnessed the onset of an unprovoked MID), they appear to do so much more effectively than alliances with no trade provisions (55 percent of buffer state-years in alliances with no trade provisions witnessed the onset of an unprovoked MID) or no alliance (66 percent of buffer state-years in no alliances witnessed the onset of an unprovoked MID).[26] These results support hypothesis 2: buffer states in alliances with trade provisions are attacked at a much lower rate than buffer states in other alliance arrangements (alliances without trade provisions or no alliance).

4.5.3 Test Three: Buffer States and Survival

If trade provisions enhance the ability of alliances to deter threats, then buffer states in alliances with trade provisions should be able to avoid invasion and occupation at a higher rate compared to buffer states in other alliance arrangements. To test this hypothesis, I operationalize invasion and occupation using the Fazal (2007) coding of violent foreign takeover (see above). The longer a buffer state avoids violent foreign takeover, the longer that state "survives." Figure 4.3 shows, for the 1815 to 1945 time period, the Kaplan-Meier survival estimates of buffer states that are part of an alliance with a trade provision, part of an alliance with no trade provision, and part of no alliance. Survival rates are computed as

$$\text{Survival Rate at time } t = \frac{\text{Number of buffer states surviving at time } t}{\text{Number of buffer states entering time } t} \tag{4.1}$$

Thus, if 4 cases enter time t, but only 3 survived during time t, then the survival rate at time t is 0.75. If the overall (or cumulative) survival rate entering time t is 100 percent (i.e. no deaths occurred in any prior period), then the new cumulative

[26]Figure 2 also shows that the distribution of the "no alliance" and the "nontrade alliance" groups is statistically distinguishable at or around the .99 confidence level.

survival rate at the end of period t is $1.00 \times 0.75 = 0.75$. Figure 4.3 shows that, by and large, buffer states in alliances with trade provisions have a higher survival rate than buffer states in alliances with no trade provisions and buffer states in no alliance.

In fact, for the majority of the graph, buffer states in alliances with trade provisions simply do not experience foreign invasion and occupation. For example, consider a state that has been a buffer for 75 years. According to Figure 4.3, such a buffer state would have a 40 percent survival rate if a member of no alliance (10 observations entering the 75th year), a 60 percent survival rate if a member of an alliance without a trade provision (3 observations entering the 75th year), and a 100 percent survival rate if a member of an alliance with a trade provision (5 observations entering the 75th year). Eventually, one buffer state in an alliance with a trade provision does die (just after 100 years of existence). This one instance is Austria-Hungary in 1918, which, along with Germany and Italy, formed a defensive alliance that included a trade provision in 1882 (atopid number 1350). In a sense, this is an exception that supports my argument: a buffer state that remained a member of an alliance with a trade provision for nearly 36 years until it died at the end of the second most devastating war in human history! After this one failure, the survival rate of buffer states in alliances with trade provisions falls to just above that of alliances without trade provisions. Thus, the results suggest that, by and large, buffer states that join alliances with trade provisions survive at a much higher rate than buffer states outside trade alliances.[27]

[27]Though one instance of failure means I can compare standard errors on a Cox proportional hazard model, computing a Cox model with only one instance of failure seems unnecessary. Instead, one should simply study the circumstances that contributed to this single case of failure.

4.5.4 Test Four: Buffer states and invasion risk at time of formation

Recall the fourth hypothesis: when entering the alliance, the likelihood of attack and invasion should not be lower for buffer states that form alliances with trade provisions compared to buffer states that form alliances without trade provisions. In light of my findings for hypotheses 2 and 3, this is a particularly important hypothesis to test as it can suggest that the observed enhanced deterrence ability of trade alliances is not due to outside states choosing to include trade provisions in alliances with buffer states that were less likely to be attacked and invaded in the first place. I will now describe my research design for testing this final hypothesis.

The dependent variable

My dependent variable in this test is *Alliance type*, which equals 0 if a buffer state is not part of an alliance in year t, equals 1 if a buffer state is part of an alliance with no trade provision in year t and not part of an alliance with a trade provision in year t, and equals 2 if a buffer state is part of an alliance with a trade provision in year t.

The independent variable

I must operationalize the likelihood that a buffer state is about to be attacked and/or invaded. Given Fazal's theory for why buffer states are susceptible to invasion and occupation (i.e. the rivals on either side invade the buffer in order to gain a strategic advantage against the other rival in a conflict), I will capture the likelihood of a buffer state being attacked using a variable that can proxy for the probability of the surrounding rival states becoming involved in war.

When will rivals be unlikely to renew hostilities? Following Fazal, I use the power asymmetry of the rivals to measure the probability of the rivals going to war. Fazal argues that asymmetry between rivals leads to stability and peace, but symmetry between rivals leads to instability and war. Whether symmetry leads to conflict because of uncertainty in the outcome of a war(Fearon 1995; Blainey, 1988) or because the rising state wishes to speed up the process of surpassing the established hegemon (Organski and Kugler 1980; Bueno de Mesquita and Lalman 1992) is beyond the scope of this paper. Instead, it is sufficient simply to recognize that buffer states will be particularly likely to die when its surrounding rivals are relatively equal in power.[28] This also means that if the rivals are relatively asymmetric in capabilities, the buffer state faces a low risk of violent state death.

The variable *Rivals' Asymmetry* will be used to capture the power asymmetry of the rivals surrounding each buffer state. It is computed by calculating, for each rivalry, the proportion of iron and steel production, energy consumption, total population, urban population, military expenditure, and military personnel held by each rival.[29] The average value of these six shares becomes each rival's *rival capability* score. Finally, the highest *rival capability* score is subtracted from the lowest *rival capability* score and this result is divided by 2. Thus, the *Rivals' Asymmetry* variable is bounded between 0 (complete symmetry) and 0.5 (complete asymmetry). I then compute the average *Rivals' Asymmetry* score for the buffer states in each alliance.

Control variables

When conducting this test, I also account for factors that are correlated with the probability of a buffer state forming an alliance and with the probability that the surrounding rivals will invade the buffer state.

First, the underlying value of the tradable good offered by the buffer state is

[28]Fazal (2007), p. 92

[29]In other words, this variable is computed just like the Correlates of War CINC score, except the "international system" is just the two states in the rivalry

compelling the other states to form an alliance with the buffer state and is making the buffer more strategically important to the surrounding rivals. How can one determine, over a large number of cases, when one state will value trading with another state? One option is to adopt a broad measure of "trade attractiveness." An example of a broad measure is economic "gravity," which is commonly used in studies of international trade.[30] According to the gravity model of international trade, bilateral trade flows are based on the economic size and distance between two units. However, there are major disadvantages to using gravity in the present study. First and foremost, the size and distance of two countries, the principle components of economic gravity, are also critical to decisions of security. For example, countries with large economies (either population or land mass) have more to protect. This could indicate a greater need for an alliance.[31] Second, distance, as mentioned above, is correlated with a country's decision to form an alliance. In short, gravity is a broad measure that can not distinguish isolate when a state will have a tradable good that will be of value to another state from the need to form an alliance.

Therefore, I choose to instead use a more precise measure of when one country will value trading with another: a state's coal production. Using coal is reasonable for two reasons. First, though coal is still a critical resource today, it was absolutely vital to many European nations' economic vitality prior to and immediately after World War II (for instance, the precursor to the European Union was the European Coal and Steel Community). Second, coal deposits prior to 1945, much like oil today, provides a rather clear and direct indicator of the extent to which one state has a tradable good that will be of value to another state.

[30] This concept of economic gravity was developed independently by Tinbergen 1962 and Pöyhönen 1963. An excellent recent example where economic gravity is used to predict trade between states is Frankel and Rose 2005.

[31] Of course, more mass could also mean a greater ability to internally balance, hence less need for an alliance. Either way, a country's size (especially relative size) would be correlated with alliance formation, not just trade.

Focusing on coal has two limitations, though neither is of major concern. First, using a narrow indicator such as coal production will constrain the domain of my analysis. Other countries may be valuable trade partners for reasons other than their coal deposits. However, as already mentioned, coal has the advantage of allowing me (and the reader) to know exactly what a variable represents (which is not the case with gravity). Second, focusing on coal constrains my analysis to European states. With the exception of a few non-European countries (such as the United States and Canada), coal production data is most accurate for the European nations during this time period.[32] The data on coal production is drawn from B.R. Mitchell's *International Historical Statistics: 1750 - 1993* (Mitchell 1998). I use this volume of statistics to code the coal production output (in thousands of metric tons) for twenty European countries. Table 4.4 reports the countries and years for which I have coal output data. I then create the variable *Coal*, which is the average recorded coal production of the buffer states in the alliance.

Second, I control for the military capabilities of the buffer state relative to the surrounding rivals. This variable was created by Fazal and it captures the ratio of the buffer states military personnel over the combined capabilities of the buffer and the two rivals. The larger the value of this variable, the less likely are the rivals to invade the buffer and the less incentive a buffer has to seek alliance partners.

Finally, I control for the buffer state being a member of an existing alliance.

[32] Of course, measurements of coal production, even in the European context, are still suspect, especially for the early 20th century and the nineteenth century. Most statistics prior to the mid twentieth century were by-products of taxation or military preparedness and, as a result, some countries would inflate their population or wealth figures in order to impress potential enemies. Additionally, government statistical services, change the detail of coverage and of concepts from one yearbook to the next. This can lead to breaks in the time series (missing data) and/or changes in the unit of measurement.

Analysis method and results

Because my dependent variable has multiple unordered categories, I estimate the model using a multinomial logit. A multinomial logit estimates the probability that the actual outcome Y will take on each of a set of discrete possible outcomes given a vector of independent variables X. Given $J + 1$ outcomes, a multinomial logit estimates J equations which show the effects of the variables on the likelihood of a particular outcome occurring. Estimates are relative to a base category, which in this case is *Alliance type*=0.

Using the buffer-state-year as the unit of analysis, I regress *Alliance type* on these variables. The results are reported in Table 4.5. One can see that *Rivals' Asymmetry* has a positive and statistically significant effect on the probability of a buffer state being part of an alliance with no trade provision in a given year. Thus, the less likely invasion becomes (i.e. the more asymmetric the rivals), the more likely the buffer state is to form an alliance without a trade provision. However, *Rivals' Asymmetry* has no statistically discernible impact on the probability of a state forming an alliance with a trade provision. This suggests that the probability of conflict between the rivals surrounding the buffer state does not impact the probability of buffer states forming alliances with trade provisions. When combined with the fact that as the surrounding rivals become less likely to attack one another (thereby reducing the likelihood of a buffer being attacked) buffer states are *more* likely to be members of alliances without trade provisions, this lends even stronger support to the above results concerning the deterrence effect of trade provisions.

4.6 Identifying the mechanism

The above quantitative analysis supports my hypotheses. However, I have not yet teased out the exact role played by the trade provisions. Consequently, my results might be due to spurious correlations. Therefore, I turn to a case study so as to more precisely identify the mechanisms. I chose the case of the Franco-Polish alliance of 1921. I chose this case because (1) it is an alliance that contains a trade provision and (2) Poland, as highlighted earlier, epitomizes a buffer state and this was the only alliance formed by Poland after it was reconstituted as a state following World War I. When exploring this case, I need to identify evidence that supports the notion that the inclusion of a trade linkage provision enhanced the credibility of the alliance and, therefore, enabled Poland to survive.

When Poland reemerged as an independent state after World War I, Poland was threatened by Germany to the West and the Soviet Union to the East, with whom it fought a war in 1920.[33] Thus, Poland sought an alliance with one of the other major powers. France proved to be the most suitable (and perhaps only) candidate, as the United States had withdrawn into isolation and the British sought to avoid too specific of commitments during peacetime.[34] The Polish minister of foreign affairs in 1920, Prince Estachy Saphieha, approached his French counter part about the alliance, but initially "found the Quai d'Orsay very reserved."[35] The sentiment of French reluctance to form a binding alliance agreement was expressed well in French diplomatic correspondence; "'[Poland had] neither frontiers, nor a government, nor an army."[36] Moreover, the French were all too aware that, in particular, its Eastern

[33] Wandycz 1962, p. 211.

[34] see Albrecht-Carrie 1958, 406 - 411 for a discussion of the interwar "French system" of alliances directed toward balancing German (and Russia) power.

[35] Quote found in Wandycz 1962, p. 213.

[36] Quote found in Wandycz 1962, p. 214.

frontiers were liable to Russian invasion.[37]

This is not to say that France had no desire for military cooperation with Poland. Both countries had a common interest in containing Germany. Moreover, France had assisted Poland in its war against Russia in 1920.[38] Even the French foreign ministry expressed an interest and desire to assist Poland with strengthening and enlarging its military.[39] However, French diplomats were opposed to forming an agreement with Poland that would obligate it to use French forces to protect Poland if attacked by Germany or Russia.[40]

However, the Quai d'Orsay [French Foreign Ministry] eventually agreed to complete an agreement in which France would "provide mutual aid in case of German aggression," but only on one condition: the alliance included commercial provisions.[41] To secure French protection, Polish officials had to include a commercial provision in the alliance treaty that granted France most favored nation trading status and gave France access to coal from the Upper Silesia region acquired by Poland from Germany as part of post-World War I German reparations. This coal was particularly important to France, as French industry was highly dependent on imports of British coal (which enabled Britain to exert influence on France). As stated in a correspondence between Polish diplomats in Warsaw and Washington, D.C. in October 1920;

> "if Poland obtains Silesia her influence on France will increase because of the coal which will allow France to become more independent of the imports of English coal. England, because of her coal can [currently] exercise

[37]Wandycz 1962, p. 214.
[38]Albrecht Carrie 1958, pp. 408.
[39]Wandycz 1962, p. 218.
[40]Ibid.
[41]Ibid. pp. 217 and 219.

political pressure on France and Italy."[42]

For its part, France lowered tariffs (but did not grant MFN status) on fifty Polish items, mainly raw materials and agricultural products. This led the *rapporteur* in the Warsaw Parliament assigned to investigate the treaty before recommending final approval to comment that the commercial component "gives to France more than France gives to us," but such discrepancies were acceptable because of the "advantages which Poland obtained in another field...[namely] a close alliance with France."[43]

Despite this evidence, one could counter the claim that the trade provision created a more credible alliance with France. First, one could argue that even though Poland was a buffer state, the two of the rivals surrounding Poland (Germany and Russia) were severely weak: Germany was under the Treaty of Versailles' constraints, while Russia was still rebuilding after internal collapse led to the Bolshevik Revolution. The other rival, the United Kingdom, was in the process of a major demilitarization as dictated by the now infamous "10 Year Rule" (whereby the armed forces' estimates should be based on the assumption that "the British Empire would not be engaged in any great war during the next ten years.").[44] Hence, Poland was in no *imminent* danger of experiencing an invasion that would lead to its demise. However, this counter argument seems unlikely to explain why France was willing to form the alliance with Poland, as Poland had been attacked by Russia the previous year (thereby demonstrating that Poland still lived in a dangerous neighborhood).

Second, one could point out, quite accurately, that Poland was invaded by Germany in 1939 and subsequently divided between Germany and the Soviet Union.

[42] Quoted found in Wandycz 1962, p. 212.
[43] Quote found in Wandycz 1962, p. 221
[44] for more on the ten year rule, see, for example, Kennedy, 2006, p. 273.

However, if one considers the views of diplomatic historians, the fact that Poland was only *eventually* invaded and partitioned, as opposed to *immediately* invaded and partitioned, suggests that the alliance with France was highly credible. As the historian Steiner states, "Given its geographic situation, between Germany and a still unsettled and unrecognized Bolshevik Russia, Poland's future was bound to be fraught with difficulties...Given the circumstances of its birth, it is almost surprising that the boundaries of the Polish state survived intact until the Fourth Partition in 1939."[45] Hence, it is very likely that without the trade alliance with France, Poland would have been invaded and partitioned *sooner* than 1939. In fact, Albrecht Carrie suggests that one of the goals of Germany forming a non-aggression pact with Poland in 1934 was precisely to weaken the credibility of the Franco-Polish pact.[46]

4.7 Conclusion

Does issue linkage enhance the credibility of treaty commitments? With respect to trade provisions in alliance treaties, the answer appears to be 'yes'. Since buffer states are highly susceptible to foreign invasion and occupation (and, hence, other states are reluctant to honor alliance commitments with the buffer state), buffer states in military alliances are an ideal 'hard case' for the ability of linkage provisions to incentivize states to adhere to their alliance commitments. I find that buffer states in alliances with trade provisions experience fewer willful violations of alliance obligations than buffer states in alliances without trade provisions, experience fewer attacks than buffer states in alliances without trade provisions, and almost never experience invasion and occupation. The case of the 1921 Franco-Polish alliance illustrates how the inclusion of trade provisions in alliance treaties are indeed the

[45] Steiner 2005, 52
[46] Albrecht Carrie 1958, pp. 469-470.

result of buffer states seeking to improve the credibility of the alliance.

I have now addressed the second of the two substantive questions posed in the introduction. Therefore, the next section will conclude with a discussion of these finding's importance and possible directions for future research.

Table 4.1: Buffer States and Alliance Commitment, 1815-1945

Group	Number of alliances	Percentage that ended due to violations
Alliances with no buffer state	113	41 percent
Alliances with at least one buffer state	132	52 percent

Difference statistically significant at $p<0.01$ (two-tailed).

Table 4.2: Buffer States, Trade Provisions, and Willful Violations, 1815-1945

Group	Number of buffer-state-years	Percentage of Years that Witnessed a violation
Alliance with trade provision	61	0 percent
Alliance with no trade provision	898	17 percent

Difference statistically significant at $p<0.01$ (two-tailed).

Table 4.3: Military Alliances, Buffer States and Alliance Duration, 1815 to 1945

Group	Number	Average Number of Years in the Alliance
Buffer states in alliance with trade provision	8	21 years
Buffer states in alliance with no trade provision	31	13 years

Table 4.4: European Countries with Coal Output Data, 1815 - 1950

Country Name	Years Available
Austria	1819-1867, 1920-1950
Austria-Hungary	1867-1915
Hungary	1850-1867, 1920-1950
Belgium	1831-1950
France	1815, 1821,1825,1824-1950
Prussia/Germany/ West Germany	1817-1950
Spain	1839-1840,1842,1844-1845, 1847-1848,1850,1854-1856, 1858,1860-1950
Britain	1815,1830-1950
Bulgaria	1894-1950
Italy	1868-1950
Netherlands	1886-1950
Romania	1898-1916, 1919-1948, 1950
Russia	1860-1940, 1945-1950
Serbia/Yugoslavia	1893-1896,1898-1912, 1920-1940,1945-1950
Sweden	1874-1950
Czechoslovakia	1913-1950
Denmark	1940 - 1950
Greece	1916-1950
Poland	1920-1950
Portugal	1915-1950

source: Mitchell (1998)

Table 4.5: Buffer States and Alliance Membership, 1815-1945

	Coef. (se)
Alliance without Trade Provision	
Rivals' Asymmetry	3.16**
	(1.45)
Coal Production	0.00
	(0.00)
Alliance in Previous Period	4.58***
	(0.41)
Buffer's military personnel	1.25**
/total military personnel held in relevant rivalry	(0.64)
Constant	-2.69***
	(0.55)
Alliance with Trade Provision	
Rivals' Asymmetry	-0.68
	(2.40)
Coal Production	0.01*
	(0.00)
Alliance in Previous Period	6.20***
	(0.79)
buffer's military personnel	1.28
/total military personnel held in relevant rivalry	(1.46)
Constant	-4.46***
	(0.79)
No.	503.

* p¡0.10, ** p¡0.05, *** p¡0.01

Figure 4.1: Percentage of Alliances with Trade Cooperation Provisions, 1815-2000 (by time period)

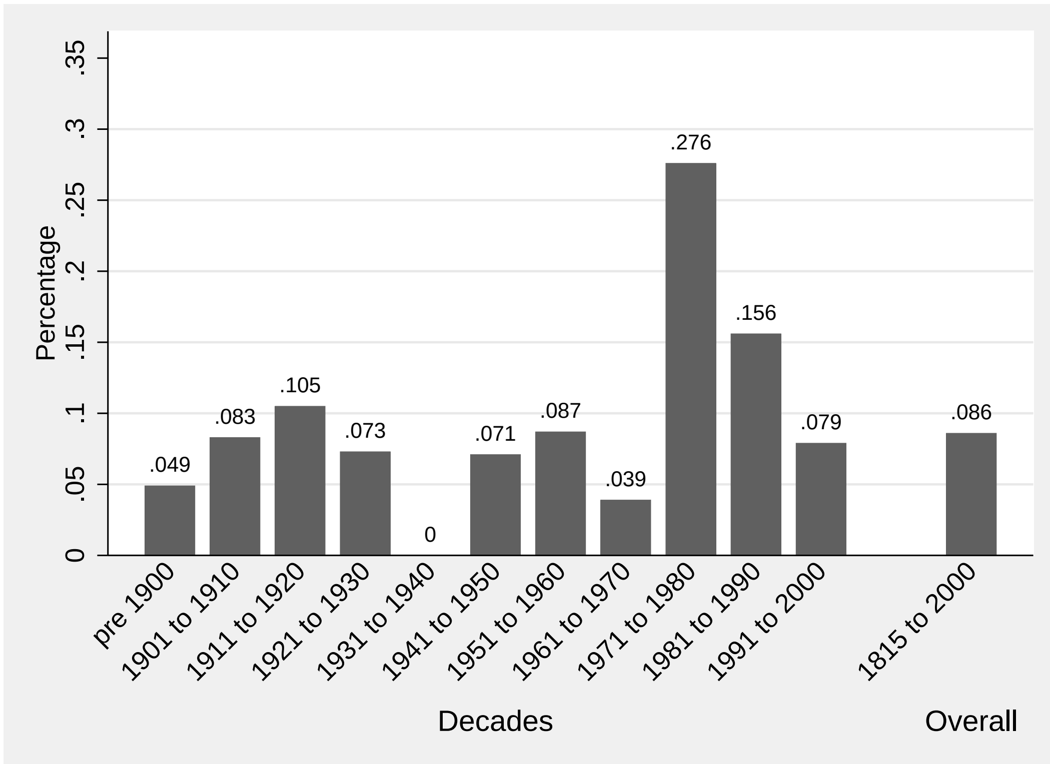
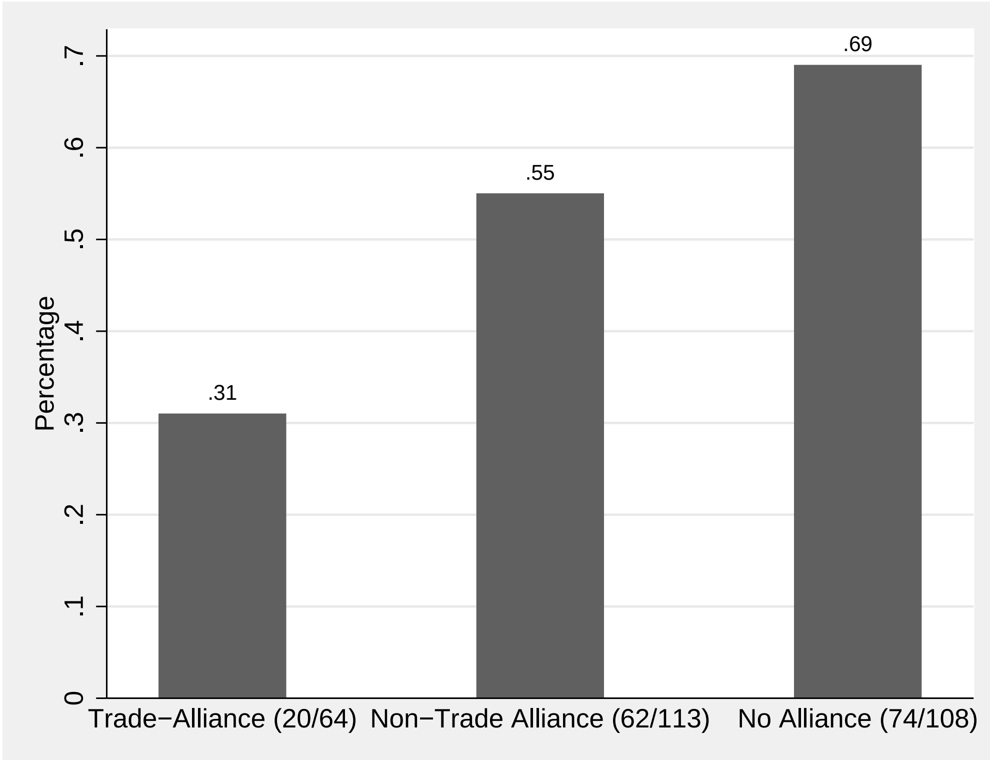
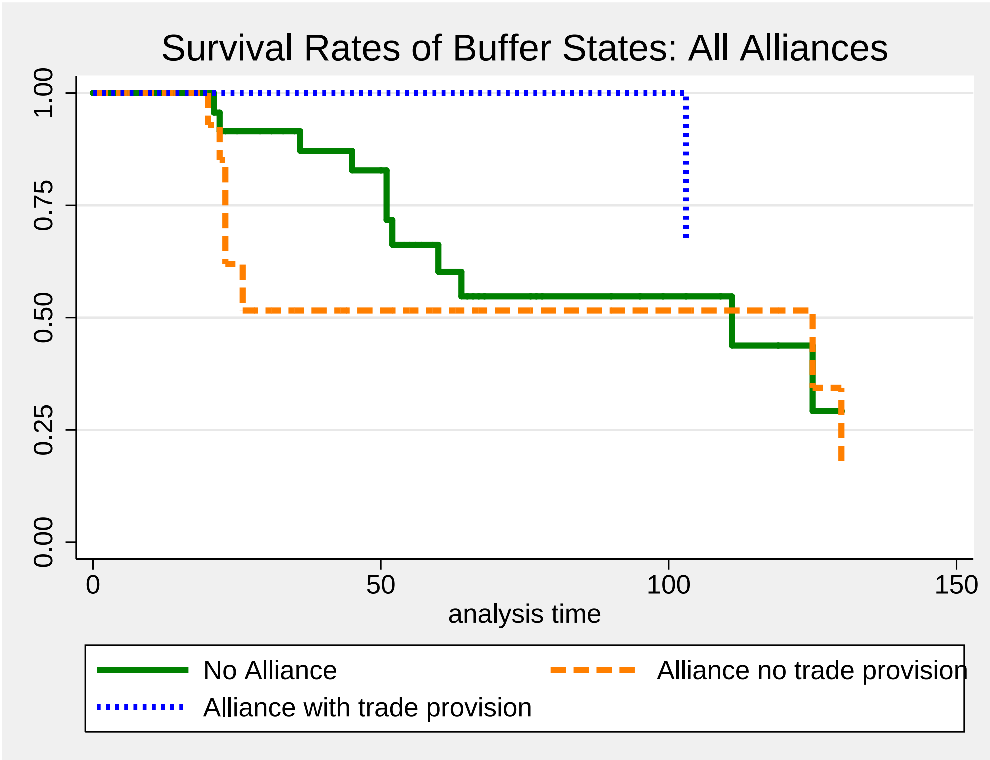


Figure 4.2: Percentage of Buffer-State-Years that Witnessed an Unprovoked MID



NOTES: Pearson chi2 comparing Trade-Alliance group and Non-Trade-Alliance group = 6.7500 (p-value 0.019). Pearson chi2 comparing Trade-Alliance group and No Alliance group = 21.1724 (p-value 0.000).

Figure 4.3: Survival Rate of Buffer States, Different Alliance Arrangements



CHAPTER V

Conclusion

5.1 What Did I do?

This dissertation sought to empirically identify the effect of issue linkage, both on the probability of agreement formation and on the likelihood of commitment. Chapter 2 explained how quantitatively exploring multilateral events (such as alliance negotiations) requires making each event the unit of analysis. This stands in contrast to the common practice in quantitative international relations research, which is to divide the participants of multilateral events into a series of dyadic combinations. Chapter 3 presented a new dataset of successful and failed alliance negotiations from 1815 to 1945 and used matching analysis to identify the causal effect of issue linkage on the probability of agreement. Chapter 4 used buffer states (states located between two recently warring rivals) as a 'hard test' of the claim that issue linkage enhances treaty credibility. The alliance relations of buffer states are ideal for testing the ability of issue linkage to enhance treaty credibility because buffer states are especially prone to invasion and occupation. This high vulnerability makes other states reluctant to remain committed to an alliance agreement with the buffer state.

5.2 What Did We Learn?

First, we learned that when analyzing a multilateral event, one should use k-adic data, where $k >= 2$. However, if the event of interest is purely bilateral, then one can use dyadic data. Second, we learned that, from 1860 to 1945 offering to expand a military alliance negotiation along an economic dimension increased the probability of agreement by 36 percentage points. Third, using a variety of analysis techniques, I found that buffer states in alliances with trade provisions avoid occupation and invasion at a higher rate than buffer states in other alliance arrangements; that third parties attack buffer states in alliances with trade provisions at a lower rate than in other alliance arrangements; and that buffer states in alliances with trade provisions experience fewer opportunistic violations of the alliance terms by their alliance partners. Since it appears that issue linkage can help buffer states to form credible commitments, issue linkage should be a useful tool in nearly any context. Overall, these findings suggest that issue linkage does work: it increases the probability of states reaching agreement and it improves the credibility of an agreement once it has been reached.

5.3 Why Does it Matter?

Since the results and evidence offered in chapters 3 and 4 confirm two widely held views regarding the role of issue linkage in international negotiations, why do they matter? In other words, what is to stop someone from saying "I already knew all of this, so why should I bother reading the dissertation?" The evidence and results in chapters 3 and 4 matter for six reasons.

First, as mentioned in the introduction, though the views that issue linkage provisions increase the probability of agreement and enhance treaty credibility are widely held, there are some notable detractors. Thus, the evidence presented in chapters 3 and 4 could lay to rest the arguments of these doubters.

Second, chapter 3 shows that over 40 percent alliance negotiations end in non-agreement during the 1815 to 1945 time period; an exceptionally high rate of failure. This finding may shed insight into the high rate of alliance compliance identified by Leeds, Long, and Mitchell (2000). Leeds, Long, and Mitchell found that, in contrast to prior work by Sabrosky (1980), alliance members failed to comply with their alliance commitments only 25 percent of the time. However, they state how "in this study, we simply identify reliable and unreliable alliances. A clear next step is to explain why some alliances are reliable and others are not" (Leeds, Long, and Mitchel 2000, p. 697). Leeds, drawing from Downs, Rocke, and Barsoom (1996), begins this work by conjecturing that, "under most conditions, leaders are reluctant to make promises that they or their partners are unlikely to uphold" and, therefore, leaders carefully select the agreements that they are willing to make (Leeds 2003, p. 808). This claim is consistent with the finding that a high percentage of alliance negotiations end in non-agreement. In particular, the surprisingly high rate of negotiation failure identified in chapter 3 suggests that states are indeed particularly selective in choosing alliance partners.

Third, the results in chapter 3 suggest that linkage offers are exceptionally costly. Specifically, though the study reveals that issue linkage works, it also finds that it is rarely used. This is particularly surprising given the high rate of alliance negotiation failure. It could be the case that the participants attempted other forms of issue linkage (though, as discussed with foreign aid offers, it is debatable whether such offers could indeed be classified as linkage). Even if this is the case, it still leaves a substantial number of negotiations that failed without witnessing an offer of linkage.

Various costs associated with linkage were highlighted in chapter 3, but the rarity of linkage offers suggests that these costs are, in many instances, quite prohibitive.

Fourth, the analysis in chapter 3 shows that scholars should rethink their approach to analyzing alliance formation. By focusing on negotiations, the results suggests that, to borrow from Gartzke, alliance formation may lie in "the error term" (Gartzke 1999). The high rate of alliance negotiation failure suggests that states decide whether or not to begin alliance negotiations and *only then* determine if the negotiations will actually result in an agreement. Presently, scholars view factors such as threat perceptions and relative capabilities as important determinants of alliance formation. The evidence here suggests these factors may may play a secondary role to the bargaining tactics (such as linkage offers) employed at the negotiating table.

Fifth, the results in chapter 4 illustrate that trade can enhance the effectiveness of deterrence. Previous studies have found similar results. Huth and Russett (1984) and Huth (1988) highlight how deterrence attempts by a defender are more likely to succeed if the defender has important but indirectly related trade interests with the target. Asyegul (2010) finds that foreign trade between the target and third-party defenders enhances the success of extended general deterrence only when it takes place within an institutional setting such as a regional economic institution. However, unlike this paper, these previous studies did not consider instances when trade cooperation is directly linked to security cooperation.

Sixth, the results in chapter 4 suggests that some buffer states can indeed 'cheat death.' While one may suppose that the presence of a valuable commodity could make buffer states even more susceptible to invasion, this research suggests quite the opposite: those buffer states endowed with a valuable tradable good can use it to ward off invasion and occupation.

5.4 Beyond Alliances?

Even if one accepts the above reasons for why these results matter, he or she may counter that the results only pertain to alliances and, moreover, only alliances formed prior to 1945. Since these results pertain to a specific type of treaty (alliances) during a specific time period (pre-1945), what relevance do they have for non-alliance negotiations today?

Arguably, the stakes are higher for alliance negotiations than for any other type of treaty (with the exception of perhaps peace agreements). Alliance negotiations deal with something as fundamental as countering external threats to the survival of a state. Concerns over state survival were particularly acute prior to 1945, since Fazal(2004, 2007) found that states were more likely to experience violent invasion and occupation during this time period.[1] This means the willingness of states to accept compromises (be it in the form of accepting a linkage proposal or otherwise) would perhaps be its lowest in pre-1945 military alliance negotiations.

Therefore, as shown in chapter 3, if issue linkage offers can secure agreement during pre-1945 alliance negotiations (where the stakes were exceptionally high), it should work for a whole host of negotiations. Moreover, as shown in chapter 4, if linkage provisions can enhance the credibility of alliance commitments for buffer states (whose high susceptibility to invasion and occupation makes other states reluctant to form alliances with them), then linkage provisions should improve treaty compliance in nearly any context.

[1] It should be noted that this propensity to be invaded and occupied was highest for buffer states - states located between two recently warring rivals.

5.5 What's Next?

Perhaps one of the more intriguing puzzles uncovered by the above analysis is the high rate of alliance treaty negotiation failure. Specifically, the data on alliance negotiations revealed that approximately 41 percent (127 out of 308) of European military alliance negotiations prior to 1945 ended without agreement. This is an exceptionally high rate of failure, especially since one would expect leaders and diplomats to be able to "look down the game tree" and, as a result, avoid negotiations that were likely to fail. What makes alliance negotiations fail? Given the observed rarity of linkage offers, why don't leaders propose linkages that make them succeed? Is there an informational problem? A commitment problem? A domestic political problem?

In some respects, this is very similar to the question of why wars occur when they are costly to all sides and could be avoided. In other words, why should the negotiations be undertaken given that beginning them is not costless and allowing them to end in failure could prove risky and costly (e.g. one might upset a major power; one is possibly forgoing the protection of another state). The rest of this chapter will offer some initial thoughts on these questions and propose an approach for exploring them.

5.5.1 Why Negotiations Fail: Preliminary Evidence

To begin, it could be useful to actually identify why negotiations fail. I begin this process by considering a random sample of 40 of the 127 failed negotiations. I re-read the accounts of these 40 negotiations in order to identify the exact cause of their failure. The reasons for negotiation failure can be classified into 5 broad categories: Exogenous Shock; Outside Offer; Two Level Game; Bargaining Failure; and Balance of Power Concerns. I will now describe each category.

- *Exogenous shock:* In this category, a negotiation fails due to a change in the state of the world that suddenly renders the negotiation of the treaty irrelevant. Examples include the onset of a major war, the unexpected defeat of one of the negotiating parties, or the unexpected onset of revolutions inside one of the negotiating parties. For instance, the 1871 negotiations between Austria, France, and Italy ended due to French forces being defeated by Prussian forces along the Rhine. Alternatively, the 1848 negotiation of a pact to not invade France between Britain, Austria, Russia, and Prussia ended with the outbreak of revolutions throughout Europe.

- *Outside Offer*: In this category, a negotiation fails due to one of the parties receiving an outside offer. Specifically, this category focuses on instances in which one of the states receives an offer to form an alliance by a state that is not party to the negotiations. For instance, Austria ended negotiations with Prussia in 1848 because Russia offered to form an alliance without conditions, while Prussia would not offer an unconditional alliance.

- *Two Level Game*: In this category, a negotiation fails due to one of the states being unable to acquire domestic support. The idea of negotiations being a two-level game, in which negotiators (level I) may be constrained by domestic political factors (level II), was introduced by Putnam (1988). This category focuses on instances in which the failure of one of the parties to acquire necessary domestic approval (perhaps because of a requirement for ratification) results in non-agreement. For instance, the British parliament refused to support an alliance with France in 1848. Realizing this, the British delegation left the negotiations.

- *Bargaining Failure*: In this category, a negotiation fails due to standard causes of bargaining failure, such as an inability to redistribute the benefits of the treaty or find a mechanism by which the treaty could be enforced (Koremenos, Lipson, and Snidal 2001; Fearon 1995). For example, the negotiations fail due to one of the parties facing a distribution problem and the other party being unwilling (unable?) to make adequate concessions. For instance, when negotiating a defensive alliance in 1854, Austria wanted Russia to guarantee that the Ottoman Empire would not be dismantled if defeated in a war by Russia. Russia was unwilling to make this concession to Austria. Therefore, the negotiations ended in a non-agreement.

- *Balance of Power Concerns*: This category focuses on a type of bargaining failure that is unique to alliance diplomacy: negotiations fail because one of the participants is concerned that the proposed terms of the alliance will provoke aggression and/or a counter alliance by a state outside the negotiations and another state in the negotiation is unwilling to weaken the terms of the alliance. For example, Prussia ended negotiations with Britain and France to form an alliance to defend Turkey in late 1870 because it did not want to induce Russia to form a counter alliance. According to British foreign secretary Granville, the agreement would "act as a powerful check against Russia."[2] However, in the words of historian A.J.P. Taylor, it was not in Prussia's interest to defend British interests against Russia "without other rewards than a gruding patronage."[3]

I will now use these categories to break down the sample of 40 military alliance negotiations. Table 5.1 shows the number of negotiation failures that fall into each category.

[2]Taylor 1956, p. 216.
[3]Ibid.

Table 5.1: Cause of Alliance Negotiation Failure, summary stats from 40 select cases

Cause	Number of Cases
Exogenous Shock	4
Outside Offer	3
Lack of Domestic Support	3
Bargaining Failure	9
Balanec of Power Concern	21

The presence of outside offers, the impact of unexpected wars or coups, and a lack of domestic political support - though not non-existent - appear to have relatively minor influence on negotiation failure. This suggests that the high non-agreement rate can not be explained as simply the result of states using one negotiation as leverage in another negotiation, the inability of leaders to successfully ‘sell’ the agreement back home, or the bad luck of a sudden coup or unexpected war. In short, alliance negotiations do not appear to fail due to factors outside the control of the diplomats engaged in the negotiation itself.

Instead, alliance negotiations appear to fail due to an inability on the part of the states in the negotiation to find an acceptable bargaining solution. Specifically, *Bargaining Failures* and *Balance of Power Concerns* explain 75 percent of the failures. Both of these categories relate to instances in which at least one of the foreign ministers finds the terms of the agreement unacceptable and is willing to end the negotiation in non-agreement, rather than sign what he feels is a bad deal.

Though bargaining failure can occur for a number of reasons (informational asymmetries, commitment problems), what is puzzling about these data is the propensity for states to even *enter* negotiations that have a high chance of failure. According to conventional bargaining models, if there is complete information, then the parties will enter a negotiation only when they know it will end in success. It is only when uncertainty is introduced that states will decide to enter a negotiation that can end in failure.

5.5.2 Next Papers

Unlike previous empirical studies of alliance formation, which treat states as having either entered or not entered an alliance, the evidence presented here suggests that alliance formation is really a two step process: states decide whether or not to enter into a negotiation and *then* states decide if the negotiation will succeed or fail. This suggests that scholars should empirically study alliances as a two stage process in which actors first make strategic decisions about whether to enter into alliance negotiations and only then begin to bargain over the terms of the alliance.

I plan to write two papers based on these findings. The first paper will be titled, "Why Negotiations Fail." It will begin by posing the question, "why is the rate of non-agreement in alliance negotiations so high?" Next, it will present the above data, except now for the entire 127 cases of failed alliance negotiations. Third, it will seek to identify the factors/conditions that explain why some negotiations succeed while others fail. For example, it will offer an explanation for why some states will allow negotiations to end in non-agreement while similarly situated states will weaken the treaty language so as to secure agreement.

The second paper will be titled, "Alliance Negotiation as a Two Stage Process." This paper will empirically model alliance negotiations as a two stage process. In the first stage, states decide whether or not to begin alliance negotiations. I argue that the factors influencing this decision are those that scholars have traditionally considered as determining alliance formation. These factors include threat perceptions and capability aggregation considerations. In the second stage, states decide whether the negotiations should end in agreement or non-agreement. Agreement at this second stage is a function of those factors that influenced the first stage (threat perceptions, capabilities, etc) and negotiation tactics: economic issue linkage; offers of territory swaps; use of ultimatums; etc. In the end, these next two papers should offer a more complete understanding of the alliance formation process and the factors that lead to diplomatic success or failure.

APPENDICES

Complete Description of Gibler and Wolford (2006) Variables

Common Threat is a dichotomous variable coded 1 if each state has participated in a Militarized Interstate Dispute (MID) against the same third state sometime in the previous 10 years, 0 otherwise. *Single Major Power* is a dichotomous variable coded 1 when one member of a dyad is a major power as indicated by the Correlates of War coding of major power status.[4] *Geographic Distance* gives the square root of the capitol to capitol distance, unless states are contiguous, in which case distance is set to 0. *Learning* has a range of -1 to 1. If a country has a lesson favoring alliance, the variable is coded 1; if it has a lesson favoring neutrality, the variable is coded as -1; and if it has no lesson, the variable is coded 0. The Learning score for the dyad is the combined score of the two states, thereby creating a variable ranging from -2 to 2.[5] *Existing Alliance* is coded 1 if the dyad members were already members of the same alliance (bilateral or multilateral) entering that year, 0 otherwise.

MID between A and B over 10 yrs is coded 1 if the two states were on opposite sides of a MID during the previous 10 years. Being on opposite sides of a MID in the past ten years is expected to diminish the probability of the dyad forming an alliance. Therefore, *Amount of Threat* counts the number of MIDs in which each state participated in the previous 10 years. The *Joint Language* variable is coded 1 if the two states in the dyad have the same predominant language, 0 otherwise.

[4]Gibler and Wolford (2006) actually use the Lai and Reiter (2000) coding of Major Power, which is coded 1 if *either or both* members of the dyad are major powers, 0 otherwise. However, the Lai and Reiter (2000) *Major Power* variable does not fully capture the level of asymmetry in the dyad (because it is coded 1 if the dyad contained 1 or 2 major powers), something that Morrow (1991) emphasizes is a key determinant of alliance formation. According to his argument, alliances serve as a type of "protection racket", whereby a large powerful state offers to protect a smaller state (give security) in exchange for access to the small state's territory, changes in the small states foreign policy, or other concessions (give autonomy).

[5]The coding depends in part on the Reiter (1996) coding for lessons learned by small powers after world wars. A state was coded as having a lesson in the postwar period favoring neutrality if it was neutral during World War I or II and was not invaded or if it was allied during such a war and was invaded. Conversely, a state had a lesson favoring alliance in the postwar period if it was allied during a world war and was not invaded or if it was not allied and was invaded. Only states in major theaters of war were coded as learning lessons. Similar coding logic is applied for nations following the Napoleonic wars.

The *Joint Ethnicity* and *Joint Religion* variables are coded similarly for ethnicity and religion. Using the Polity IV polity score (with a value of -10 to +10), *Polity Difference* uses the absolute value of the difference between the Polity IV scores of the two states in the dyad to create a measure of regime similarity, while *Joint democracy* is a dummy variable coded 1 if both states have polity scores of 5 or higher.[6]

[6]There is also a variable, *Trade*, that measures the bilateral trade flows between the two states. However, the inclusion of this variable in the model is problematic for two reasons. Methodologically, the limited availability of trade data means much data is lost when this variable is included in the regression model. Substantively, the literature on trade and alliances has been wrapped in a "chicken-or-the-egg" dilemma. In other words, does trade between two states lead to an alliance or does the presence of an alliance lead to trade? On one side is work such as Gowa and Mansfield (1993) who incorporate security externalities into the relative gains considerations of states. Because a portion of the wealth gained from international trade can now be spent on the military, the subsequent increase in military size proves beneficial to security allies ("my friend's gain is my gain"), but detrimental to security adversaries ("my enemy's gain is my loss"). On the other hand are studies such as Fordham (2007), who argues that the fear of losing a valuable trading relationship deters bilateral conflict between trading partners. In turn, this gives states a motive to defend their trading partners from external threats that might disrupt commerce. Consequently, trade will lead to alliances.

Alternative Triadic-Dyadic Dataset Simulation

A major downside to the Bilateral-Trialateral simulation in the main text is that it did not produce a mixture of bilateral alliances and multilateral alliances that matched that of a real dataset. The reality is that multilateral alliances are nearly 5 times more prominent than bilateral alliances. It is not possible for a DGP with capability ratio alone to create such a dataset. One way to compel the DGP to provide the proper distribution of bilateral and multilateral alliances is to return to the simpler model (with capability ratio as the only independent variable), but allow the parameter on capability ratio to differ for triads and dyads. Specifically, the model estimated is

$$xb = \text{cons} + \beta_1 * \text{cap ratio} * \text{dyad} + \beta_2 * \text{cap ratio} * \text{triad} + u$$

where *dyad* is a binary variable coded 1 if an observation is a dyad, 0 otherwise and *triad* is a binary variable coded 1 if an observation is a triad, 0 otherwise. The true value on β_1 is 10, while the true value on β_2 is 5 and the constant is -12. Admittedly, a model with different **cap ratio** coefficients for triads and dyads lacks a theoretical motivation. However, it allows one to test the ability of logit estimation to recover the true parameter estimates of a choice-based sample containing an accurate distribution of bilateral and multilateral alliances. Specifically, a typical simulation produces about 10 to 20 trilateral alliances and 60 to 70 bilateral alliances. Table A.1 reports the results from 500 Monte Carlo simulations of the DGP. Applying logit estimation to each simulation of the full dataset results in average parameter estimates close to the true parameter values. Additionally, the estimates from the choice-based sample perform quite well relative to the estimates from the full sample.

	Triadic-Dyadic DGP estimated with Triadic-Dyadic Data	Triadic-Dyadic DGP estimed with Choice-Based Sample
β on Cap Ratio for Bilateral Alliances ($\beta_1 = 10$)		
Bias	-0.01	0.07
Root Mean Squared Error	0.80	0.96
Over Confidence	1.07	1.64
β on Cap Ratio for Trialateral Alliances ($\beta_2 = 5$)		
Bias	-0.08	0.14
Root Mean Squared Error	5.05	5.16
Over Confidence	6.31	7.09

Alternative Choice-based Sampling Approach

I also tried choice-based sampling on both Y and X (cap ratio), which I will describe in this footnote. Since the population of triads is large (161,700), but the number of triads where Y=1 is comparatively small (approximately 120), this leaves a rather large population from which to draw my rather small random sample of triads where Y=0 (240). Given the small size of this random sample, it is quite possible to draw a random sample with little variation on **cap ratio**. In turn, this would limit the power of my test. I use the following procedures to stratify on X and Y:

1. Multiply the number of Y=1 triads by 2 to determine the number of Y=0 triads I need for my estimation. Call this value N_0

2. Divide N_0 by 4 to determine the number of observations I need to draw from each **cap ratio** quartile of the Y=0 triads. Call this value $\frac{N_0}{4}$

3. Starting with the first quartile of the Y=0 triads, take a random sample equal to size $\frac{N_0}{4}$.

4. Assign a weight to the observations in this sample. Assuming the first quartile has M observations, the weight, pi, is pi $= \frac{\frac{N_0}{4}}{M}$

5. Repeat for each quartile of triads.

6. Combine this sample of Y=0 triads with my Y=1 triads and estimate.

7. Repeat 500 times

This procedure lowers the bias from 1.27 to 1.13, lowers the Root Mean Squared Error from 3.59 to 3.49, and lowers Over Confidence from 7.34 to 6.55. Though these are reductions, I do not view them as substantial enough reductions to justify describing this procedure in the main text.

The Feasibility of Using Spatial Interdependence Weighting Matrix to Model the Dependent Variable

In the main text, I discuss how spatial interdependence regression models do not yet offer a viable method for estimating the creation of multilateral alliances. I offer one reason in the text (the weighting matrix captures whether or not i and j have a connection, which is precisely what one needs to estimate) There is a second reason not discussed in the text.

Even if one can model the dependent variable, a second issue is sheer complexity. The state's decision to enter a particular alliance depends on other state's decisions, but not in a linear-additive way, as would be most-easily and directly captured by a $\mathbf{Wy}$ term. Instead, it would depend on which combinations of others entered that alliance, which is some large set of possible alliances that a state may enter or not enter. Conceptually, each possible alliance is an outcome. This would result in having N observations of 1 or 0 for entering that alliance for each of N states, or an $N * N!$ matrix. Moreover, one would want to specify a W of dimensions $N * N! \times N * N!$ that reflected properly how the dependence of each specific state's entry to that specific alliance depended on other states' entry decisions on that alliance and on this and other states' entry decisions on other possible alliances. Finally, one would need one such enormous and complicated W for each combination of others' decisions on their sets of entry decisions that we thought could be important to determining whether this specific state enters this specific alliance.

Matching Results for 1860 to 1913 and 1919 to 1945

I am also concerned that the positive results may be driven by my decision to test the entire 1860 to 1945 time period. Specifically, one could argue that the international environment during the war years of 1914 to 1919 and 1939 through 1945 and during the interwar years of 1919 to 1939 were dramatically and substantively different from those during the the period of 1860 to 1913.

It is also commonly recognized that the beginning of World War I marked the end of this first "golden age" of economic globalization and the beginning of the inter-war period of economic crisis.[7] The early part of the post-World War I time period was dominated by the post-war recessions within the belligerent powers, while the 1930s experienced a sharp decline in international trade due to the global Great Depression. The global international trade system then broke into various trade blocs and was mared by the imposition of retaliatory tariffs, sparked by the implementation of the Smoot-Hawley Tariff by the United States in 1930. Consequently, one should conduct separate analysis on the 1860 to 1913 and 1919 to 1939 time periods.

The below table shows the average treatment effect when I conduct matching only on those observations from 1860 to 1913 time period and only those observations from the 1919 to 1939 time period. It should be noted that since each time period has a relatively small sample size, I conduct this analysis using the unrestricted sample (meaning it is not restricted to just negotiations over offensive/defensive alliances).[8] The table reveals two important observations. First, it shows that the effect of trade linkage on the probability of negotiation agreement is positive and statistically significant for both the 1860 to 1913 and 1919 to 1939 time periods. Second, the

[7] Pahre 2008, p. 14. See, for example, Eichengreen 1996; Bordo, Eichengreen, and Irwin 1999, and Oatley 2004
[8] Results from conducting matching on the restricted samples are substantively the same.

effect for the 1919 to 1939 time period is substantially smaller (0.15) than for the 1860 to 1913 time period (0.56). Therefore, it does appear that considering the 1860 to 1945 time period in its entirety masks the rather large positive effect found in the earlier portion of this time period and the more modest effect found in the later portion of this time period. However, considering the 1860 to 1945 time period in its entirety does not drive the estimated positive effect of issue linkage.

Time Period	ATE	Lower Bound .95 CI	Upper Bound .95 CI	Observations
1860 to 1913	0.56	0.47	0.66	101
1919 to 1939	0.15	0.08	0.22	87

Results contingent on 1860 starting year?

Building from the previous analysis, one might be concerned that my results for the 1860 to 1913 time period are contingent on using 1860 as a start year for my analysis. To address this concern, table 5 shows the average treatment effect when I allow the beginning point of the 1860 to 1913 time period to move back 5 years at a time.

Table 5 reveals two important observations. First, it shows that the positive effect of trade linkage on the probability of negotiation agreement from 1860 to 1913 is not dependent on the 1860 cutoff. Second, the sudden change that occurs once the time is taken back to 1845 suggests that it was perhaps the British unilateral free trade policy, not the adoption of the Cobden-Chevalier trade pact, that constituted the major "break point" for the international system.

Time Period	ATE	Lower Bound .95 CI	Upper Bound .95 CI	Observations
1860 to 1913	0.57	0.47	0.66	99
1855 to 1913	0.56	0.47	0.65	105
1850 to 1913	0.55	0.47	0.64	120
1845 to 1913	0.23	-0.06	0.51	136

Probit Results for 1860 to 1945 time period

The main text showed results matching algorithms. Though I view matching as the most method given my data, I offer here results from probit analysis. Due to small sample size for the sub-periods, I must conduct the probit analysis over the entire 1860 to 1945 time period (in order to have enough power and to avoid perfect separation in the data). The results are shown in table A1. One can see that the coefficient is positive and significant at the 0.10 confidence level. Subtantively, the marginal effect of a trade provision is 0.30 (it increases the probability of agreement by 30 percentage points).

Variable	Coefficient
ECON	1.12*
	(0.59)
ally	-0.55***
	(0.18)
jdem	0.09
	(0.31)
min mil	0.06
	(0.16)
min dist	-0.27
	(0.19)
peak year	-0.10
	(0.19)
N	-0.19
	(0.17)
buffer	-0.01
	(0.19)
prior negotiation	-1.07***
	(0.19)
Constant	1.86***
	(0.48)
No.	256.00

* p¡0.10, ** p¡0.05, *** p¡0.01

Genetic Matching Algorithm on 1860 to 1945 time period

To illustrate that my results are not driven by the propensity score matching algorithm, I conduct analysis using genetic matching[9]. Genetic matching uses a search algorithm to iteratively check and improve covariate balance, and it is a generalization of propensity score and Mahalanobis Distance (MD) matching.[10] Specifically, rather than minimizing a single metric, such as the MD metric, the algorithm searches amongst a range of metrics to find the particular distance measure which optimizes post-matching covariate balance. Each potential distance metric considered corresponds to a particular assignment of weights, W, for all matching variables. The algorithm weights each variable according to its relative importance for achieving the best overall balance.

Column 1 of Table A.2 reports the results from a difference of means test after genetic matching was used to achieve balance on the entire 1860 to 1945 time period. The estimated effect of 0.33 is smaller than the estimated effect of 0.36 when applying the Abadie and Imbens algorithm to the restricted sample, but larger than the than the effect of 0.28 estimated when applying the Abadie and Imbens algorithm to the unrestricted sample (or 0.30 from the probit model estimated in Appendix A).

	Genetic Matching
ATE	0.33
p-value	0.00
N	235

[9]Diamond and Sekhon 2008
[10]Rosenbaum and Rubin 1985.

Matching Results for 1815 to 1859 time period

Below are the results for the 1815 to 1859 time period. One can see that though the effect of trade linkage offers is negative for the 1815 to 1859 time period (decreases the probability of agreement by nearly 51 percentage points), the 95 percent confidence intervals are quite large and include zero. Thus, the effect is not statistically distinguishable from zero. Moreover, I was not able to achieve covariate balance for this sample of data.

What explains the statistically insignificant negative effect for trade linkage between 1815 and 1859? Is it due to (1) the trade linkage offers being truly meaningless during the 1815 to 1859 time period, (2) the negative effect of mercantilist trade demands being offset by the positive effect of trade cooperation offers by, for example, German states seeking to join the *Zollverein* customs union, (3) trade demands actually having a negative effect on the probability of agreement, but the effect is measured with a great deal of error, or (4) trade linkage offers being negatively associated with the probability of alliance agreement because they were only used in negotiations that were unlikely to succceed in the first place? Space constraints prohibit this article from exploring which of these is the true explanation for the null effect. However, understanding exactly why issue linkage does not appear to increase the probability of agreement during the 1815 to 1859 time period could greatly enhance our knowledge of when linkage offers are likely to work.

Time Period	Effect of Trade Linkage	Lower Bound .95 CI	Upper Bound .95 CI	Observations
1815 to 1859	-0.51	-1.40	0.38	74

Match On: Military Size, Buffer, Contiguity, Crisis Period, Democracy, Offensive & Defensive alliances, prior negotiation

"Peer-Effects" impacting the 1860 to 1945 Results?

One might also be concerned that the presence of strategic interdependence between the countries in my data. Specifically, the fact that the United Kingdom offered economic issue linkage to Austria might influence the negotiations between the United Kingdom and France (since France might then also demand an offer of issue linkage).[11] Such 'peer-effects' are a common concern for causal inference, but the solutions are typically unclear. As Manski (2000) states, "when observed outcomes constitute the only empirical evidence available, a researcher who conjectures the presence of endogenous interactions within any hypothesized group cannot be proved wrong."[12] Ultimately, the interdependencies between groups of countries might be so subtle that one can never truly account for all interdependencies. For instance, Suppose A and B are negotiating an alliance and A and C have MFN status with one another. If A grants to B a trade concession in order to form the alliance, then C will also be given this tariff rate (assuming the rate offered to B is lower than what is currently offered to C, which may not be the case). As a result, if C is in negotiations on an alliance with D, the new rate C receives from A could alter any rate accepted by C from D or offered from C to D. Because such subtleties could undermine the application of nearly all estimation methods to any data, scholars, such as Graham, Imbens, and Ridder (2009), will often recognize such 'peer-effects', but then simply assume them away.[13]

Rather than assuming away the problem, Manski (2000) suggests that the researcher think hard about the source of the violation in the data brought to bear

[11] Scholars working within the 'causal inference' framework, would recognize such strategic interdependencies as violating the Stable Unit Treatment Value Assumption (SUTVA).

[12] Manski 2000, p. 130.

[13] Graham, Imbens, and Ridder 2009, pp. 9-10.

on the research question.[14] Therefore, I will attempt to address the most obvious manifestation of a 'peer-effect' in my data. Specifically, if the concern is that country C will demand an economic linkage offer from country A because country A offered such a linkage to county B, then the solution is to ensure that country A is not in *both* the treated and control case in a match pair. Therefore, I inspect the matches from the above analysis and throw out matches that have the same country. This led to 75 matches (out of 140) being removed. After removing these matches, the ATE was estimated to be 37 percent points (with .95 confidence intervals that do not include zero), compared to an original ATE of 36 percentage points. Additionally, I considered the opposite case: a worse case scenario in which I estimated the ATE on just the 75 that were the *most* likely to violate SUTVA. This led to an estimated ATE of .45 (with .95 confidence intervals that do not include zero).[15] Since neither procedure substantively altered the estimated average treatment effect, we can continue under the assumption that the data is closely approximating SUTVA.

[14]Manski 2000, 132.

[15]STATA do files for conducting these and other robustness tests are available upon request.

Inter-Coder Relibility Check

In order to test my application of the coding instrument, I hired two graduate students to code select chapters from Taylor's *Struggle for Mastery of Europe.* Specifically, I had the students code chapters 10, 11, and 12 (approximately 80 pages), as these three chapters contained perhaps the largest concentration of failed negotiations.

I instructed the students to read the chapters looking for evidence of (1) a **meeting** (correspondance of letters, physical meeting) at the diplomatic level (between ambassadors, heads of state, foreign ministers) in which (2) a **proposal** of a formal (i.e. written) alliance (defensive, offensive, neutrality, consultative, or non-aggression) is made and there is evidence of (3) a **rejection** (one side must decline forming the alliance).

All three coders (myself and the two graduate students) individually identified the same cases of failed negotations with economic linkage. With respect to the total number of failed alliance negotiations, I coded 15 failures, student 1 coded 19 failures and student 2 coded 17 failures. Though the two graduate students coded more cases of alliance failure than me, my failures are a subset of the failures identified by the students. This generates an intercoder reliability rate of between $\frac{15}{19} = 0.79$ and $\frac{15}{17} = 0.89$, which is well above the 0.75 rate of acceptability.

The inter-coder reliability check suggests that another individual could obtain a slightly different set of failed negotiations following my criteria. That descrepancies in coding could arise is no surprising. Even when coding successful negotiations (for which actual treaty text exists), scholars disagree. For example, the Correlates of War (COW) project and ATOP both provide datasets of the alliances that have existed since 1815. However, whereas COW identifies just under 500 alliances, ATOP identifies nearly 650. Coding failed negotiations requires a larger degree of judgement than coding successful negotiations (due to lack of a treaty), so the real question is not "will descrepancies arise?", but "are the inferences I draw from the data sensitive to such descrepancies?"

To determine if this is the case, I compare two estimates of the ATE for economic linkage offers during the 1870 to 1881 time period (the period covered by the chapters coded by the graduate student coders): the ATE from using my coding of failed negotiations and the ATE from using graduate student 1's coding of failed negotiations. I chose the coding of graduate student 1 as the number of failed negotiations identifeid by this graduate student (19) serves as an upper bound on the number of failed negotiations a coder could have identified for this time period. The ATE using my coding is -0.25, while the ATE using graduate student 1's coding is -0.275. Thus, my coding of failed negotiations could be 10 percent larger ($\frac{0.275-0.25}{0.25}$. IN fact, if one assume that the bias could be in either direction, then the estimated positive ATE for the post-1880 time period could be as large as $0.24 * (1 + .10) = 26.4$ or as small as $0.24 * (1 - 0.10) = 21.6$, while the negative effect identified for the pre-1880 time period (using the sample that removes k-ads with states that share borders) could be as large as $-0.38 * (1 + .10) = -41.8$ or as small as $-0.38 * (1 - 0.10) = -34.2$. In short, such coding descrepancies are unlikely to substantively influence my results.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Abadie, Alberto and Imbens Guido "Simple and Bias-Corrected Matching Estimators." Working Paper. Department of Economics, UC Berkley. 2002

[2] Abadie, Alberto and Imbens, Guido "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica.* 74(1) 2006. pp. 235 - 267

[3] Abadie, Alberto; Drukker, David; Herr, Jane; and Imbens, Guido. "Implementing Matching Estimators for Average Treatment Effects in Stata." *The STATA Journal.* 4(3) 2003. pp. 290-311.

[4] Achen, Christopher. 'Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong." *Conflict Management and Peace Science.* 22, 2005. pp. 327-339.

[5] Aggarwal, Vinod K. ed. *Institutional Designs for a Complex World: Bargaining, Linkages, and Nesting.* Ithaca, NY: Cornell University Press. 1998.

[6] Albrecht-Carrié, René. *A Diplomatic History of Europe Since the Congress of Vienna.* New York: Harper & Brothers. 1958.

[7] Axelrod, Robert and Keohane, Robert, "Achieving Cooperation under Anarchy: Strategies and Institutions," *World Politics*, October 1985. pp. 226-254.

[8] Aysegul, Aydin. "The deterrent effects of economic integration." *Journal of Peace Research.* September 2010. pp.523-533.

[9] Bairoch, P. "European trade policies, 1815-1915." in (eds.) Mathias, P. and Pollard, S. *The Cambridge Economic History of Europe, Volume VIII.* Cambridge: Cambridge University Press. 1989.

[10] Basinger, Scott and Hallerberg, Mark. "Remodeling the Competition for Capital: How Domestic Politics Erases the Race to the Bottom." *American Political Science Review.* 98:2. 2004. pp. 261-267

[11] Beck, Nathaniel and Katz, Jonathan. "Throwing Out the Baby with the Bath Water: A Comment on Green, Kim, and Yoon." *International Organization.* Spring 2001. pp. 487-495.

[12] Beck, Nathaniel; Katz, Jonathan; and Tucker, Richard, "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science*, October 1998, pp. 1260 - 1288.

[13] Beck, Nathaniel; King, Gary; and Zeng, Lanche. "Improving Quantitative Studies of International Conflict: A Conjecture." *The American Political Science Review* March 2000. pp. 21-35

[14] Becker, Sascha and Caliendo, Marco. "Sensitivity Analysis for Average Treatment Effects." *The Stata Journal.* 2007 7:1 pp.71-83.

[15] Benedick, R.E. *Ozone Diplomacy: New Directions in Safeguarding the Planet*. Cambridge, MA: Harvard University Press. 1991.

[16] Bennett, Scott. "Measuring Rivalry Termination." *Journal of Conflict Resolution*. April 1997. pp. 227-254.

[17] Bennett, Scott, and Allan Stam. 1997-2007. *EUGene: Expected Utility Generation and Data Management Program*. Web site: http://eugenesoftware.org. Latest version: 3.2, released October, 2007.

[18] Bennett, Scott and Stam, Allan. "Research Design and Estimator Choices in the Analysis of Interstate Dyads: When Decisions Matter." *The Journal of Conflict Resolution*. October 2000. pp. 653-685.

[19] Bennett, Scott and Stam, Allan. *The Behavorial Origins of War* Ann Arbor, MI :University of Michigan Press. 2004

[20] Bentley, A.F. *The Process of Government*. Chicago: University of Chicago Press. 1907

[21] Bernheim, Douglas and Whinston, Michael. "Multimarket Contract and Collusive Behavior." *RAND Journal of Economics*. Spring 1990. pp. 1 - 26.

[22] Bianco, William T. and Bates, Robert H. "Cooperation by Design: Leadership, Structure, and Collective Dilemmas." *The American Political Science Review*. March 1990, pp. 133-147.

[23] Blainey, Geoffrey. *The Causes of War*. Third Edition. London: MacMillian Press. 1988

[24] Buchanan, J.M. and Tullock, G. *The Calculus of Consent*. Ann Arbor: University of Michigan Press. 1962

[25] Bueno de Mesquita, Bruce and Lalman, David. *War and Reason: Domestic and International Imperatives*. New Haven, CT: Yale University Press. 1992

[26] Clarke, Kevin. "The Phantom Menace: Omitted Variable Bias in Econometric Research." *Conflict Management and Peace Science*. 22, 2005. pp. 341-352.

[27] Cochran, William G. "Matching in Analytical Studies." *American Journal of Public Health*. 43 1953. pp. 684-691.

[28] Coleman, J.S. "The Possibility of a Social Welfare Function." *American Economic Review*. December 1966. pp. 815-827.

[29] Colaresi, Michael; Rasler, Karen; Thompson, William. *Strategic Rivalries in World Politics: Position, Space and Conflict Escalation*. Cambridge: Cambridge University Press. 2007.

[30] Croco, Sarah and Teo, Tze Kwang. "Assessing the Dyadic Approach to Interstate Conflict Processes: A.k.a. "Dangerous" Dyad-Years." *Conflict Management and Peace Science*. 22:5 - 18, 2005. pp. 5 - 18.

[31] Davis, Christina, "International Institutions and Issue Linkage: Building Support for Agricultural Trade Liberalization," *American Political Science Review*. February 2004. pp. 153-169.

[32] Davis, Christina, "Linkages in Economic and Security Bargaining: Evidence from the Anglo-Japanese Alliance," *International Security*. 2009.

[33] Denemark, Robert and Hoffmann, Matthew. "Just Scraps of Paper? The Dynamics of Multilateral Treaty-Making." *Cooperation and Conflict*. 43(2) 2008. pp. 185-219.

[34] Dixon, William J. "Reciprocity in United States-Soviet Relations: Multiple Symmetry or Issue Linkage?" *American Political Science Review*. May 1986. pp. 421-425.

[35] Downs, George; David Rocke and Peter Barsoom. "Is the good news about compliance good news about cooperation?" *International Organization.* Summer 1996. pp. 379-406.

[36] Drezner, Daniel. "Bargaining, Enforcement, and Multilateral Sanctions: When Is Cooperation Counterproductive?" *International Organization* 54:1 2000. pp. 73-102

[37] Drezner, Daniel. "The Hidden Hand of Economic Coercion" *International Organization* 57:3 2003. pp. 643-659

[38] Eichengreen, Barry. *Globalizing Capital.* Princeton, NJ: Princeton University Press. 1996.

[39] Eichengreen, Barry and Frieden, Jeffrey. "The Political Economy of European Monetary Unification: An Analytical Introduction." *Politics and Economics.* July 1993. pp. 85-104.

[40] Eichengreen, Barry; Michael Bordo and Douglas Irwin. "Is Globalization Today Really Different than Globalization a Hundred Years Ago?" in *Brookings Trade Policy Forum* Susan M. Collins and Robert Lawerence (eds.) Washington, D.C.: Brookings Institute. 1999.

[41] Evans, Diana. "Policy and Pork: The Use of Pork Barrel Projects to Build Policy Coalitions in the House of Representatives." *American Journal of Political Science.* November 1994. pp. 894-917.

[42] Evans, Diana. *Greasing the Wheels: The Use of Pork Barrel Projects to Build Majority Coalitions in Congress.* New York, NY: Cambridge University Press. 2004

[43] Fazal, Tanisha. "State Death in the International System." *International Organization.* April, 2004. pp. 311-344

[44] Fazal, Tanish. *State Death: The Politics and Geograph of Conquest, Occupation, and Annexation* Princeton, NJ: Princeton University Press. 2007

[45] Fearon, James D. "Rationalist Explanations for War." *International Organization* 1995. pp. 379-414

[46] Fearon, James D. "Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs," *The Journal of Conflict Resolution* February 1997. pp. 68-90

[47] Ferejohn, J. *Pork Barrel Politcs.* Stanford: Stanford University Press. 1974

[48] Fearon, James D. "Bargaining, Enforcement, and International Cooperation." *International Organization.* Spring 1998. pp. 269-305.

[49] Firman, H. Richard "Side-Payments Versus Security Cards: Domestic Bargaining Tactics in International Economic Negotiations." *International Organization.* Summer 1993. pp. 387-410.

[50] Frankel, Jeffrey and Rose, Andrew. "Is Trade Good or Bad for the Environment? Sorting out the Causality." *The Review of Economics and Statistics.* February 2005. pp. 85-91.

[51] Franzese, Robert and Hays, Jude, "Empirical Models of Spatial Interdependence." in *Oxford Handbook of Political Methodology.* 2007a. pp. 570 - 604.

[52] Franzese, Robert and Hays, Jude, "Empirical Models of International Capital-Tax Competition." in *International Taxation Handbook.* 2007b. pp. 43-72

[53] Franzese, Robert and Hays, Jude, "Spatial-Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data." *Political Analysis.* 15(2) 2007c. pp. 140-64

[54] Franzese, Robert and Hays, Jude, "Interdependence in Comparative Politics" *Comparative Political Studies.* March/April 2008. pp. 742-80

[55] Franzese, Robert; Hays, Jude; Kachi, Aya. "The m-STAR Model as an Approach to Modeled, Dynamic, Endogenous Interdependence in Comparative and International Political Economy." Working Paper. 2009

[56] Frieden, Jeffrey. *Global Capitalism* Norton. 2006

[57] Gaddis, John Lewis. *The United States and the Cold War: Implications, Reconsideration, Provocations.* New York: Oxford University Press. 1992

[58] Gaddis, John Lewis. "Conclusion" *Cold War Statesmen Confront the Bomb: Nuclear Diplomacy since 1945* editors John Lewis Gaddis, Phillip Gordon, Ernest May and Jonathan Rosenberg. Oxford: Oxford University Press. 1999. pp. 260 - 271.

[59] George, Alexander and Smoke, Richard. *Deterrence in American Foreign Policy.* New York, NY: Columbia University Press. 1974.

[60] Ghosen, Faten; Palmer, Glenn; and Bremer, Stuart. "The MID3 Data Set, 1993-2001: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science.* 21. 2004. pp. 133-154.

[61] Gibler, Douglas and Sarkees, Meredith. "Measuring Alliances: The Correlates of War Formal Interstate Alliance Dataset, 1816 - 2000," *Journal of Peace Research*, March 2004, pp. 211-222.

[62] Gibler, Douglas and Wolford, Scott. "Alliances, then Democracy: An Examination of the Relationship between Regime Type and Alliance Formation." *Journal of Conflict Resolution.* February 2006. pp. 1-25

[63] Goertz, Gary and Diehl, Paul. "The Empirical Importance of Enduring Rivalries" *International Interactions*, 1992 18(2). pp. 15163.

[64] Green, Donald; Kim, Soo Yeon; and Yoon, David H. "Dirty Pool." *International Organization.* Spring 2001. pp. 441-468.

[65] Grieco, Joseph. "Anarchy and the Limits of Cooperation: A Realist Critique of the Newest Liberal Institutionalism." *International Organization.* Summer 1988. pp. 485-507.

[66] Haas, Ernst. *When Knowledge is Power.* Berkeley, CA: University of California Press. 1990.

[67] Haas, Ernst. "Why Collaborate? Issue-Linkage and International Regimes." *International Organization.* April 1980. pp. 357-405.

[68] Haefele, E.T. "A Utility Theory of Representative Government." *American Economic Review.* June 1971. pp.350-367.

[69] Hays, Jude. "Globalization and Capital Taxation in Consensus and Majoritarian Democracies." *World Politics.* 56(3) 2003. pp. 79-113.

[70] Held, David, Anthony McGrew, David Goldblatt, and Jonathan Perraton. *Global Transformations.* Stanford, CA: Stanford University Press. 1999.

[71] Ho, Daniel; Kosuke Imai, Gary King, and Elizabeth Stuart. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15. 2007 pp. 199-236.

[72] Hoekman, Bernard M. ''Determining the Need for Issue Linkages in Multilateral Trade Negotiations", *International Organization* 43(4) 1989. pp. 693-714.

[73] Hoff, Peter. "Bilinear Mixed-Effects Models for Dyadic Data." *Journal of the American Statistical Association.* March 2005. pp. 286-295.

[74] Hoff, Peter and Ward, Michael. "Modeling Dependencies in International Relations Networks." *Political Analysis*. 12:2. 2004. pp. 160 - 175.

[75] Hufbauer, Gary Clyde; Schott, Jeffrey J.; Elliott, Kimberly Ann; and Oegg, Barbara. *Economic Sanctions Reconsidered* 3rd Edition. Peterson Institute. 2007

[76] Huth, Paul. "Extended Deterrence and the Outbreak of War." *American Political Science Review* 82. 2. 1988. pp. 423-443.

[77] Huth, Paul and Russett, Bruce. "What Makes Deterrence Work? Cases from 1900 to 1980." *World Politics*. July 1984. pp. 496-526.

[78] Imai, Kosuke; King, Gary; Stuart, Elizabeth. "Misunderstandings between experimentalists and observationalists about causal inference." *Journal of the Royal Statistical Society*. Vol. 171, Part 2. 2008. pp.1-22.

[79] Jahn, Detlef. "Globalization as 'Galton's Problem': The Missing Link in the Analysis of Diffusion Patterns in Welfare State Develoment." *International Organization*. 60, 2006. pp. 401-431.

[80] Jervis, Robert. *The Logic of Images in International Relations*. Princeton, NJ: Princeton University Press. 1970

[81] Jervis, Robert. "The Political Effects of Nuclear Weapons." *International Security*. 13 (2). pp. 28-38.

[82] Kadera, Kelly and Mitchell, Sara McLaughlin. "Manna from Heaven or Forbinnen Fruit? the (Ab)Use of Control Variables in Research on International Conflict." *Conflict Management and Peace Science*. 22, 2005. pp. 273-275.

[83] Kau, J.H. and Rubin, P.H. "Self-Interest, Ideology, and Logrolling in Congressional Voting." *Journal of Law and Economics*. October 1979. pp.365-384.

[84] Keele, Luke "An overview of rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data." 2009.

[85] Kenwood, A.G. and Lougheed, A.L. *The Growth of the International Economy, 1820-2000*. Fourth Edition. New York, NY: Routledge. 1999.

[86] Kennedy, Paul. *The Rise and Fall of British Naval Mastery* Second Edition. Humanity Books. 2006

[87] Keohane, Robert and Nye, Joseph. *Power and Interdependence*. Boston, MA: Little Brown. 1977.

[88] King, Gary. "Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Data." *International Organization*. Spring 2001a. pp. 497-507.

[89] King, Gary and Zeng, Langche, "Explaining Rare Events in International Relations," *International Organization*, Summer 2001b. pp. 693-715.

[90] King, Gary and Zeng, Langche, "Logistic Regression in Rare Events Data," *Political Analysis*, Spring 2001. pp. 137-163.

[91] Koremenos, Barbara; Lipson, Charles and Snidal, Duncan. "The Rational Design of International Institutions." *International Organization*. Autumn 2001. pp. 761-799.

[92] Koremenos, Barbara. "Contracting around International Uncertainty." *American Political Science Review*. November 2005. pp. 549-565.

[93] Kramer, G.H. "A Dynamic Model of Political Equilibrium." *Journal of Economic Theory.* December 1977. pp.310-334.

[94] Kuziemko, Ilyana and Werker, Eric. "How Much Is a Seat on the Security Council Worth? Foreign Aid and Bribery at the United Nations." *Journal of Political Economy.* 114:5 2006.

[95] Lai, Brian and Reiter, Dan, "Democracy, Political Similarity, and International Alliances, 1816 - 1992," *Journal of Conflict Resolution*, April 2000, pp. 203-227.

[96] Langer, William. *European Alliances and Alignments.* New York, NY: Alfred A. Knopf. 1966.

[97] Leeds, Brett Ashley. *Alliance Treaty Obligations and Provisions CodeBook.* 2005.

[98] Leeds, Brett Ashley; Ritter, Jeffrey M; Mitchell, Sara McLaughlin; and Long, Andrew, "Alliance Treaty Obligations and Provisions, 1815 - 1944," *International Interactions*, 2002, pp. 237 - 260.

[99] Leeds, Brett Ashley and Savun, Burcu. "Terminating Alliances: Why Do States Abrogate Agreements?" *The Journal of Politics.* November 2007. pp. 1118-1132.

[100] Long, Andrew and Leeds, Brett Ashley, "Trading for Security: Military Alliances and Economic Agreements," *Journal of Peace Research*, 2006, pp. 433-451.

[101] Limão, Nuno. "Trade Policy, Cross-border Externalities and Lobbies: Do Linked Agreements Enforce More Cooperative Outcomes?" *Journal of International Economics.* September 2005. pp. 175-199.

[102] Limão, Nuno. "Are Preferential Trade Agreements with Non-trade Objectives a Stumbling Block for Multilateral Liberalization." *Review of Economic Studies.* July 2007. pp.821-855.

[103] Lohman, Susanne, "New Games: Modeling Domestic-International Linkages," *Journal of Conflict Resolution*, February 1997, pp. 38-67.

[104] Long, Andrew; Nordstrom, Timothy; and Baek, Kyeonghi. "Allying for Peace: Treaty Obligations and Conflict between Allies." *Journal of Politics* 2007. 69:1103-1117.

[105] Lyall, Jason. "Do Democracies Make Inferior Counterinsurgents? Reassessing Democracy's Impact on War Outcomes and Duration." *International Organization.* Winter 2010. pp. 167-192.

[106] Mansfield, Edward; Milner, Helen; and Rosendorff, Edward. "Why Democracies Cooperate More: Electoral Control and International Trade Agreements." *International Oraganization* Summer 2002. pp. 477-513.

[107] Manski, Charles "Monotone Treatment Response." *Econometrica* 65:6 1997. pp.1311-1334.

[108] Marshall, Monty G. and Jaggers, Keith. *Polity IV project.* 2002.

[109] Martin, Lisa, "Credibility, Costs, and Institutions: Cooperation on Economic Sanctions," *World Politics*, April 1993, pp. 406-432.

[110] Maxwell, Stephen. *Rationality in Deterrence.* Adelphi Paper 50. Londo: International Institute for Strategic Studies. 1968.

[111] Mayer,Frederick. "Managing Domestic Differences in International Negotiations: The Strategic Use of Internal Side-Payments." *International Organization* 46 (4) 1992. pp. 793-818.

[112] Mayhew, David. *Party Loyalty Among Congressmen.* Cambridge: Havard University Press. 1966

[113] McGinnis, Michael, "Issue Linkage and the Evolution of International Cooperation," *Journal of Conflict Resolution*, March 1986, pp. 141-170.

[114] McKelvey, R.D. "Intransitives in Multidimensional Voting Models and Some Implications for Agenda Control." *Journal of Economic Theory.* June 1976. pp.472-482.

[115] Mearsheimer, John. *Conventional Deterrence.* Ithaca, NY:Cornell University Press. 1983.

[116] Mearsheimer, John. *The Tragedy of Great Power Politics* New York: Norton. 2001

[117] Mitchell, Brian R. *International Historical Statistics: 1750 - 1993* London:Macmillan. 1998

[118] Molinari, Francesca "Missing Treatments." *Journal of Business and Economic Statistics* January 2010. pp. 82-95.

[119] Moravcsik, Andrew. *The Choice for Europe: Social Purpose and State Power from Messin ato Maastricht.* Ithaca, NY: Cornell University Press. 1998.

[120] Morgan, T. Clifton, "Issue Linkages in International Crisis Bargaining," *American Journal of Political Science*, May 1990, pp. 311-333.

[121] Morgan, Stephen and Winship, Christopher. *Counterfactuals and Causal Inference: Methods and Principles for Social Research.* New York, NY: Cambridge University Press. 2007.

[122] Morgenthau, Hans. *Politics Among Nations* 1973.

[123] Morrow, James, "Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances," *American Journal of Political Science*, November 1991, pp. 904-933.

[124] Morrow, James, "Signaling Difficulties with Linkage in Crisis Bargaining," *International Studies Quarterly*, June 1992, pp. 153-172.

[125] Morrow, James D. "Alliances, Credibility, and Peacetime Costs." *Journal of Conflict Resolution* June 1994. pp. 270 - 297

[126] Mueller, D.C. "The Possibility of a Social Welfare Function: Comment." *American Economic Review.* December 1967. pp.1304-1311.

[127] Mueller, John. *The Remnants of War.* Ithaca: Cornell University Press. 2004

[128] Oatley, Thomas. *International Political Economy.* New York, NY: Pearson. 2004.

[129] Oneal, John and Russett, Bruce. "The classic liberals were right: Democracy, Interdependence, and conflict, 1950-1985." *International Studies Quarterly.* June 1997. pp. 267-293.

[130] Oneal, John and Russett, Bruce. "Rule of Three, Let it Be? When More Really is Better." *Conflict Management and Peace Science.* 22:4 2005. pp. 293-310.

[131] Organski, A.F.K. and Kugler, Jacek. *The War Ledger.* Chicago: University of Chicago Press. 1980.

[132] Oye, Kenneth. (ed) *Cooperation Under Anarchy.* Princeton, NJ: Princeton University Press. 1979.

[133] Oye, Kenneth. *Economic Discrimination and Political Exchange: World Political Economy in the 1930s and 1980s.* Princeton, NJ: Princeton University Press. 1993.

[134] Pahre, Robert. *Politics and Trade Cooperation in the Nineteenth Century: The 'Agreeable Customs' of 1815-1914.* Cambridge: Cambridge University Press. 2008.

[135] Pearson, Glenda. "Rohn's World Treaty Index: Its Past and Future." *International Journal of Legal Information.* Vol. 29:3 2001. pp. 543-559.

[136] Peceny, Mark; Beer, Caroline; and Sanchez-Terry, Shannon. “Dictatorial Peace?” *American Political Science Review*. March 2002. pp. 15-26.

[137] Powers, Kathy. “Regional Trade Agreements as Military Alliances.” *International Interactions*. 30. 2004. pp. 373 - 395.

[138] Powers, Kathy. “Dispute Initiation and Alliance Obligations in Regional Economic Institutions.” *Journal of Peace Research*. 43. 2006. pp. 453-471.

[139] Powers, Kathy and Goertz, Gary. “The Evolution of Regional Economic Institutions into Security Institutions.” Unpublished Manuscript. July 2006.

[140] Pöyhönen, Pentti. ”A Tentative Model for the Volume of Trade Between Countries.” *Weltwirtschaftliches Archiv* 90 (1), pp. 93-99. 1963

[141] Putnam, Robert, “Diplomacy and Domestic Politics: The Logic of Two-Level Games,” *International Organization*, Summer 1988, pp. 427-60.

[142] Ray, James Lee. “Constructing Multivariate Analyses (of Dangerous Dyads).” *Conflict Management and Peace Science*. 22, 2005. pp. 277-292.

[143] Reiter, Dan and Stam, Allan. *Democracies At War*. Princeton, NJ: Princeton University Press. 2002

[144] Reiter, Dan and Stam, Allan. “Indentifying the Culprit: Democracy, Dictatorship, and Dispute Initiation.” *American Political Science Review*. 97:2 2003. pp. 333-337

[145] Remmer, Karen. “Does Democracy Promote Interstate Cooperation? Lessons from the Mercosur Region.” *International Studies Quarterly*. March 1998. pp. 25-51.

[146] Riker, William and Brams, Steven. “The Paradox of Vote Trading.” *The American Political Science Review*. December 1973. pp.1235-1247.

[147] Robins, Garry; Morris, Martina. “Advances in exponential random graph models.” *Social Networks* 29. 2007. pp. 169-172.

[148] Rogowski, Ronald. *Commerce and Coalitions: How Trade Affects Domestic Political Alignments*. Princeton: Princeton University Press. 1989.

[149] Rosenau, James N. *Linkage Politics: Essays on the Convergence of National and International Systems*. New York: The Free Press. 1969.

[150] Rosenbaum, Peter. *Observational Studies*. 2nd ed. New York: Springer. 2002.

[151] Rubin, Donald. *Matched Sampling for Causal Effects*. New York, NY: Cambridge University Press. 2006.

[152] Schelling, Thomas, *The Strategy of Conflict*, Harvard University Press, Boston, MA, 1960.

[153] Schoeder, Paul. *The Transformation of European Politics, 1763 - 1848* Oxford: Oxford University Press. 1994.

[154] Schonhardt-Bailey, Cheryl. *Free Trade: The Repeal of the Corn Laws*. Thoemmes Press. 1996.

[155] Sebenius, James, “Negotiation Arithmetic: Adding and Subtracting Issues and Parties,” *International Organization*, Spring 1983, pp. 281-316.

[156] Sekhon, Jasjeet and Diamond, Alexis. “Genetic Matching for Estimating Causal Effects” Working Paper. 2008.

[157] Signorino, Curtis “Strategic Interaction and the Statistical Analysis of International Conflict.” *The American Political Science Review*. June 1999. pp. 279-297.

[158] Simmons, Beth. “International Law and State Behavior: Commitment and Compliance in International Monetary Affairs.” *American Poliitical Science Review.* Dec. 2000. pp. 819-835.

[159] Simmons, Beth. “Treaty Compliance and Violation.” *Annual Review of Political Science.* 2010. pp. 273-296.

[160] Singer, J. David and Small, Melvin. “Formal Alliances, 1815-1939: A Quantitative Description.” *Journal of Peace Research.* 3(1). 1966. pp.1-32.

[161] Smith, Alastair. “Alliance Formation and War.” *International Studies Quarterly.* December 1995. pp. 405 - 425.

[162] Spagnolo, Giancarlo. “Issue Linkage, Credible Delegation, and Policy Cooperation.” *CEPR Discussion Paper* 2778. 1999.

[163] Stam, Allan and McGillvray, Fiona. “Political Institutions, Coercive Diplomacy, and the Duration of Economic Sanctions.” *Journal of Conflict Resolution.* April 2004. pp. 154-172.

[164] Steglich, Christian; Snijders, Tom; and Pearson, Michael. “Dynamic Networks and Behavior: Separating Selection from Influence.” Working Paper. Interuniversity Center for Social Science Theory and Methodology. 2007

[165] Stein, Arthur, “The Politics of Linkage,” *World Politics*, October 1980, pp.62-81.

[166] Steiner, Zara. *The Lights That Failed: European International History, 1919-1933.* New York, NY: Oxford University Press. 2005.

[167] Stone, Randall. “The Scope of IMF Conditionality.” *International Organization.* 62:4 2008.

[168] Stone, Randall; Slantchev, Branislav; and Tamar, London. “Choosing How to Cooperate: A Repeated Public-Goods Model of International Relations.” *International Studies Quarterly* June 2008. pp. 335-362.

[169] Stratmann, Thomas. “The Effects of Logrolling on Congressional Voting.” *American Economic Review.* December 1992. pp.1162-1176.

[170] Stratmann, Thomas. “Logrolling in the U.S. Congress.” *Economic Inquiry.* July 1995. pp. 441-456.

[171] Starr, Harvey. “Cumulation from Proper Specification: Theory, Logic, Research Design, and ‘Nice’ Laws.” *Conflict Management and Peace Science.* 22, 2005. pp. 353-363.

[172] Taylor, A.J.P. *Struggle For Mastery In Europe, 1848 - 1918.* Oxford: Oxford University Press. 1954.

[173] Thacker, Strom. “The High Politics of IMF Lending.” *World Politics.* October 1999. pp. 38-75.

[174] Thompson, William. “Identifying Rivals and Rivalries in World Politics.” *International Studies Quarterly.* December 2001. pp.557-586.

[175] Tinbergen, Jan. *Shaping the World Economy: Suggestions for an International Economic Policy* New York: Twentieth Century Fund. 1962.

[176] Tollison, Robert and Willett, Thomas, “An Economic Theory of Mutually Advantageous Issue Linkages in International Negotiations,” *International Organization*, Autumn 1979, pp. 425-449.

[177] Tomz, Michael. *Reputation and International Cooperation: Sovereign debt across Three Centuries.* Princeton, NJ: Princeton University Press. 2007.

[178] Von Stein, Jana. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *American Political Science Review* November 2005. pp. 611-622

[179] Wagner, Harrison. "Economic Interdependence, Bargaining Power, and Political Influence." *International Organization*. Summer 1988. pp. 461-483.

[180] Wallace, William. "Atlantic Relations: Policy Coordination and Conflict." *International Affairs* April 1976.

[181] Walt, Stephen. *The Origins of Alliances* Cornell University Press. 1985.

[182] Waltz, Kenneth. *Theory of International Politics* McGraw-Hill. 1979.

[183] Wandycz, Piotr. *France and Her Eastern Allies, 1919 - 1925.* Minneapolis, MN: University of Minnesota Press. 1962

[184] Ward, Michael; Siverson, Randolph; and Cao, Xun. "Disputes, Democracies, and Dependencies: A Reexamination of the Kantian Peace." *American Journal of Political Science* July 2007. pp. 583-601.

[185] Warren, T. Camber, "The Geometry of Security: Modeling Interstate Alliances as Evolving Networks." *Journal of Peace Research* forthcoming 2009

[186] Wikenfeld, Jonathan (ed.) *Conflict Behavior and Linkage Politics.* New York: David McKay. 1973.