Assessing the Variation of Formal Military Alliances

Brett V. Benson[1]
and Joshua D. Clinton[1]

Abstract

Many critical questions involving the causes and consequences of formal military alliances are related to differences between various alliances in terms of the scope of the formal obligations, the depth of the commitment between signatories, and the potential military capacity of the alliance. Studying the causes and consequences of such variation is difficult because while we possess many indicators of various features of an alliance agreement that are thought to be related to the broader theoretical concepts of interest, it is unclear how to use the multitude of observable measures to characterize these broader underlying concepts. We show how a Bayesian measurement model can be used to provide parsimonious estimates of the scope, depth, and potential military capacity of formal military alliances signed between 1816 and 2000. We use the resulting estimates to explore some core intuitions that were previously difficult to verify regarding the formation of the formal alliance agreement, and we check the validity of the measures against known cases in alliances as well as by exploring common expectations regarding historical alliances.

Keywords

---

[1]Department of Political Science, Vanderbilt University, Nashville, TN, USA

Corresponding Author:
Brett V. Benson, Department of Political Science, Vanderbilt University, PMB 505, 230 Appleton Place, Nashville, TN 37203, USA.
Email: brett.benson@vanderbilt.edu

Understanding the formation and consequences of formal military alliances is a research program that is central to the study of international relations. Military alliances help define and shape the nature of interactions between countries, and by structuring international obligations, they help construct the nature of the international system. Better understanding not only the ways in which formal alliances vary but also why this variation occurs is critically important for understanding interstate relations. So too, of course, is the importance of exploring their actual impact on international behavior.

Formal military alliances differ a great deal. Consider, for example, the differences between the 1920 Franco-Belgian agreement and the 1915 multilateral alliance among France, Russia, United Kingdom, and Italy. Both contain defensive obligations, include France as an alliance member, target a specific adversary, and are signed in approximately the same time period. Beyond these similarities, the agreements differ a great deal. In addition to specifying precisely the amount of troops to be committed under certain conflict scenarios, the Franco-Belgian agreement publicly requires both sides to organize a system of defense including the merging of their military forces, the joint occupation of the Rhineland, and mutual assistance in the provision of war matériel. By contrast, the 1915 multilateral agreement was signed in secret during wartime and contains offensive and consultation obligations. Yet, the formal terms of the 1915 agreement do not specifically describe costs required of the signatories to organize and integrate their militaries.

How should we compare these two alliances? On one hand, the 1915 agreement obligates alliance members to take military action under a greater set of circumstances, because it contains commitments of military intervention under both offensive and defensive conditions. It also involves several relatively powerful countries as signatories. On the other hand, the 1920 Franco-Belgian accord requires alliance members to pay high costs to form the commitment by stipulating that both sides will actively integrate their militaries, jointly base their troops, and mutually develop and provide war matériel. Moreover, the agreement is a public declaration that may entail some reputation costs if alliance obligations are not kept. With so many points of difference, examining these two alliances together seems, in many ways, to be an apples-to-oranges comparison. Is there a straightforward way to compare various characteristics of formal military alliances that will help us begin to study the reasons for, and the implications of, core distinctions?

We focus on two motivating questions: First, can we use observable features of formal alliance agreements and characteristics of the signatories themselves to provide a more parsimonious characterization of the extent to which alliances differ from one another? Second, can we explore the variation to help us better understand the politics of alliance formation? Better understanding the willingness of signatories to enter into different types of alliances depending on their situation and that of their fellow signatories is important for understanding the construction of the international arena.

Our approach begins with the observation that formal military alliances vary in coherent ways. Alliances vary in the breadth of the circumstances to which the obligations of a military alliance have application (hereafter scope) as well as the costliness of the obligations to which signatories commit themselves when they join the alliance (hereafter depth). Agreements with sweeping offensive and defensive provisions, for example, are obviously broader in the scope of the military obligations contained in the terms of the agreement than neutrality pacts. Defensive commitments that formalize joint military planning as well as requirements for peacetime military integration, the provision of aid, and military basing impose deeper costs on the alliance members than agreements that only contain defensive obligations. Alliances also vary in terms of their pooled military strength (hereafter potential military capacity)—multiple powerful and ideologically aligned major powers likely have more potential military capacity than a bilateral alliance between minor powers with divergent preferences.

Identifying the range of alliances that exist and probing the conditions under which various types of alliances are likely to be formed along the dimensions of scope, depth, and potential military capacity of formal military alliances is key to understanding the role of military alliances in structuring the international system. Scholars have emphasized the importance of these concepts for characterizing and understanding the formation and consequences of military alliances (see Snyder 1997; Leeds et al. 2002; Schelling 1966; Benson 2011, 2012; and especially Leeds and Anac 2005). Research on these topics would benefit from a single parsimonious measure of each concept, but an empirical characterization remains elusive in spite of a wealth of data. The difficulty is due to the multifaceted aspect of these alliance features, which requires that any characterization is best inferred from multiple observable attributes.

We use a statistical measurement model to provide a principled framework for estimating the scope, depth, and military capacity of formal military alliances. Methodologically, our approach is similar to the approach taken by scholars interested in measuring the ideology of elected and unelected officials (e.g., Poole and Rosenthal 1997; Martin and Quinn 2002; Clinton, Jackman, and Rivers 2004), the positions of a political party in an underlying policy space (Budge et al. 2001), the ideology of a congressional district in the United States (Levendusky, Pope, and Jackman 2008), the extent to which a country is democratic (Pemstein, Meserve, and Melton 2010), the positions taken by a country in the United Nations General Assembly (Voeten 2000), or the depth of preferential trade agreements (Dur et al. 2014).

Our approach makes several contributions. First, we show how a Bayesian latent trait model (Quinn 2004) can recover a theoretically informed, multidimensional estimate of alliance variation that reflects the scope, depth, and potential military capacity of formal military alliances while also quantifying the sometimes-substantial uncertainty that we have about the resulting estimates (Jackman 2009b). We also demonstrate how the resulting measures quantify the influence of various features and allow us to discriminate between alliances belonging to broad classifications

that previous scholarship has been unable to disentangle when relying on particular observable alliance features (e.g., having to equate "offensive alliances" with more a expansive scope of commitment).

In validating our measures using both qualitative and quantitative information, we are also able to validate several important substantive conclusions regarding the politics of alliance formation. We demonstrate that notions regarding the scope, depth, and potential military capacity of a military alliance are distinct and meaningful measures. We also highlight that our measures of notable alliances comport well with the existing expectations, and we show how prominent alliances within a theater of operations show evidence of "balancing" in terms of both the potential military capacity of the signatories involved and the scope and depth of the formal treaty commitments.

We also find that many, but certainly not all, of the alliances that are exceptional in one respect are less so in others. This suggests that there are likely meaningful trade-offs associated with designing deeper and broader alliance agreements. For example, relatively weak signatories may occasionally wish to strengthen their alliance through deeper treaty terms that are designed to expand their combined capabilities through costly peacetime military coordination and sweeping wartime obligations, but militarily powerful alliance signatories may wish to curtail allies' access to the aggregate capabilities of the alliance by designing conditions that limit military intervention or make it costless to escape.

Finally, we use our measures to analyze the formation conditions of formal military alliances. That is, we examine whether factors that are predicted to affect the scope and depth of alliances when they are initially formed covary as expected. Our analysis not only helps confirm the validity of the measures that we recover but also contributes to our understanding of alliance formation. Whereas existing empirical work is often forced to deal with many coarse, but relevant alliance features, the measurement model we present summarizes the information contained in the multiple measures and, in so doing, provide a more nuanced characterization of the way in which alliance agreements vary.

Conceptualizing Variation in Alliances

A natural starting point for conceptualizing how formal military alliances vary involves considering the formal terms of the agreement and the characteristics of the signatories involved. Each of these, however, consists of many different factors and a wealth of available data. Our goal is to integrate the many measures that have been collected in a principled way so as to provide a meaningful characterization of alliance agreements that reflects variation along conceptually distinct dimensions. We focus on the following three dimensions: scope of the obligations, depth of commitment, and potential military capacity. Scope accounts for the variation in the breadth of the circumstances under which the terms of the agreement obligate alliance members to commit military action. Depth reflects the degree to which the alliance

agreement imposes peacetime and related costs on the signatories. Potential military capacity measures the total adjusted potential military power or strength of using characteristics of alliance members.

These three concepts are important for many fundamental questions related to military alliances. The scope of the obligations contained in alliance agreements is related to questions on the formation of military alliances. Scholars, for example, have argued that alliance members have incentives to limit their obligations when there are entrapment concerns (Snyder 1984, 1997; Fearon 1997; Zagare and Kilgour 2003; Kim 2011; Benson 2012). The depth of alliance commitments is also relevant for alliance formation, and some argue that incentives for opportunism may lead to deeper agreements (Leeds and Anac 2005; Abbott and Snidal 2000; Lake 1999). For example, Snyder (1997, 11) points out that states often include costly actions in their alliance commitments to ''validate'' agreements ''when interests of allies are somewhat divergent, or when they are subject to change. . . . '' According to Snyder, when signatories' preferences diverge, ''The partner will need to be reassured that one's underlying interests are not so at odds with the contract that the alliance is no more than a 'scrap of paper.''' In general, scholars believe that alliances entail deeper commitments to incur peacetime costs when concerns about opportunism exist and when signatories may have divergent preferences.

The concepts of scope and depth are also relevant for better understanding the effects of military alliances. Are agreements containing a broader set of agreed-upon obligations related to conflict initiation and war? Are deeper alliance agreements any more credible? If the imposition of costs signals information and facilitates the coordination of warfighting abilities (Morrow 1994; Fearon 1997), alliances that impose greater costs during peacetime may be more reliable.

The potential military capacity of an alliance is another critical concept. Balance of power theories, for example, considers how the distribution of power resulting from the joint military strength of competing alliance networks affects the stability of the international system (Morgenthau 1948; Organski 1968; Waltz 1979; Walt 1987), and there is a long debate with an extensive empirical literature about the effect of alliances in various theories of the distribution of power and stability.[1] Potential military capacity is also relevant for understanding the deterrent effect of military alliances and whether the combined potential military force of the allies might affect an adversary's calculation to challenge one of the allies (Morrow 1994; Smith 1995; Leeds 2003b; Zagare and Kilgour 2003; Yuen 2009; Johnson and Leeds 2011; Benson 2012; Benson, Meirowitz, and Ramsay 2014; Benson, Bentley, and Ray 2013).

Measuring the Scope of Obligations

Our characterization of scope draws on conventional conceptualizations. Scholars generally agree that alliance agreements typically specify the primary obligations of alliance members, some of which require members to become involved militarily

in a broad set of circumstances while others are more limited in scope. For example, Snyder (1997) explains that offensive alliance agreements obligate alliance members in a wide range of circumstances compared to those written to secure a third party's neutrality in the case of a military conflict. This is the standard view of alliances—agreements with offensive and defensive provisions obligate members to commit military action to a broader range of circumstances than defensive agreements alone, and defensive agreements are broader in military scope than, say, consultation pacts or neutrality agreements, which do not bind signatories to commit militarily to any conflict and may even require states not to become involved militarily.

Following the Alliance Treaty Obligations and Provisions (ATOP) categorization of obligations that commit alliance members to military action or nonaction (Leeds et al. 2002), we begin by including in our estimation of scope the variables from the ATOP project that indicate whether an alliance is offensive or defensive.[2] The difference between these alliances in the ATOP coding is that offensive provisions obligate at least one member to commit military support in conflicts involving an alliance member even if the conflicts were not triggered by an attack by a nonalliance member on an alliance member. Alliances coded as defensive include provisions that condition military action on an attack by a nonalliance member on an alliance member. We also include agreements with neutrality and consultation provisions, because such obligations require nonmilitary actions of alliance members. We do not include agreements that include only nonaggression provisions. These alliances may be fundamentally different in their design and purpose from those we aim to analyze (Mattes and Vonnahme 2010). Nonaggression pacts focus on avoiding war between signatories themselves, whereas the alliances we focus on include provisions that specify promises of actions related to conflicts with third parties.

In addition to these traditional classifications of alliance agreements, we follow ATOP rules and include variables indicating whether an alliance contains more specific conditional provisions. Accordingly, we include the following indicator variables: military action depends on war environment, nonmilitary action depends on war environment, military action depends on noncompliance, nonmilitary action depends on noncompliance, military action depends on nonprovocation, nonmilitary action depends on nonprovocation, nonmilitary action required only if requested, and conditional other. The two war environment variables indicate whether military (in the case of agreements where the primary obligations are offensive or defensive) or nonmilitary (in the case of agreements where the primary obligations are neutrality or consultation) action is conditional on some factor relating to the war environment such as a specific adversary, location, ongoing conflict, or number of adversaries. The two noncompliance variables indicate whether military/nonmilitary action is conditional on the noncompliance with a certain demand. The two nonprovocation variables indicate whether military/nonmilitary action is conditional on one of the alliance members being attacked without provocation. The variable nonmilitary action required only if requested indicates whether consultation is required only if

requested by an alliance member. The variable conditional other refers to any other condition specified in the agreement that is not covered by these other variables. Each of these provisions delineates scope conditions for military action.

We also include the ATOP variables that identify conditions in the agreement for renouncing the obligations. These variables include renunciation allowed, renunciation prohibited, and renunciation conditional. Renunciation allowed permits any alliance member to renounce at any time without advance notice. Renunciation prohibited stipulates that renunciation is strictly prohibited. Renunciation conditional indicates whether an agreement permits renunciation if another member takes an aggressive action. Renunciation conditions are critical for informing the conception of scope, because they stipulate whether the primary obligations for military or nonmilitary action are firm or flexible. Flexibility of terms may limit the scope. Comparing two offensive agreements similar in every way except one permits renunciation and another does not, the more flexible agreement may be more limited because it allows for the inapplicability of military action conditional on factors that may be determined by the alliance members on an ad hoc basis.

The scope of the obligations agreed to in an agreement also varies depending on whether the alliance is designed to deter or to compel. Schelling (1966) distinguished between deterrent military commitments that are intended to prevent changes to the status quo and compellent commitments that are designed to induce or coerce changes in the status quo. Compellent threats condition military action on noncompliance with an explicit or implicit demand on the target to make a concession. Consequently, following Benson (2012), we include a variable indicating whether an agreement includes a compellent provision. We also use Benson's (2012) conceptualization of probabilistic commitments to create a variable indicating whether an alliance agreement is deterministic. Our coding indicates whether any agreement promising active military support allows escape through voluntary or probabilistic reasons (Benson 2012). The reason for including this variable is similar to the justification provided for including the renunciation conditions. An alliance is considered deterministic if it commits members to military action for either compellent or deterrent purposes without the option of escape. We also include Benson's (2011, 2012) coding of unconditional alliances. An alliance is considered unconditional if it commits members to military action for either compellent or deterrent purposes without any conditions on *casus foederis*. It is reasonable to expect that military obligations that do not allow for escape and do not impose conditions on military involvement beyond specifying whether the objective is compellence or deterrence likely bind members to a broader range of military circumstances than those that impose conditions on military action or allow members to escape.

Measuring the Depth of Commitments

Many scholars agree that alliances also vary in the depth of the commitments included in the agreement. That is, the content of formal alliances differs in the

degree to which the agreement contains formalized and binding commitments.[3] Our conception of the depth of an alliance commitment follows from the view that alliance commitments themselves impose varying levels of costs on alliance members beyond those associated with the risks of conflict. Such costs include sunk formation costs (Fearon 1997; Smith 1995) and peacetime military coordination costs (Morrow 1994; Snyder 1997). Deeper alliance agreements impose higher costs while shallower commitments impose lower costs.

To estimate a single measure of depth, we use several existing variables of different provisions in alliance agreements that impose costs on alliance members. Building on Leeds and Anac (2005), we include in our characterization of depth several measures of provisions that formalize the imposition of peacetime costs and that institutionalize certain aspects of a military alliance. Accordingly, we include the following variables: military contact, common defense policy, integrated command, military aid, military basing, specific contribution, organization, economic aid, and secret. Military contact indicates whether the agreement requires contact between the militaries of the alliance members during peacetime. Common defense policy indicates whether alliance members are required to conduct a common defense policy including common doctrine, coordination of training and procurement, joint planning, and so on. Integrated command is a variable that indicates whether alliance members are obligated to integrate military command among allies both in peacetime and in wartime. Military aid indicates whether alliance members are required to provide unspecified military aid, grants or loans, and/or military training or technology transfer. The variable military basing captures whether the alliance agreement contains provisions stipulating that members agree to jointly place troops in a neutral territory or station troops in a members' territory. Specific contribution indicates whether the agreement specifies details about the contributions to be made by one or more of the allies or how the costs of the alliance should be divided. Organization is a variable that specifies whether the agreement requires members to create any stand-alone organizations or other organizations that provide for regular meetings of government officials or the named organization. Economic aid indicates whether the agreement includes obligations for providing economic aid for nonspecific reasons, postwar recovery, or for trade concessions. Secrecy is a variable that specifies whether alliances are public, public but contain some secret provisions, or totally secret.

Measuring the Potential Military Capacity

Conventional approaches for measuring the potential joint military capacity of an alliance include summing the capabilities of the alliance partners using the Composite Index of National Capabilities (CINC scores; Singer, Bremer, and Stuckey 1972). Accordingly, we include capabilities, which is the log of the summed CINC scores of all alliance members. While aggregate capabilities provide one estimate of the raw potential military capability of the alliance, other

factors related to specific characteristics of the signatories might also enhance or constrain the military capacity of the alliance. We also include major power, which is a variable indicating whether at least one alliance member is a major power (Correlates of War 2011). The presence of a major power in an alliance may affect the overall military capacity of an alliance. Scholars claim that major powers possess unique characteristics that give such alliances a distinctive military advantage—for example, they possess significantly greater economic resources, have more economic and security interests, possess advanced weapons systems (such as nuclear weapons in the post–World War II [WWII] era), and influence in international institutions such as the United Nations Security Council (Gibler and Vasquez 1998). Although scholars generally agree that major powers are qualitatively distinct from other powers, measuring their impact on the military capacity of an interstate alliance is not straightforward. Scholars typically use a separate indicator to control for the presence of a major power (Levy 1981; Siverson and Tennefoss 1984; Morrow 1991; Leeds 2003a; Benson 2011).

The distance between alliance members is another factor that might influence the overall potential military capacity of an alliance. In particular, the distance between allies may degrade the signatories' combined capabilities because of the cost related to projecting military forces and coordinating long-distance military actions (Boulding 1962; Starr and Most 1976; Bueno de Mesquita 1983; Bueno de Mesquita and Lalman 1986; Smith 1996; Weidmann, Kuse, and Gleditsch 2010; Bennett and Stam 2000b). We include a variable for distance (Bennett and Stam 2000a). Our measure includes the mean of all paired combinations of alliance members.[4] The distance to a target country may also be relevant for assessing the strength of some alliances, but given the difficulty of identifying threats and the aspiration to estimate the strength of alliances lacking a specified threat, we omit this variable.[5]

Because the size of the alliance may also be a correlate of its military capacity, we also include a measure for ally count, which is the log of the number of alliance members according to ATOP (Leeds et al. 2002). Multiple signatories may provide advantages in conflict bargaining, yield potential gains from division of labor and specialization, and enhance the credible use of allies' military capabilities beyond the additive advantage of simply summing individual military capabilities. However, the relationship is somewhat unclear as more signatories may complicate logistical coordination and increase the chances of that allies' opinions and interests will diverge.

Finally, the commonality of security interests may affect the resolve of signatories to contribute in a war. One common approach is to use "s-scores" to measure of the closeness of foreign policy interests (Signorino and Ritter 1999) based on the similarity of countries' alliance portfolios. We include a measure for sglo, which we calculate by taking the mean s-score of all paired combinations of alliance members. Regime type may also affect signatories' willingness to contribute in war. More democratic alliances may or may not also produce stronger alliances. Democratic

allies may lead to common security interests, and domestic audience costs may lead democratic alliances to be more credible (Lai and Reiter 2000; Leeds et al. 2002; Gibler and Sarkees 2004; Mattes 2012). On the other hand, the relationship is not entirely clear because democracies may prefer not to ally with one another (Simon and Garzke 1996; Gibler and Wolford 2006) because the veto points created by domestic political institutions may create difficulties for taking action (e.g., Tsebelis 2002) or because election-induced leadership turnover may make them unreliable (Gartzke and Gleditsch 2004). We include a measure of how democratic, on average, the signatories are. We calculate POLITY by taking the mean Polity IV score for all alliance members (Marshall, Jaggers, and Gurr 2002).

A Statistical Measurement Model

The challenges scholars confront when characterizing the nature of military alliances are endemic to the social sciences. How do we use several multiple observable features to estimate a parsimonious measure that summarizes the structure of the common variation we observe? We know, for example, that an alliance that commits the signatories to establish joint military bases, integrate military commands, and provide economic and military aid is one that requires deep commitments. How do we compare such an alliance with such provisions to an agreement made in secret and which requires ongoing military contact and establishes a formal organization? All of these aspects are related to the depth of the commitment entailed by the alliance, but it is not clear which set of obligations imposes greater costs. Similar difficulties arise when describing the scope of an alliance—is an offensive alliance with specific conditions for its invocation broader in scope than a unconditional and open-ended defensive pact? Even assessing the potential military capacity of an alliance can pose difficulties when making comparisons—does an alliance between two major powers with very dissimilar alliance portfolios have a greater military potential than an alliance among many like-minded countries that are located in close proximity to one another?

Scholars currently have three unsatisfying options for resolving these measurement issues. One possibility is to focus the analysis on a single proxy variable at a time. This approach is problematic because it fails to account for the considerable variation that may exist within the values of the chosen variable. For example, while the terms of the average offensive alliance may obviously reflect broader commitments than the terms of an average neutrality agreement, there may still be important variation within alliance agreements that are categorized as offensive. For example, the 1939 Pact of Steel alliance between Germany and Italy is widely viewed as an example of an aggressive military alliance with sweeping terms. As offensive alliances goes, it is indeed both expansive and relatively deep, obligating alliance members to commit military support under both offensive and defensive circumstances as well as establishing ''standing committees'' in both countries for the purpose of

remaining in constant military contact and consulting about actions to take if ''the common interests of the contracting parties [are] injured.''

It may be difficult to imagine an offensive alliance requiring more of its members than the Pact of Steel. There are examples of offensive alliances that are more explicit in imposing specific costs. There are, as well, alliances that contain similarly aggressive text but contain few to no costly provisions beyond the primary offensive and defensive obligations. The 1941 WWII Axis agreement among Germany, Italy, and Japan is an example of the latter category. It simply states that ''Germany, Italy and Japan jointly and with every means at their disposal will pursue the war forced upon them by the United States of America and Britain to a victorious conclusion.'' The content of agreement does not contain provisions that impose additional costs such as joint military contact, military aid, basing, or integrated military command. Of course, the agreement was signed during WWII, and so it may make sense that the signatories would not specify peacetime costs in the agreement and wartime costs may be implied by the phrasing of the primary obligation: ''with every means at their disposal will pursue the war . . . '' But a few years later in 1944, the United Kingdom and Ethiopia also signed an offensive alliance during WWII, which exceeded both the Pact of Steel and the WWII Axis in the level of detail it used to specify the explicit obligation created by the alliance. In addition to the primary offensive and defensive obligations, it also required signatories to maintain official military contact in both wartime and peacetime; it stipulates that the United Kingdom would organize, train, and administer the Ethiopian army; and it allows for British basing in Ethiopia. This variation in the depth of the agreement terms across these three ''offensive'' alliances underscores the challenges associated with using a single coarse variable to proxy for concepts with more subtle distinctions.

A second approach for accounting for the variation in alliances is to create an additive index based on multiple characteristics. This method is problematic because there is no theoretical guidance for combining the measures or interpreting the resulting scale. Even if we think that the establishment of joint military bases, integrated military commands, and the provision of economic and military aid signals a deeper level of commitment between signatories than an alliance that lacks these features, how do we evaluate the magnitude of the differences in the level of depth? For example, is an alliance with two of these features twice as ''deep'' as an alliance with only one? Moreover, how do we compare an alliance that commits signatories to both economic and military aid to one that only provides for joint military bases and an integrated military command? It seems difficult to rationalize the relationships that are assumed by an additive index, and the assumed equivalences increase as the number of variables used to construct the measure increases.

A third approach is to use a regression specification to predict the effect of some features of an alliance on an outcome of interest y, while controlling for multiple other features relevant to the concept in question. For example, if we are predicting the effect of alliance scope, for example, on outcome y, the typical regression

specification $y = \alpha + \beta_1 x_1 + \beta_2 x_2$ allows the left-hand side to measure the scope of an alliance—as a linear function of x_1 and x_2—and its relation to y.[6] Including multiple measures in a regression changes measurement issues into specification issues. Additionally, including a host of variables to account for variation in the nature of alliances may also adversely affect the number of degrees of freedom that scholars have, given the number of potential indicators of alliance strength. Interpreting the effects from a saturated regression model can pose difficulties (Ray 2003; Achen 2005), particularly if the model includes multiple interactions (Braumoeller 2004; Brambor, Clark, and Golder 2006).

If the question of interest relates to alliance formation—for example, what accounts for the willingness of signatories to sign wide-ranging alliances rather than more limited alliances—the regression approach provides no help. Scholars interested in such questions are forced to choose to focus on a particular measure (e.g., Benson 2012) despite knowing that any single measure is an imperfect proxy. Multiple regressions using multiply proxies may obviously be run, but there is no simple way of providing a parsimonious assessment of the relationship of interest in such circumstances. Moreover, a shortcoming of all of these approaches is that, as indirect measures of a latent concept, they all fail to reflect our uncertainty about how the observed concepts relate to the actual concepts of interest and to account for the precision with which we are able to characterize such concepts.

In contrast to these three approaches, a Bayesian latent variable model provides a principled framework for extracting concepts that are theoretically related to observable features of alliances using weaker assumptions than the alternatives noted above. Non-Bayesian methods are certainly available, but for both theoretical (see the arguments of Gill 2002 and Jackman 2009a) and practical reasons we adopt a Bayesian approach. Most notably, unlike a frequentist approach, a Bayesian latent variable approach allows us easily to quantify our uncertainty about the resulting estimates using the posterior distributions of estimated parameters.

To focus our exposition, suppose we are interested in measuring the scope of alliance obligations and let x_i^* denote the scope of alliance i at the time of its formation. Even if we cannot know the actual scope, we can use characteristics of the agreement that are theorized to be related to the scope of alliance i—for example, whether the alliance is a commitment to offensive actions, whether there are specific conditions placed on the commitment, and whether signatories are committed to military action without the flexibility of escape—to estimate x_i^*. In so doing, we want to describe the relative scope of various alliances and also quantify our level of uncertainty about these characterizations. If we have $k \in 1 \ldots K$ observable measures of a dimension of interest, let the observed value for variable k for alliance i be denoted as x_{ik}. Observed measures may include continuous (e.g., the average distance between signatories), binary (e.g., whether a major power is involved), and ordinal measures.

The statistical measurement model we use to relate observable aspects to the underlying dimension of interest is identical to that used to estimate how
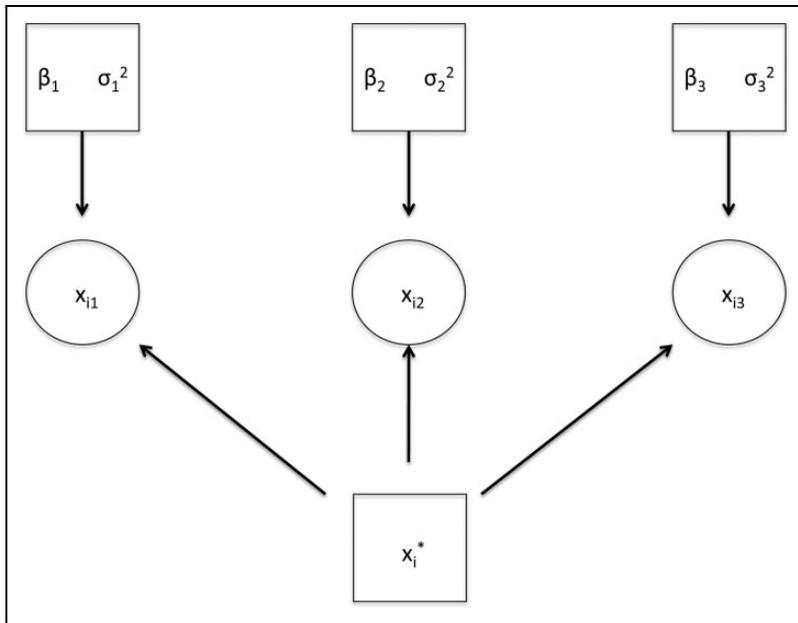
Figure 1. Directed Acyclic Graph: Bayesian latent variable model.
Note: Circles denote observed variables and squares denote parameters to be estimated.

democratic a country is or how liberal a district or a member of the US Congress is. A Bayesian latent variable model provides a principled way of relating observed features to a latent dimension that is thought to be responsible for generating the association between the observed characteristics (see, e.g., Quinn 2004; Jackman 2009b). The idea is neither new nor controversial, and while these models have been used to measure concepts critical for studying the politics of the United States (e.g., Clinton and Lewis 2008; Levendusky and Pope 2010) and comparative politics (e.g., Rosenthal and Voeten 2007; Rosas 2009; Pemstein, Meserve, and Melton 2010; Treier and Jackman 2008; Hoyland, Moene, and Willumsen 2012), scholars have only recently begun to apply the models to concepts in international relations (see, e.g., Schnakenberg and Fariss 2014; Gray and Slapin 2011; Dur et al. 2014).

Figure 1 provides a graphic representation for the three measures to provide an intuition for the measurement model. As Figure 1 makes clear, the model assumes that x_i^* is related to x_{i1}, x_{i2}, and x_{i3} across alliances, but it allows the relationship to differ between variables. For example, x_{i1} and x_{i2} may be related to x_i^* in different ways, and these differences are captured by β_1, β_2, σ_1^2, and σ_2^2.

Given the number of estimated parameters, estimating alliance characteristics (x^*) from the matrix of observed characteristics (x) requires additional structure. The

Bayesian latent variable specification (see, e.g., Jackman 2009a, 2009b) we use assumes that for all alliances:

$$x_i \sim N\big(\beta_{k0} + \beta_{k1} x_i^*, \sigma_k^2\big). \tag{1}$$

The measurement model of equation (1) assumes that the observed correlates of the alliance characteristic x are related to that characteristic in identical ways across the N alliances but that different measures may be related to alliance strength in different ways. For example, the relationship between the scope of an alliance and whether it entails offensive characteristics is identical across alliances—that is, β_{k1} does not vary by i—but offensive objectives may be more related to the scope of an alliance than whether there are specific provisions regarding the conditions under which the agreement may be renounced by the signatories (i.e., β_{k1} may be greater than β_{k2}).

The model allows for differences in both the mean value of the observed measure x_k and the latent concept x^* (as this will be reflected in the estimate of β_{k0}), and it also allows the scale of the observed and latent variables to differ (accounted for by β_{k1}). The β parameters therefore allow us to probe whether observed factors are related to the underlying concept of interest—$\beta_{k1} > 1$ implies that a one-unit change in the latent scale of x^* corresponds to more than a one-unit change in the observed measure x_k, $\beta_{k1} < 1$ implies that a one-unit change in the latent scale corresponds to less than a one-unit change in the observed measure, and $\beta_{k1} < 0$ implies that positive values of x_k correspond to negative values of x^*. The model can also account for the possibility that a measure is unrelated to the latent dimension, if $\beta_{k1} = 1$. In addition to the estimating the nature of the correlation between the observed and unobserved variables, the σ_k^2 term allows for varying amounts of error in the precision of this relationship. Finally, because we estimate a version of equation (1) for each of the K observed measures, the relationship may vary across observed traits, and we can use all available measures to help uncover the underlying latent trait.

These assumptions are silent about causality—nothing requires that the latent trait x^* causes the observed phenomena or vice versa. All that is assumed is that there is a correlation between the observed and unobserved traits that can be used to learn about the unobserved trait. For example, the Unified Democracy Scores of Pemstein, Meserve, and Melton (2010) measure democracy using twelve expert assessments even though the analyzed experts certainly do not cause democracy. Similarly, Levendusky, Pope, and Jackman (2008) use various aspects of a congressional district that are related to district ideology but which do not necessarily cause it.

Given the unknown parameters x^* and β to be estimated from the observed covariate matrix x, the likelihood function is:

$$L(x*, \beta) = p(x|[x*, \beta]) \propto \Sigma_{i=1}^{N} \Sigma_{k=1}^{K} \phi\left(\frac{x_i - (\beta_{k0} + \beta_{k1} x_i^*)}{\sigma_k}\right), \tag{2}$$

where $\phi(\bullet)$ is the probability density function of the normal distribution. To complete the specification and form the posterior distribution of x^* and β, we assume standard diffuse conjugate prior distributions.[7]

Given the discussion in the first section, we seek to characterize alliances (x^*) along three dimensions. Let $x[1]_i^*$ denote the potential military capacity of the alliance (with estimates given by $\hat{x}[1]_i$), let $x[2]_i^*$ denote the depth of the alliance commitments created by the provisions of the alliance (with estimates $\hat{x}[2]_i$), and let $x[3]_i^*$ denote the scope of conditions falling under the purview of the alliance (with estimates $\hat{x}[3]_i$). To identify the center of the latent space, we innocuously assume that the means of $x[1]^*$, $x[2]^*$, and $x[3]^*$ are all 0. To fix the scale of the recovered space, we assume that the variance of $x^*[1]$, $x^*[2]^*$, and $x^*[3]^*$ are also all 1. To fix the rotation of the space and define the meaning of positive values, we assume that higher values of the summed capacity of signatories correspond to positive values in the first dimension, alliances that stipulate for joint military bases reflect a deeper and more costly commitment for signatories, and compellent alliances receive positive values in the third dimension.

We do not need to know the precise nature of the relationship between the observed characteristics and the strength of the alliance to implement the model, but we do need to identify which measures are, and are not, potentially related to each of the three dimensions we are interested in. Following the discussion in the first section, for every characteristic pertaining to either the depth or the scope of the commitments created by the alliance, we assume that $\beta[1] = 0$; for any characteristic not related to the depth of the alliance, we assume that $\beta[2] = 0$; and for every characteristic not related to the scope of the alliance, we assume that $\beta[3] = 0$. That is, to define the meaning of the dimensions we recover, we assume that only those measures that are thought to be theoretically related to the dimension influence the estimated alliance score on that dimension.

Because we identify the latent dimensions by making assumptions about alliance characteristics, our statistical measurement model can shed important insights into the relationship between the these alliance characteristics and a question we consider later is how alliance characteristics $x^*[1]$, $x^*[2]$, and $x^*[3]$ are related.

Given these measures and identification constraints, we use the Bayesian latent factor model that can accommodate both continuous and ordinal measures described by Quinn (2004) and implemented via MCMCpack (Martin, Quinn, and Park 2011). We use 100,000 estimates as "burn-in" to find the posterior distribution of the estimated parameters, and we used one of our every 1,000 iterations of the subsequent 1,000,000 iterations to characterize the estimates' posterior distribution. Parameter convergence was assessed using the diagnostics implemented in CODA (Plummer et al. 2006).

Estimates of Alliance Features

Our Bayesian latent variable model not only produces estimates about the scope of the obligations, depth of the commitments, and the potential military capacity but also reveals how the various observable features described in the first section are related to each. Exploring these relationships helps assess the construct validity of our measurement model.

Estimating three dimensions of alliance characteristics enables the inspection of the relationship between alliances across dimensions to locate the estimated alliance scores in the recovered space. We began this article by considering the difficulty in making comparisons between alliances such as the 1920 Franco-Belgian accord and the 1915 alliance between France, UK, Russia, and Italy. Using our measures, we find that the 1915 alliance is among the most wide-ranging alliance agreements with an estimated scope score of 2.40 (on a variable that is assumed to have a mean of 0 and a standard deviation of 1). It was formed during World War I, contained both offensive and defensive provisions, and did not impose limiting conditions on alliance members' use of military force to achieve the war objective. Yet, with a depth score of only 0.225, it was not a deep agreement in the sense that it did not formalize costly commitments regarding military integration and coordination. In contrast, the content of the 1920 Franco-Belgian accord provided a depth score of 2.63 and a scope score of only 0.207. Accordingly, the terms of the 1920 agreement formalized many costly commitments that the 1915 agreement lacked, but it was not nearly as sweeping in terms of the circumstances under which alliance members' were obligated to use actual military force.

Figure 2 plots the distribution of alliance estimates in the dimensions of *potential military capacity* (x - axis and the *scope of the obligations* contained in the alliance agreement (y - axis). A score is estimated for each of the 489 alliances signed between 1816 and 2000 for which we have data on the observable characteristics (plotted in gray), but we focus our attention on a few selected alliances to illustrate the face validity of our estimates. (The Online Appendix contains the full set of estimates and standard errors, and it also contains an extensive discussion of alliances in the World Wars and East Asia along those two dimensions.)

As Figure 1 shows, the most powerful alliance in terms of potential military capacity is the Allied agreement in WWII. This alliance is a joint declaration by 39 countries, including the United States, Russia, the United Kingdom, and China. It is notably also one of the most sweeping agreements on the dimension of scope. The terms of the agreement explicitly target Germany, Italy, and Japan and contain no limiting conditions on the scope of alliance members' military obligations. It is a broad declaration of war in both offensive and defensive circumstances. The estimated location of this alliance is a reassuring starting point for assessing the output of the estimates. We might expect that the winning alliance in the world's most widespread and deadliest war—one that included all of
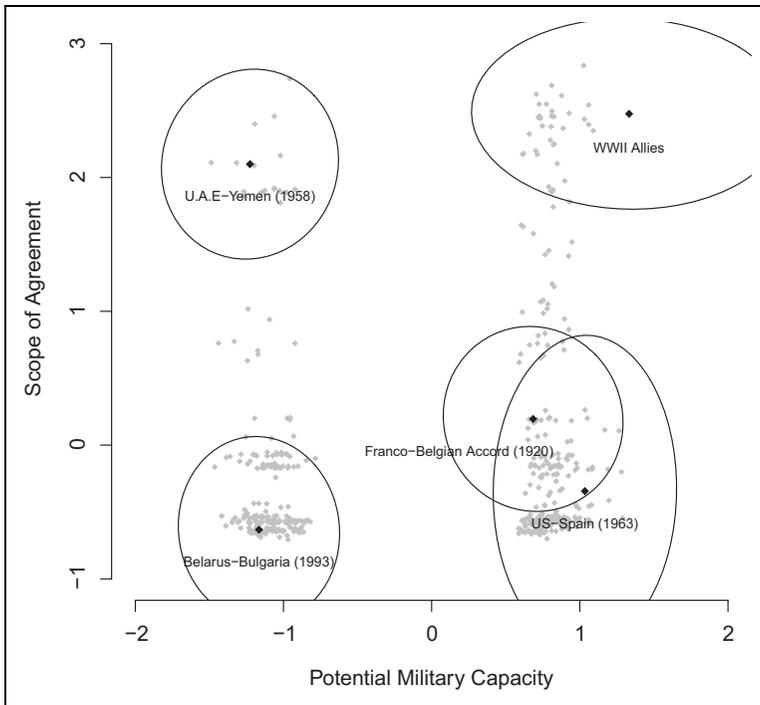
Figure 2. Potential Military Capacity and Scope of Alliance Agreements, 1815–2000.
Note: Points denote the posterior mean of the estimated alliance strength of each of the 489 alliances we analyze. The ellipses denote the 95 percent regions of highest posterior density for the selected alliances.

the world's major powers—is likely to contain some of the most aggressive obligations and rank among the most militarily powerful alliances in history.

The estimate of other notable alliances further reflects reassuring differences in Figure 1. Note the orientation of the 1958 United Arab Republic–Yemen (UAR) alliance compared to the WWII Allied agreement. While the scope of the UAR contains similarly broad military obligations, it does not have nearly same military might as the WWII Allied agreement. The UAR agreement, which included Egypt, Syria, and Yemen, was formed to unite the Arab community against the expansion of communism in Syria and elsewhere in the Arab world (Walt 1987, 71-80). Gamal Abdel Nassar, former president of Egypt and the president of the United Arab Republic, insisted on a full union and control over both countries in exchange for his agreement to use all offensive and defensive military capabilities to halt the rising influence of the Syrian Communist Party. Nassar then seized the control of Syria and banned all political parties.

Even though the UAR alliance was weaker in terms of potential military capacity than most other alliances in the data, it was sufficiently powerful to satisfy the

primary objective of the Union, which was to crack down on the Syrian Communists. It is noteworthy also that Jordan and Iraq formed the short-lived Iraq–Jordan Federal Union at the same time. According to Walt (1987), the Federal Union was designed to balance against the UAR.[8] Our measure suggests that the estimates for the potential military capacity of the UAR and Federal Union were indeed evenly matched. In fact, our measurement model estimates that the UAR and Federal Union are nearly indistinguishable on all three dimensions—scope, depth, and potential military capacity.

In contrast to both the WWII Allies and the UAR, the 1963 US–Spain alliance is relatively limited in scope. In 1963, both governments codified a formal defense pact, which was an extension of their defensive obligations formed in the Madrid Pact of 1953. This alliance was replaced by new bilateral alliances in 1970 when the scope of military obligations was further limited by the inclusion of a consultation provision and then again in 1976. In 1982, Spain officially joined North Atlantic Treaty Organization (NATO). As depicted in Figure 2, the scope of the US–Spain alliance is relatively limited even though the alliance was relatively strong in terms of its potential military capacity. Not only does it not contain offensive provisions, but it also provides flexibility and the potential for limiting military action on alliance members' primary obligation. The primary obligation states that "a threat to either country, and to the joint facilities that each provides for the common defense, would be a matter of common concern to both countries, and each country would take such action as it may consider appropriate within the framework of its constitutional processes." This commitment is soft both on the amount of military support to be provided and whether military action might even be guaranteed.

Also at the lower end of the scope dimension is the 1993 Belarus–Bulgaria agreement. Although it is significantly less militarily powerful than the US–Spain alliance and even though it is classified by ATOP as having no defensive or offensive military obligations (in contrast to the US–Spain alliance which is classified as defensive), the Belarus–Bulgaria accord, like the US–Spain agreement, is limited in scope. The formal terms of the primary obligation state, "In case of a situation which, according to one of the parties threatens its security or world peace, the parties promise to hold immediate consultations on the measures of resolution of such a situation." While it is clear that the agreement aims at resolving security threats as they arise, the agreement does not specify military obligations that members must follow.

We now evaluate how notable alliances compare to one another when we relate the potential military capacity of the signatories to the depth of the alliance agreement. Figure 3 presents this relationship and reveals a clear and meaningful variation in the alliances we observe along these two dimensions. First, consider again the comparison between the WWII Allied agreement and 1958 UAR Union. Whereas before we saw that the terms of these two agreements were similar in scope, the depth of the two agreements differs significantly. The terms of the UAR agreement called for the integration of the allies' militaries, unified command over those forces,
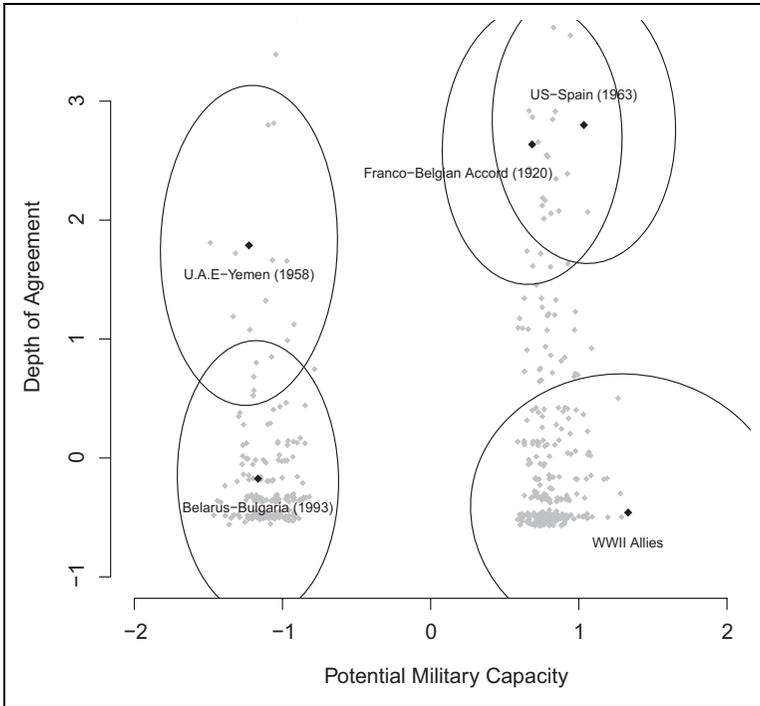
Figure 3. Potential Military Capacity and Depth of Commitment, 1815–2000.
Note: Points denote the posterior mean of the estimated alliance strength of each of the 489 alliances we analyze. The ellipses denote the 95 percent regions of highest posterior density for the selected alliances.

and the full grant of military power and foreign policy authority to a unified command. By contrast, the Allied agreement contained no such terms.

Much of this difference is likely due to the fact that the Allied agreement was signed in wartime, making the institutionalization of costly peacetime organizations and military integration less relevant. However, the presence of war does not always account for the lack of the depth of different agreements. As an illustration of this point, note that the 1944 alliance between the United Kingdom and Ethiopia was also signed during WWII. It is approximately equal to the WWII Allied agreement in terms of scope (with a scope score of 2.51). Yet, even as a wartime agreement, it ranks among the deepest agreements (with a depth score of 2.06), committing members to expend costs to maintain military contact in both wartime and peacetime. Additionally, the agreement specified that British must organize, train, and administer the Ethiopian army, while the Ethiopian government will cede certain of its territory to be administered by the British military.

Another interesting comparison is the 1963 US-Spain alliance and the 1993 Belarus–Bulgaria accord. We saw that both alliances were at the low end of the

scope dimension but significantly different from one another in terms of potential military capacity. Figure 3 shows that they are different in terms of both depth and military capacity. Whereas the Belarus–Bulgaria agreement includes no terms for the expenditure of peacetime costs, the US–Spain alliance specifies that both militaries will remain in contact and requires the US government to provide military equipment as well as funding for Spain's defense. Additionally, it calls for the formation of a consultative committee between the two sides. In terms of the depth of the commitment, the Belarus–Bulgaria agreement is much more similar to the WWII Allied agreement, in spite of having fundamentally different military obligations, than the US–Spain alliance is to the WWII Allied agreement, despite the fact that both of these latter two alliances include the US as a signatory and, therefore, contain significant potential military capacity.

Components of Alliance Dimensions

To validate our measure further it is instructive to compare how the various inputs included in the model relate to the estimated dimensions. By doing so highlights the following two advantages of our measurement approach: (1) our estimates reflect the characteristics of multiple measures without having to specify the precise nature of the relationship and (2) we can recover variation between alliances who possess the same value for any particular measure because our measure does not necessarily depend on a single measure (e.g., "offensive alliances"). We are particularly interested in assessing the inputs of the dimensions of scope and depth, since these measures are truly novel.[9]

Figure 4 reveals that many aspects that are thought to be related to the depth of an alliance commitment are related to the estimated dimension in expected ways. In particular, the most prominent feature of an alliance with costly obligations is the requirement for extending military aid and establishing military bases. These two provisions are followed closely by maintaining military contact, establishment of a formal organization, the specification of particular obligations, and an integrated military command.

Terms of an alliance agreement that provide for economic aid and the secrecy of the agreement have either little or no relationship to the depth of the obligations specified by the alliance according to our measurement model. A feature of our measurement model is that the variability in the relationship between these measures and the underlying dimensions means that we can use these results to compare the depth of alliance agreements containing different sets of obligations—an agreement specifying military aid and a integrated military command, for example, is estimated to be associated with imposing a more costly commitment on signatories than an agreement containing only stipulations for the exchange of economic aid and a common defense.

In terms of the scope of the alliance agreement, we again find reassuring correlations between the measures we observe and the dimension we estimate. Figure 5
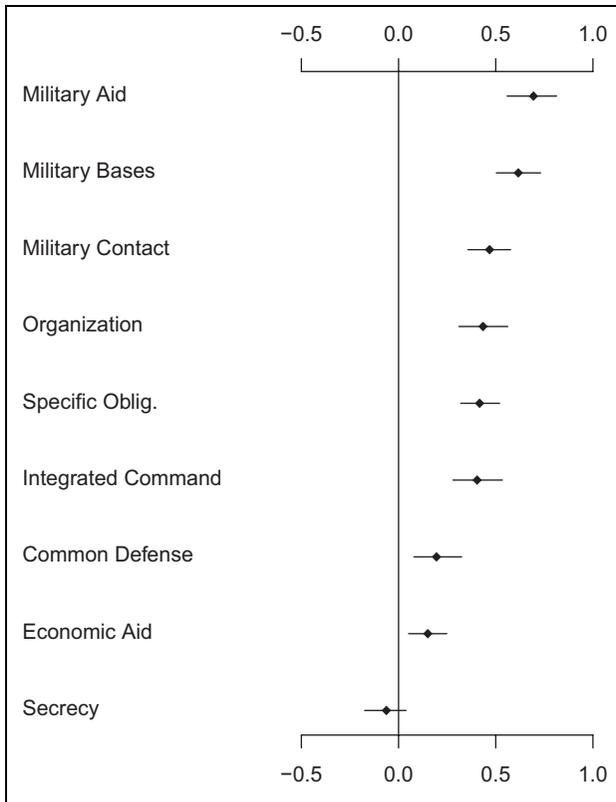
Figure 4. Factor Loadings: "Depth of Specified Obligations."
Note: Circles denote posterior mean and lines denote 95 percent HPD regions. HPD = highest posterior density.

shows the factor loadings for the inputs included in the model estimation for the scope of obligations. As expected, agreements that contain unconditional pledges of offensive and compellent military support are all estimated to be more expansive in scope than defensive agreements that condition military action on circumstances related to the war environment such as a particular target or location. It is also reassuring that our measurement model predicts that the alliance with the most limited scope would be a consultation or neutrality pact.

In sum, not only does our measurement model produce estimates of prominent alliances that comport well with historical and qualitative investigations, but the alliance characteristics that are thought to be theoretically related to each of the three dimensions are also revealed to be related as expected. These examinations suggest the validity of our approach and the resulting estimates.
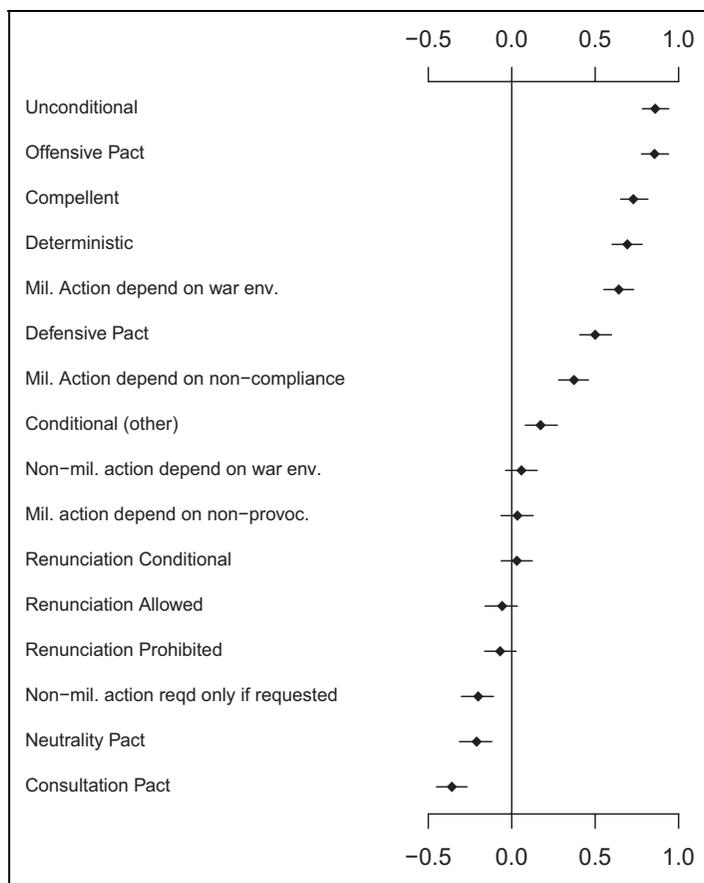
Figure 5. Factor Loadings: "Scope."

Note: Circles denote posterior mean and lines denote 95 percent HPD regions. HPD = highest posterior density.

Construct Validity

Understanding the conditions under which various types of alliances are formed and the decision of how broad and deep to design the terms of the agreement are important questions for better understanding the way in which countries construct the international arena through their choices of actions and our measures provide the necessary characterization to begin these important investigations. In addition, applying our measures in an applied analysis of alliance formation helps validate the construct validity of our measures by showing that features that are theoretically thought to be related to the formation of different alliance types are revealed to be so using our measures.

Scholars of alliances have theorized that different factors might be expected to give rise to different alliances on the three dimensions that we estimate. Existing work focuses on explaining variation in the observed measures—Benson (2012), for example, disaggregates deterrent types of alliances and studies why differences exist among these alliances—but we can use our measures to explore such question with more confidence. In particular, we specify regression models that test whether factors identified by scholars of alliance formation are, as theorized, associated with alliance formation as predicted. In particular, we estimate models of scope and depth with robust standard errors where all measures are taken in the year that the military alliance was formed.[10] In each case, the goal is to use factors in the regressions that have not already been used to construct the measures of scope and depth themselves. In contrast to the measurement model that seeks to include aspects inherent to an alliance that are thought to affect the underlying concept of interest, the goal in these regressions is to evaluate whether the resulting estimates reflect theoretically expected correlations in terms of aspects that are thought related to each concept but that do not themselves comprise it.

We first estimate a model of the *scope of the obligations* in an alliance agreement. Theories of alliance formation explain that concerns of entrapment create incentives for alliance members to design agreements to include conditions that limit military obligations and/or create flexible options for escape (Snyder 1984, 1997; Fearon 1997; Benson 2012). Benson's (2012) theory of moral hazard and alliance formulation predicts that alliance members' design the content of alliance agreements to be more flexible and to allow for probabilistic escape as alliance members' preferences and capabilities diverge (Benson 2012). However, empirical tests of this theoretical model are limited by the unavailability of a measure, such as our estimate of scope, that captures continuous variation of limits on the obligations that signatories include in the design of their agreements. Benson (2012) uses a coarse ordinal measure of particular conditions and qualifications included in alliance agreements.

Using estimates generated from our measurement model, we estimate a regression of the effects of divergent preferences and capabilities on the scope of an agreement. *Divergent preferences* are measured using the average UN affinity scores between pairs of alliance members, and *difference in capabilities* is measured by subtracting the CINC scores of the strongest and weakest pair of alliance members. We also include variables for the year in which the alliance was formed (*start year* and *start year*[2]), because alliance agreements become noticeably more limited in scope over time. Finally, we also include in the regression our measure of depth, since it may be associated with scope.

Table 1 presents the results of model 1. Using our measure of scope, it is clear that alliances agreements are designed to be more limited as signatories' preferences and capabilities diverge. The coefficients for divergent preferences and difference in capabilities are both negative and significant. These results confirm our expectations from theory and provide further reassurance that our measures correctly capture

Table 1. Regression of the Formation Conditions for the Scope and Depth of Alliance Agreements.

	(1) Scope post-1945	(2) Scope pre-1946	(3) Depth post-1945
Constant	1,417.91† (859.15)	361.70 (317.87)	−2,747.45* (1,259.62)
Divergent preferences (UN score)	−0.57** (0.14)		0.63* (0.30)
Difference in capabilities	−1.15* (0.46)	3.24** (1.07)	
Depth	0.09† (0.05)	0.46** (0.14)	
Democracy			0.004 (0.17)
Major power involved			0.57** (0.14)
Avg. log(capacity)			−9.61** (2.44)
Scope			0.30* (.14)
Start year	−1.42 (0.87)	−0.37 (0.34)	2.80* (1.27)
Start year^2	0.0004 (.0002)	0.0001 (.0001)	−0.0007* (0.0003)
N	267	189	267
R^2	.32	.21	.24

Standard errors in parentheses.
Note: $^\dagger p < .10$, $*p < .05$, $**p < .01$.

valid and meaningful concepts. Additionally, our measure of depth is also positively related to scope, though the effects are only significant at the $p \leq .10$ level.

It is important to note, however, that the finding on *divergent capabilities* only holds after 1945. As shown in model 2, prior to 1945, not only does the result not hold, it is actually positively associated with scope, suggesting that, at least for some early time period, alliances were designed to be broader when members' capabilities were more divergent.[11] It is possible that the advent of nuclear weapons dramatically exacerbated concerns of moral hazard in the post-WWII era by significantly widening the capabilities gap between nuclear and nonnuclear states and significantly increasing the risks associated with being involved in a conflict. Using our measure of scope, we have not only been able to evaluate an existing theory of moral hazard and alliance formation but also identified an intriguing puzzle about different effects of the divergence of alliance members' preferences on agreement design that may raise new questions about moral hazard, military technology, and alliance formation.

In addition to estimating a model of scope, we regress depth on several covariates that are predicted to be associated with the extent to which an agreement imposes peacetime and formation costs on alliance members. Many scholars argue that individual incentives for opportunism may lead to deeper agreements (Leeds and Anac 2005; Abbott and Snidal 2000; Lake 1999). Prospective allies who have divergent preferences may have greater incentives to behave opportunistically. Snyder (1997, 11) claims that divergent preferences often lead states to include

costly actions in their alliance commitments to "validate" agreements "when interests of allies are somewhat divergent, or when they are subject to change . . . " According to Snyder, when signatories' preferences diverge, "The partner will need to be reassured that one's underlying interests are not so at odds with the contract that the alliance is no more than a 'scrap of paper.'" Additionally, states with divergent preferences may need to come to an agreement on how their militaries will coordinate. Formalizing these terms in an alliance enhances the depth of the commitment. Consequently, we use UN affinity scores to measure the divergence of foreign policy preferences of alliance members in the year that the alliance was formed.

Other factors, which also might create worries of opportunism, may lead alliance members to design agreements with deep commitments. If major powers and nondemocracies pay relatively lower costs for violating agreements (Leeds 2003a), then at the alliance formation stage such states may need to provide additional assurances to signal commitment (Fearon 1997; Morrow 1994). If so, then when states with these attributes are party to an alliance, we might expect the terms of the agreements to impose deeper costs as a signal of commitment. Lake (2011) provides an additional explanation for why the presence of a major power might increase the depth of an alliance agreement. According to Lake, dominant states benefit from establishing rules that both discipline subordinates and credibly commit to limit their own power. Accordingly, we might expect that alliance agreements with major power allies will formalize treaty terms that establish security organizations and impose military coordination costs on members. Thus, in model 3, we include a variable for *major power*, which indicates whether at least one major power is party to an alliance agreement at the time of signing. We also include a variable for *democracy*, which is measured by creating an indicator for each alliance member that is a democracy according to Polity IV (Marshall, Jaggers, and Gurr 2002) and then taking the average across all alliance members at the time of formation.[12]

Another factor that may be associated with the depth of an agreement is the military strength of the signatories. A primary advantage of military alliances is the synergistic benefit that comes from realizing economics of scale through joint military coordination (Lake 1996, 1999; Conybeare 1992, 1994). That is, deep military coordination may expand alliance members' joint capabilities beyond states' additive raw capabilities. States with already relatively low capabilities may especially benefit from such coordination, and relatively stronger states may feel less of a need to pay peacetime military costs to gain coordination benefits. We explore this relationship by including a variable for the *average log(capacity)* of the alliance using the members' average CINC scores at the time of formation. Finally, we also include our measure of scope as a variable, because we might expect that alliances with broader scope of alliance obligations require, on average, costlier commitments both to coordinate more extensive military obligations and to signal the credibility of a broader range of commitments.[13]

As Table 1 reveals, all of these variables except for *democracy* are associated with depth as expected. The result for divergent preferences is especially interesting, given that the coefficient is positive when estimating depth. Recall that divergent preferences were negatively related to scope in model 1. The opposite relationship with these dimensions reassures us that the measures of scope and depth capture distinct concepts. In the regressions of both scope and depth that we have estimated in this section, not only have the results added to our confidence that the measurement model has recovered conceptually valid aspects of alliance variation from our measurement model but also yielded substantively relevant findings about theoretically meaningful predictions of the formation of formal military alliances.

Conclusion, Caveats, and Implications

Scholars of alliances have been blessed with a tremendous amount of data due to the generous and impressive efforts of those who have collected data on both the formal terms of the various alliances and the characteristics of the signatories involved. The amount of available data prompts the question, How can we best measure the strength of international alliances, given the wealth of available data and our theoretical conceptions of alliance strength while also accounting for the inevitable ambiguity that must necessarily accompany such a determination?

We show how a Bayesian latent variable trait model—a model that has been applied to many other important concepts in political science—can be used to integrate observable measures with theoretical arguments about the determinants of alliance strength to provide estimates of how the various observable factors relate to the theoretically implied dimensions, how the strength of alliances in the two dimensions relate to one another, and how certain we are about all of the estimated parameters. Applying the model to all alliances in the international system between 1816 and 2000 provides estimates of alliance strength in terms of the strength of signatories and the strength of the formal terms of the alliance agreement. We include alliances that are multilateral and lack a specific target. The estimates have strong face validity—familiar alliances are located as we would expect, and exploring the post-WWII alliances of East Asia provides reassuringly reasonable estimates.

Not only do our estimates provide scholars with measures that can be used to explore questions such as the determinants of alliance formation, they are also relevant for other pressing topics in the study of military alliances such as the persistence of alliances, their relationship with conflict, and their reliability. Additionally, the estimates also reveal important insights into the nature of alliances. For example, although there is a relationship between the depth and scope of military agreements, the correlation is modest (.112), and there is considerable variation of potential interest in the relationship. Moreover, when comparing the estimates of temporally and geographically proximate alliances involved in the World Wars and East Asia, the estimates suggest that many of the

alliances that are formed are nearly balanced. Perhaps reflecting the theoretical claims of balance of power theorists (Morgenthau 1948; Waltz 1979), alliances formed in response to one another are very similar in military capacity and often bear similarities in the scope of the terms of the obligations. Third, the ability to assess the certainty with which we are able to estimate the location of alliances provides the ability to account for the precision of our estimates when comparing alliances or using the estimates in secondary analyses. Finally, although we think our estimates are theoretically sound and possess strong face validity, the measurement model we employ is sufficiently general that scholars can use the model to generate their own estimates using different assumptions or different measures. Such flexibility in the applicability of the method may, for example, be of value if scholars wish to evaluate questions that call for the examination of only active military alliances or just those with specific targets.

We think our measure and the method provide an important advance for scholars interested in alliances and the international system, but some caveats are worth noting. First, while our statistical model provides a principled way of extracting information from multiple measures related to theoretically relevant dimensions with a minimal number of assumptions, the resulting estimates are still dependent on the relationship between the observable measures to extract the latent dimensions. For example, the fact that CINC scores are relative to the global capacity in each year means that comparisons across long periods of time may be difficult and the estimates are likely most appropriate for temporally comparisons or when temporally related differences are accounted for (similar concerns can emerge whenever these scores are used in any regression that explores variation over time).

Second, the ability to use a Bayesian latent variable model to extract the latent dimensions structuring the strength of an alliance does not ameliorate potential concerns about the endogeneity of alliances. While we have been careful not to include input variables in the measure of our three dimensions of scope, depth, and potential military capacity that scholars may seek to correlate in regressions of these same measures, nothing in our analysis discounts the fact that the formation of an alliance is presumably a strategic act based on the assessment of expected consequences, and those interested in using the estimates should be careful of the potential pitfalls.

While no measure is perfect, our model is able to use theoretical insights to identify how features are related to the scope, depth, and potential military capacity of an alliance in theoretically relevant dimensions. It also generates the measure for these dimensions from observable characteristics of the signatories involved and the formal terms of the agreement while also quantifying how certain we are about the resulting estimates. Moreover, it also provides a principled way to integrate alternative measures and assumptions into the measurement task that produces estimates that allow us to focus on better understanding the determinants and consequences of alliances rather than continually having to grapple with the question of how best to measure the variation across the alliances themselves.

Acknowledgment

Declaration of Conflicting Interests

Funding

Notes

1. See Powell (1996) and Levy (1989) for a review of the debate over the relationship between alliances, distribution of power, and stability.
2. The fifth section in the Online Appendix lists all of the factor inputs for each dimension along with a description of the variable and its coding rule.
3. Scholars often refer to the depth of agreements and international institutions. For a review of the literature regarding the formalization of cooperative security institutions and alliances in particular, see Leeds and Anac (2005). Similar to our approach with military alliance agreements, Dur et al. (2014) performs a factor analysis on 50 substantive provisions to estimate the depth of preferential trade agreements.
4. Scholars have long recognized the need for adjusting aggregate military capacity for distance. Most existing research degrades strength linearly (see e.g., Bueno de Mesquita and Lalman [1986] and Smith [1996]), but research has not established the actual mathematical relationship between distance and capabilities, and we also do not know if the rate of degradation is sensitive to the technological sophistication and geography of a country. Consequently, we make no assumptions about how distance degrades military capacity.
5. That said, scholars could certainly use our framework to estimate a more limited measure for alliances with identified targets.
6. We can treat $y = \alpha + \beta_1 x_1 + \beta_2 x_2$ as accounting for the regression of y on the unobserved alliance strength x^*, given the true specification $y = \alpha_0 + \beta_0 x^*$ if we can assume that x^* is a linear function of x_1 and x_2. If, for example, $x* = \gamma_1 x_1 + \gamma_2 x_2$ and $y = \alpha_0 + \beta_0 x*$, the regression of $y = \alpha + \beta_1 x_1 + \beta_2 x_2$ is equivalent to the regression of $y = \alpha_0 + \beta_0 (\gamma_1 x_1 + \gamma_2 x_2)$ because $\alpha = \alpha_0$, $\beta_1 = \beta_0 \times \gamma_1$, and $\beta_2 = \beta_0 \times \gamma_2$ even though we do not observe x^*! Note, however, that this decomposition relies on the extremely strong—and implausible—assumption that $x* = \gamma_1 x_1 + \gamma_2 x_2$. This requires that not only must the latent trait be a function of observables that are correctly specified in the regression specification, but also that the relationship between x^* and the observables is without error. If there is error in this relationship—say $x^* = \gamma_1 x_1 + \gamma_2 x_2 + \varepsilon$ —then we are in a classic error-in-variables situation and the estimated regression coefficients are inconsistent (see, e.g., the nice review by Hausman 2001).

7. Specifically, the prior distribution of β_k conditional on σ_k^2 is normally distributed and the prior distribution for σ_k^2 is an inverse Gamma distribution (Jackman 2009a).

8. In the second section of the Online Appendix, we provide additional analysis of other notable alliances that show further evidence of balancing based on our estimates.

9. Our measure of potential military capacity is a straightforward estimate of the combined capabilities of the signatory states. It correlates highly with the summed CINC scores of alliance members.

10. We also specify a regression of potential military capacity, which we include in section 4 of the Online Appendix.

11. Note models 1 and 3 include all alliances formed after data for United Nations affinity scores is available in 1946. Model 2 does not include data for divergent preferences, because the time period in this model is prior to the formation of the United Nations.

12. In other specifications, we also use other measures including average polity, a dummy variable for at least one democracy, and the standard error of the mean polity score. None of these measures has a significant effect on depth.

13. The effects of the other variables in the model are robust to the exclusion of scope.

Supplemental Material

The online appendix is available at http://jcr.sagepub.com/supplemental.

References

Abbott, Kenneth W., and Duncan Snidal. 2000. "Hard and Soft Law in International Governance." *International Organization* 54 (3): 421-56.

Achen, Christopher H. 2005. "Let's Put Garbage-can Regressions and Garbage-can Probits Where They belong." *Conflict Management and Peace Science* 22 (4): 327-39.

Bennett, D. Scott, and Allan C. Stam. 2000a. "EUGene: A Conceptual Manual." *International Interactions* 26 (2): 179-204.

Bennett, D. Scott, and Allan C. Stam. 2000b. "A Universal Test of an Expected Utility Theory of War." *International Studies Quarterly* 44 (3): 451-80.

Benson, Brett V. 2011. "Unpacking Alliances: Deterrent and Compellent Alliances and Their Relationship with Conflict, 1816–2000." *The Journal of Politics* 73 (4): 1111-27.

Benson, Brett V. 2012. *Constructing International Security: Alliances, Deterrence, and Moral Hazard*. New York: Cambridge University Press.

Benson, Brett V., Patrick Bentley, and Jim Ray. 2013. "Ally Provocateur: Why Allies Do Not Always Behave." *Journal of Peace Research* 50 (1): 47-58.

Benson, Brett V., Adam Meirowitz, and Kris Ramsay. 2013. "Inducing Deterrence through Moral Hazard in Alliance Contracts." *Journal of Conflict Resolution* 57 (6): 307-35.

Boulding, Kenneth Ewart. 1962. *Conflict and Defense: A General Theory*. Newport, UK: Harper.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14 (1): 63-82.

Braumoeller, Bear F. 2004. "Hypothesis Testing and Multiplicative Interaction Terms." *International Organization* 58 (4): 807-20.

Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, Eric Tanenbaum, Richard C. Fording, Derek J. Hearl, Hee Min Kim, Michael McDonald, and Silvia Mendez. 2001. *Mapping Policy Preferences. Estimates for Parties, Electors, and Governments 1945-1998*. Oxford, UK: Oxford University Press.

Bueno de Mesquita, Bruce. 1983. *The War Trap*. New Haven, CT: Yale University Press.

Bueno de Mesquita, Bruce, and David Lalman. 1986. "Reason and War." *The American Political Science Review* 80 (4): 1113-29.

Clinton, Joshua D., Simon Jackman, and Doug Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98 (2): 355-70.

Clinton, Joshua D., and David E. Lewis. 2008. "Expert Opinion, Agency Characteristics, and Agency Preferences." *Political Analysis* 16 (1): 3-20.

Conybeare, John A. C. 1992. "A Portfolio Diversification Model of Alliances: The Triple Alliance and Triple Entente, 1879-1914." *The Journal of Conflict Resolution* 36 (1): 53-85.

Conybeare, John A. C. 1994. "The Portfolio Benefits of Free Riding in Military Alliances." *International Studies Quarterly* 38 (3): 405-19.

Correlates of War. 2011. "State System Membership List, v2011." *Correlates of War Project*. Accessed January 15, 2014. http://correlatesofwar.org.

Dur, Andreas, Leonardo Baccini, Manfred Elsig, and Karolina Milewicz. 2014. "The Design of International Trade Agreements: Introducing a New Dataset." *The Review of International Organizations*. 9 (3): 353-75.

Fearon, James D. 1997. "Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs." *The Journal of Conflict Resolution* 41 (1): 68-90.

Gartzke, Erik, and Kristian Skrede Gleditsch. 2004. "Why Democracies May Actually Be Less Reliable Allies." *American Journal of Political Science* 48 (4): 775-95.

Gibler, Douglas M., and Meredith Reid Sarkees. 2004. "Measuring Alliances: The Correlates of War Formal Interstate Alliance Dataset, 1816-2000." *Journal of Peace Research* 41 (2): 211-22.

Gibler, Douglas M., and John A. Vasquez. 1998. "Uncovering the Dangerous Alliances, 1495-1980." *International Studies Quarterly* 42 (4): 785-807.

Gibler, Douglas M., and Scott Wolford. 2006. "Alliances, Then Democracy: An Examination of the Relationship between Regime Type and Alliance Formation." *The Journal of Conflict Resolution* 50 (1): 129-53.

Gill, Jeff. 2002. *Bayesian Methods: A Social and Behavioral Sciences Approach*. Boca Raton, FL: Chapman & Hall/CRC.

Gray, Julia, and Jonathan Slapin. 2011. "How Effective Are Preferential Trade Agreements? Ask the Experts." *The Review of International Organizations* 7:1-25.

Hausman, Jerry. 2001. "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left." *Journal of Economic Perspectives* 15 (4): 57-67.

Hoyland, Bjorn, Karl Moene, and Fredrik Willumsen. 2012. "The Tyranny of International Index Rankings." *Journal of Development Economics* 97 (1): 1-14.

Jackman, Simon. 2009a. *Bayesian Analysis for the Social Science*. New York: Wiley & Sons.

Jackman, Simon. 2009b. "Measurement." In *The Oxford Handbook of Political Methodology*, chapter 6, edited by Henry E. Brady, Janet M. Box-Steffensmeier, and David Collier, 119-51. Oxford, UK: Oxford University Press.

Johnson, Jesse C., and Brett Ashley Leeds. 2011. "Defense Pacts: A Prescription for Peace." *Foreign Policy Analysis* 7 (1): 45-65.

Kim, Tongfi. 2011. "Why Alliances Entangle but Seldom Entrap States." *Security Studies* 20:350-77.

Lai, Brian, and Dan Reiter. 2000. "Democracy, Political Similarity, and International Alliances, 1816-1992." *The Journal of Conflict Resolution* 44 (2): 203-27.

Lake, David A. 1996. "Anarchy, Heirarchy, and the Variety of International Relations." *International Organization* 50 (1): 1-33.

Lake, David A. 1999. *Entangling Relations*. Princeton, NJ: Princeton University Press.

Lake, David A. 2011. *Hierarchy in International Relations*. Ithaca, NY: Cornell University Press.

Leeds, Brett Ashley. 2003a. "Alliance Reliability in Times of War: Explaining State Decisions to Violate Treaties." *International Organization* 57 (4): 801-27.

Leeds, Brett Ashley. 2003b. "Do Alliances Deter Aggression? The Influence of Military Alliances on the Initiation of Militarized Interstate Disputes." *American Journal of Political Science* 47 (3): 427-39.

Leeds, Brett Ashley, and Sezi Anac. 2005. "Alliance Institutionalization and Alliance Performance." *International Interactions* 31 (3): 183-202.

Leeds, Brett, Jeffrey Ritter, Sara Mitchell, and Andrew Long. 2002. "Alliance Treaty Obligations and Provisions, 1815-1944." *International Interactions* 28 (3): 237-60.

Levendusky, Matthew S., and Jeremy C. Pope. 2010. "Measuring Aggregate-level Ideological Heterogeneity." *Legislative Studies Quarterly* 35 (2): 259-82.

Levendusky, Matthew S., Jeremy C. Pope, and Simon D. Jackman. 2008. "Measuring District-level Partisanship with Implications for the Analysis of U.S. Elections." *American Journal of Political Science* 70 (3): 736-53.

Levy, Jack S. 1981. "Alliance Formation and War Behavior: An Analysis of the Great Powers, 1495-1975." *The Journal of Conflict Resolution* 25 (4): 581-613.

Levy, Jack. 1989. "The Causes of War: A Review of Theories and Evidence." In *Behavior, Society, and Nuclear War*, vol. 1, edited by Philip E. Tetlock, Jo L. Husbands, Robert Jervis, Paul C. Stern, and Charles Tilly, 209-333. New York: Oxford University Press.

Marshall, Monty G., Keith Jaggers, and Ted Robert Gurr. 2002. *Polity IV Project Political Regime Characteristics and Transitions, 1800-2002*. College Park: Center for International Development and Conflict Management, University of Maryland.

Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the US Supreme Court, 1953-1999." *Political Analysis* 10 (2): 134-53.

Martin, Andrew D., Kevin M. Quinn, and Jong Hee Park. 2011. "MCMCpack: Markov Chain Monte Carlo in R." *Journal of Statistical Software* 42 (9): 1-21.

Mattes, Michaela. 2012. "Democratic Reliability, Precommitment of Successor Governments, and the Choice of Alliance Commitment." *International Organization* 66 (1): 153-72.

Mattes, Michaela, and Greg Vonnahme. 2010. "Contracting for Peace: Do Nonaggression Pacts Reduce Conflict?" *Journal of Politics* 72 (4): 925-38.

Morgenthau, Hans J. 1948. *Politics among Nations: The Struggle for Power and Peace*. New York: Knopf.

Morrow, James D. 1991. "Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances." *American Journal of Political Science* 35 (4): 904-33.

Morrow, James D. 1994. "Alliances, Credibility, and Peacetime Costs." *Journal of Conflict Resolution* 38 (2): 270-97.

Organski, Abramo F. K. 1968. *World Politics*. New York: Knopf New York.

Pemstein, Daniel, Stephen A. Meserve, and James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18 (4): 426-49.

Plummer, Martyn, Nicky Best, Kate Cowles, and Karen Vines. 2006. "CODA: Convergence Diagnosis and Output Analysis for MCMC." *R News* 6 (1): 7-11.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-economic History of Roll Call Voting*. New York: Oxford University Press.

Powell, Robert. 1996. "Stability and the Distribution of Power." *World Politics* 48 (1): 239-67.

Quinn, Kevin M. 2004. "Bayesian Factor Analysis for Mixed Ordinal and Continuous Responses." *Political Analysis* 12 (4): 338-53.

Ray, James Lee. 2003. "Explaining Interstate Conflict and War: What Should Be Controlled for?" *Conflict Management and Peace Science* 20 (2): 1-31.

Rosas, Guillermo. 2009. "Dynamic Latent Trait Models: An Application to Latin American Banking Crises." *Electoral Studies* 28 (3): 375-87.

Rosenthal, Howard, and Erik Voeten. 2007. "Measuring Legal Systems." *Journal of Comparative Economics* 35 (4): 711-28.

Schelling, Thomas C. 1966. *Arms and Influence*. New Haven, CT: Yale University Press.

Schnakenberg, Keith E, and Christopher J. Fariss. 2014. "Dynamic Patterns of Human Rights Practices." *Political Science Research & Methods* 2 (1): 1-31.

Signorino, Curtis S., and Jeff M. Ritter. 1999. "Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions." *International Studies Quarterly* 43 (1): 115-44.

Simon, Michael W., and Erik Gartzke. 1996. "Political System Similarity and the Choice of Allies: Do Democracies Flock Together, or Do opposites Attract?" *The Journal of Conflict Resolution* 40 (4): 617-35.

Singer, J. David, Stuart Bremer, and John Stuckey. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820-1965." In *Peace, War, and Numbers*, edited by Bruce Russett, 19-48. Beverly Hills, CA: Sage.

Siverson, Randolph M., and Michael R. Tennefoss. 1984. "Power, Alliance, and the Escalation of International Conflict, 1815-1965." *American Political Science Review* 78 (4): 1057-69.

Smith, Alastair. 1995. "Alliance Formation and War." *International Studies Quarterly* 39 (4): 405-25.

Smith, Alastair. 1996. "To Intervene or Not to Intervene: A Biased Decision." *The Journal of Conflict Resolution* 40 (1): 16-40.

Snyder, Glenn H. 1984. "The Security Dilemma in Alliance Politics." *World Politics* 36 (4): 461-95.

Snyder, Glenn Herald. 1997. *Alliance Politics*. Ithaca, NY: Cornell University Press.

Starr, Harvey, and Benjamin A. Most. 1976. "The Substance and Study of Borders in International Relations Research." *International Studies Quarterly* 20 (45): 581-620.

Treier, Shawn, and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52 (1): 201-17.

Tsebelis, George. 2002. *Veto Players: How Political Institutions Work*. Princeton, NJ: Princeton University Press.

Voeten, Erik. 2000. "Clashes in the Assembly." *International Organization* 54:185-215.

Walt, Stephen M. 1987. *The Origins of Alliances*. Ithaca, NY: Cornell University Press.

Waltz, Kenneth Neal. 1979. *Theory of International Politics*. Jamesport, NY: Addison-Wesley Pub. Co.

Weidmann, N. B., D. Kuse, and K. S. Gleditsch. 2010. "The Geography of the International System: The Cshapes Dataset." *International Interactions* 36 (1): 86-106.

Yuen, A. 2009. "Target Concessions in the Shadow of Intervention." *Journal of Conflict Resolution* 53 (5): 745-73.

Zagare, F. C., and D. Marc Kilgour. 2003. "Alignment Patterns, Crisis Bargaining, and Extended Deterrence: A Game-theoretic Analysis." *International Studies Quarterly* 47 (4): 587-615.