

# **Appendix: Democracy, Electoral Competition, and Alliance Treaty Depth**

This appendix checks the findings in the manuscript in six ways. In the first section, I show results with several alternative model specifications, including controls for US membership, a high electoral democracy dummy, polity scores of the most capable alliance member, and raw data about the proportion of electoral democracies when the alliance formed and alliance treaty design. Then I estimate the association between allied democracy and alternative measures of treaty depth. After that, I show that findings of a negative correlation between military institutionalization and alliance performance are sensitive to how treaty depth is measured. In the fourth section, I assess whether accounting for non-random selection into alliances changes the relationship between democracy and treaty design. I also consider how uncertainty in the latent depth measure affects inferences about the association between democracy and treaty depth. Last, I add issue linkages to create a trivariate model of the sources of alliance treaty credibility.

## **1 Alternative Model Specifications**

Throughout this appendix, I consider and estimate a series of modifications to the models in the paper. In addition to the robustness checks with alternative measures of electoral democracy, I fit two models—one with a crucial control and an alternative expression of the lexical index of electoral democracy (LIED). First, perhaps the positive association between electoral democracy and treaty depth is driven by the United States, which formed multiple deep and conditional alliances

after 1945. To assess this possibility, I add a dummy indicator of US membership to the generalized joint regression model specification in the manuscript. Table 1 summarizes the resulting estimates. Even after adjusting for US membership, the same patterns hold- electoral democracy decreases the probability of unconditional support, but decreases treaty depth.

	Uncond. Mil. Support		Latent Depth	
	Estimate	Std. Error	Estimate	Std. Error
Executive Constraints	0.7827697	0.2497400	-0.2816211	0.19945208
Electoral Democracy	-0.1926309	0.0464984	0.1018848	0.03968870
Economic Issue Linkage	0.2317009	0.1538630	-0.0916942	0.13180424
FP Concessions	-0.1325221	0.0702642	0.0194565	0.06684578
Number of Members	-0.1372209	0.0165604	0.0292460	0.01389707
Wartime Alliances	-0.5381560	0.1826230	-0.0829389	0.17755572
Asymmetric Obligations	-0.0821881	0.1894113	0.2290210	0.17380222
Asymmetric Capability	1.4378521	0.2782333	0.5737740	0.20675204
Non-Major Only	2.0806578	0.2699037	0.1896873	0.21443238
FP Disagreement	0.4528560	0.2478126	0.2675286	0.20833092
US Member	-0.8133331	0.5911568	0.1073361	0.28787374
(Intercept)	-1.3690869	0.3141911	-1.4500026	0.27104533

Table 1: Results from a joint generalized regression model of treaty depth and unconditional military support. This model includes a dummy variable to control for US alliance membership. All smoothed terms report the effective degrees of freedom and the chi-squared term. The unconditional military support model is a binomial GLM with a probit link function. The treaty depth model is a beta regression. I model the error correlation between the two processes with a T copula.

As a further robustness check, I coarsen the lexical index of electoral democracy into a dummy indicator. I set this high electoral democracy indicator equal to one if the most capable alliance member scores a four or higher on the ordinal lexical index. Thus, this variable sets a threshold for electoral democracy to impact alliance treaty design. Table 2 shows the estimates from this model specification. I find the same results with a binary measure of electoral democracy- high electoral democracy increases treaty depth and decreases the probability of unconditional military support.

	Uncond. Mil. Support		Latent Depth	
	Estimate	Std. Error	Estimate	Std. Error
Executive Constraints	0.7810906	0.3356106	-0.5993968	0.27879422
Electoral Democracy	-0.9638878	0.3777303	0.8417301	0.26915378
Economic Issue Linkage	-0.0187766	0.1305101	0.0444618	0.12030579
FP Concessions	0.0671109	0.0780041	-0.0197488	0.06569849
Number of Members	-0.1241376	0.0235545	0.0231010	0.01433187
Wartime Alliances	-0.7444694	0.4019197	-0.0335950	0.16507610
Asymmetric Obligations	0.0188003	0.2143709	0.2946093	0.16280497
Asymmetric Capability	1.2704918	0.6958357	0.5169070	0.26184475
Non-Major Only	1.9809759	0.7008937	0.1857511	0.26012603
FP Disagreement	0.3142098	0.2221990	0.5270110	0.18483542
US Member	-0.3842154	0.5130693	0.0368238	0.28682336
(Intercept)	-1.3618016	0.7646655	-1.5596016	0.29809180

Table 2: Results from a joint generalized regression model of treaty depth and unconditional military support. This model replaces the ordinal lexical index of democracy with a dummy variable for states with a LIED score of four or higher. All smoothed terms report the effective degrees of freedom and the chi-squared term. The unconditional military support model is a binomial GLM with a probit link function. The treaty depth model is a beta regression. I model the error correlation between the two processes with a T copula.

## 1.1 Aggregate Polity Score

As an alternative expression of the impact of democracy, I estimated a model with the Polity indicator of regime type in place of the lexical index of democracy and executive constraints variables. Polity scores are a function of electoral competition, executive constraints, and open political competition, so this aggregate index is more focused on regime type than a specific characteristic of democracy. Thus, the results below should not be interpreted as evidence for or against the two hypotheses, because the hypotheses emphasize electoral politics. Rather, this analysis gives a sense of the possible overall impact of democratic regimes, when different components are aggregated and combined based on the Polity project's criteria.

Replacing the two democracy components with the Polity Index leads to slightly different inferences. Increasing the Polity score of the most capable alliance member leads to greater treaty depth. Thus, the next effect of elections, competition and executive constraints on treaty depth is

	Uncond. Mil. Support		Latent Depth	
	Estimate	Std. Error	Estimate	Std. Error
Most Capable Member Polity	-0.0157811	0.0136509	0.0219941	0.0094106
Economic Issue Linkage	-0.1235628	0.1322331	-0.1474833	0.1266920
FP Concessions	-0.1010907	0.1022996	0.0267593	0.0738329
Number of Members	-0.1079909	0.0235576	0.0253535	0.0141773
Wartime Alliances	-0.4556664	0.1911211	-0.2360622	0.1795052
Asymmetric Obligations	-0.0419032	0.2310118	0.2717387	0.1440629
Asymmetric Capability	1.5734008	0.2439077	0.4705337	0.1933493
Non-Major Only	2.2426598	0.2526901	0.1316914	0.1844885
FP Disagreement	0.2686585	0.2149443	0.4093991	0.1817255
s(Mean Threat)	6.9506656	100.3110358	1.2937629	18.9361851
s(Start Year)	8.2043321	116.6182702	7.9446766	37.4211195
(Intercept)	-1.7110218	0.3143552	-1.1940559	0.2319091

Table 3: Results from a joint generalized regression model of treaty depth and unconditional military support. Most capable member Polity is the key independent variable. All smoothed terms report the effective degrees of freedom and the chi-squared term. The unconditional military support model is a binomial GLM with a probit link function. The treaty depth model is a beta regression. I model the error correlation between the two processes with a T copula.

positive, and this likely reflects the impact of elections. The Polity coefficient for unconditional support is negative, but it cannot be distinguished from zero at conventional levels of statistical significance. This may reflect the competing effects of elections and executive constraints, as well as how Polity aggregates its component indicators.

## 1.2 Proportion of Democracies

This section offers a set of supporting estimates for inferences about the proportion of states with electoral competition reported in the manuscript. First, raw data on the proportion of states with electoral competition are consistent with my argument. Differences in the average proportion of states with open electoral competition across treaty depth and unconditional military support are consistent with the hypotheses, as Figure 1 shows. Deep and conditional alliances have the highest average share of members with open electoral competition.

This section also shows the parameter estimates which I translate into substantive effect calcu-

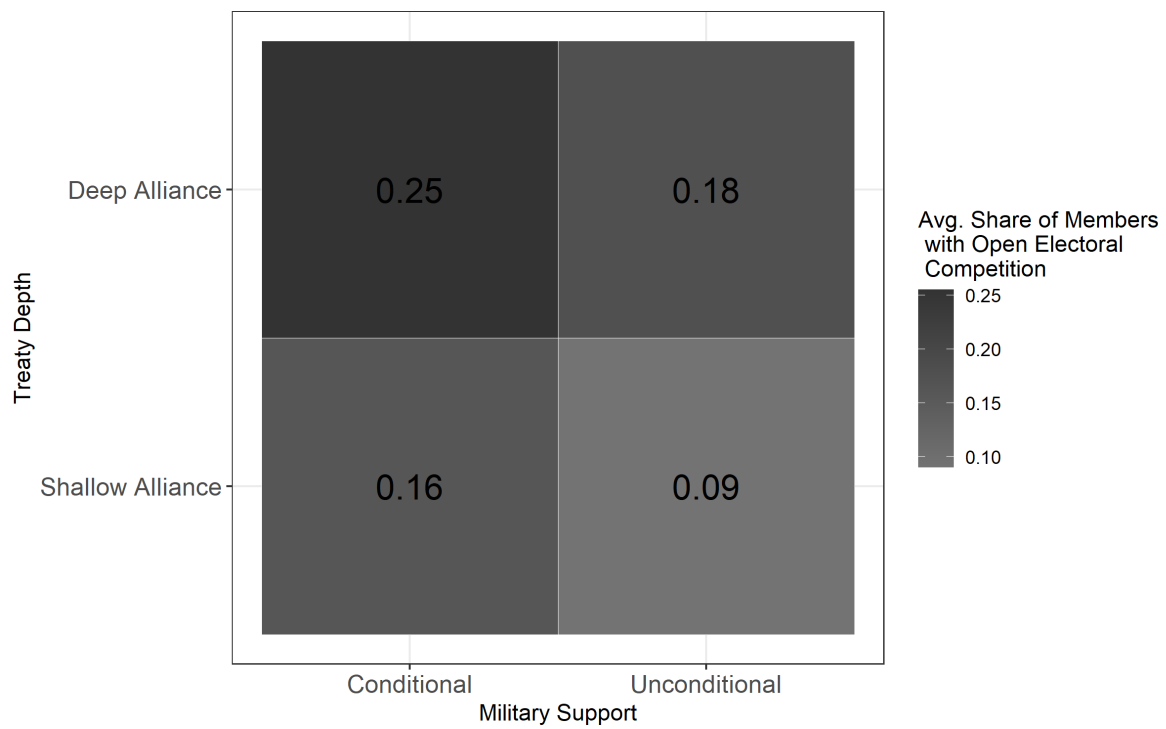


Figure 1: Average proportion of alliance members with open electoral competition in four groups of alliances. Alliances groups are based on treaty depth and unconditional military support.

lations in the paper in Table 4. As in the model that emphasizes competition and constraints in the most capable alliance member, the proportion of competition variable is negative for unconditional support and positive for treaty depth. Similarly, the share of states with executive constraints increases the probability of unconditional support, but has no clear relationship with treaty depth.

	Uncond. Mil. Support		Latent Depth	
	Estimate	Std. Error	Estimate	Std. Error
Proportion Executive Constraints	0.5939340	0.3092504	-0.0272315	0.2523820
Proportion Competitive Elections	-0.8218134	0.3858689	0.6506311	0.2936681
Economic Issue Linkage	-0.0571900	0.1459478	-0.1264786	0.1272376
FP Concessions	-0.0267587	0.1031483	0.0441473	0.0651197
Number of Members	-0.0986130	0.0236908	0.0252992	0.0132093
Wartime Alliances	-0.6846609	0.1995169	-0.0278939	0.1843955
Asymmetric Obligations	-0.2628870	0.2524742	0.1500064	0.1648030
Asymmetric Capability	1.7102946	1.0419926	0.5080621	0.2004186
Non-Major Only	2.3963606	1.0535705	0.1917110	0.2313839
FP Disagreement	0.4571224	0.2355976	0.3032570	0.1796349
s(Mean Threat)	7.8969292	101.7615169	1.0000056	21.9366853
s(Start Year)	6.8750132	93.5147005	8.4786452	49.1388890
(Intercept)	-1.9684451	1.0498710	-1.3479056	0.2660591

Table 4: Results from a joint generalized regression model of treaty depth and unconditional military support. All smoothed terms report the effective degrees of freedom and the chi-squared term. The unconditional military support model is a binomial GLM with a probit link function. The treaty depth model is a beta regression. I model the error correlation between the two processes with a normal copula.

## 2 Other Measures of Treaty Depth

To address a highly skewed distribution and facilitate inferences across a wide range of estimation strategies, results in the paper are based on rescaled mean latent treaty depth, which ranges between 0 and 1. I modeled the rescaled outcome with a beta distribution. I make similar inferences about treaty depth and the electoral democracy of the most capable alliance member when I estimate regression models of treaty depth without this transformation, however.

Latent mean treaty depth is extremely skewed. This in turn leads to assumption violations

and poor model fit with ordinary least squares estimators. To address this problem, I fit skew-t and skew-cauchy models of treaty depth. Both these estimators model the skewed errors, and the skew-cauchy model allows for larger outliers. The skewed regressions accommodate non-normal residuals and influential observations that might bias inferences with OLS. I summarize the estimates from the skew models in Figure 2.

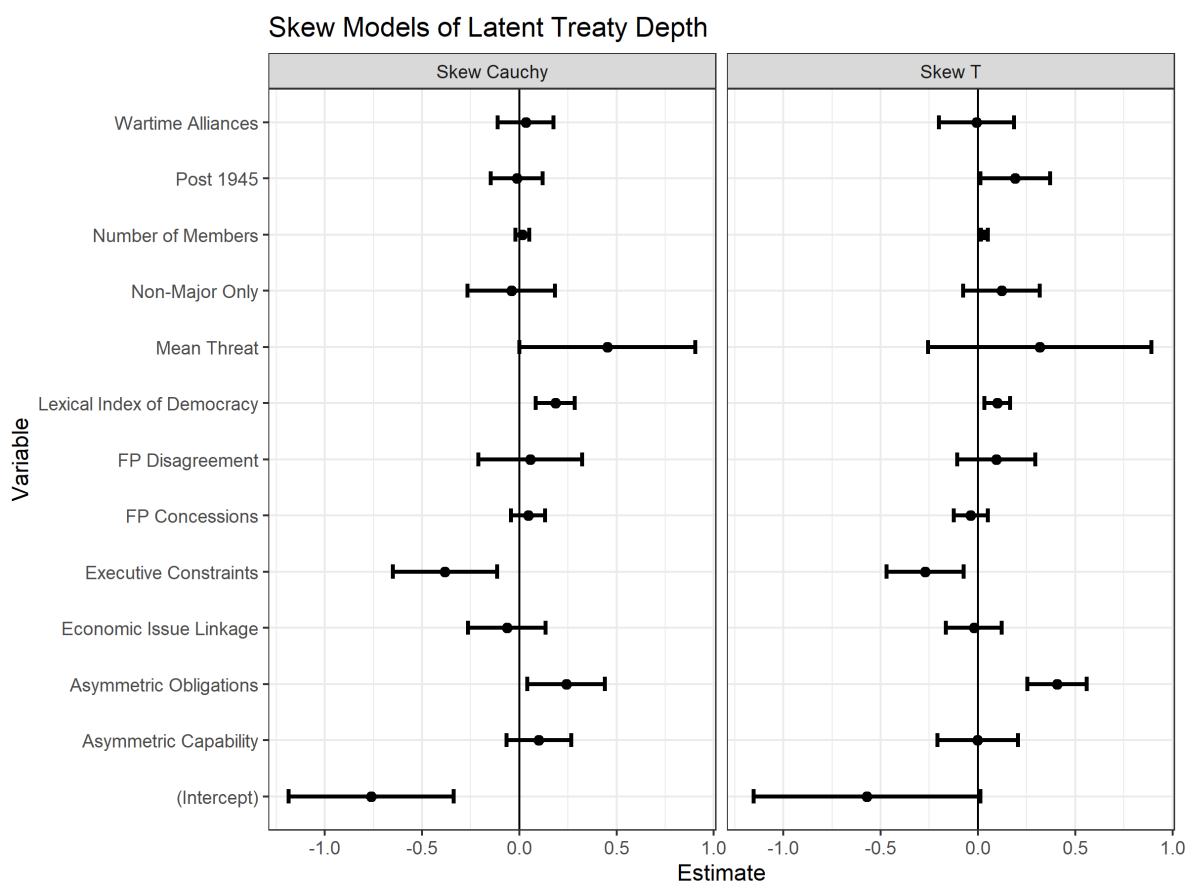


Figure 2: Coefficient plot of estimates from skew-t and skew-cauchy models of alliance treaty depth in offensive and defensive ATOP alliances from 1816 to 2012. Both estimators find a positive association between electoral democracy and treaty depth.

Without rescaling treaty depth, I still find a positive relationship between electoral democracy and alliance treaty depth. This inference also does not depend on how extreme the assumed skewness in the error terms is— both the skew-t and skew-Cauchy estimators give similar inferences.

Besides estimating the association between democracy and depth without rescaling the out-

come, I examine the relationship between alliance leader democracy and two measures from earlier scholarship that capture parts of treaty depth. First, Leeds and Anac (2005) develop an ordinal measure of military institutionalization. Second, Benson and Clinton (2016) use a latent variable model to measure treaty depth, but they conceptualize depth as the overall costliness of the alliance obligations. Although these two measures have limitations that make my measure a better fit for this paper, I find that the electoral democracy of the alliance leader is positively correlated with both variables.

Greater alliance leader electoral democracy increases the Leeds and Anac measure of military institutionalization in separate and joint models. The ordinal military institutionalization measure is a useful first step towards measuring defense cooperation, but it understates the amount of variation in treaty depth. Leeds and Anac (2005) argue that commitments of an integrated military command, common defense policy, or any basing rights generate high military institutionalization. Official contact between military officials, formal organizations, providing training or technology, subordination of forces, or specific contributions reflect moderate institutionalization. If at least one of the relevant factors is present, Leeds and Anac assign the alliance the highest corresponding level of institutionalization. To give an example, an alliance with basing rights and subordination of forces has high institutionalization and is just as institutionalized as an alliance with only basing rights. This approach assumes that alliances with multiple sources of depth as just as deep or institutionalized as alliances with one source. As a result, it understates variation in alliance treaty depth.

As Figure 3 shows, the ordinal variable is positively correlated with my latent measure, but the latent depth measure varies widely within each level. The deepest alliances on the latent measure also have the highest military institutionalization score. There are substantial differences within each category and overlap in the latent scores across the categories, however. For example, some alliances that Leeds and Anac (2005) assign a moderate institutionalization score have more depth than alliances with high institutionalization scores because these alliance treaties contain multiple



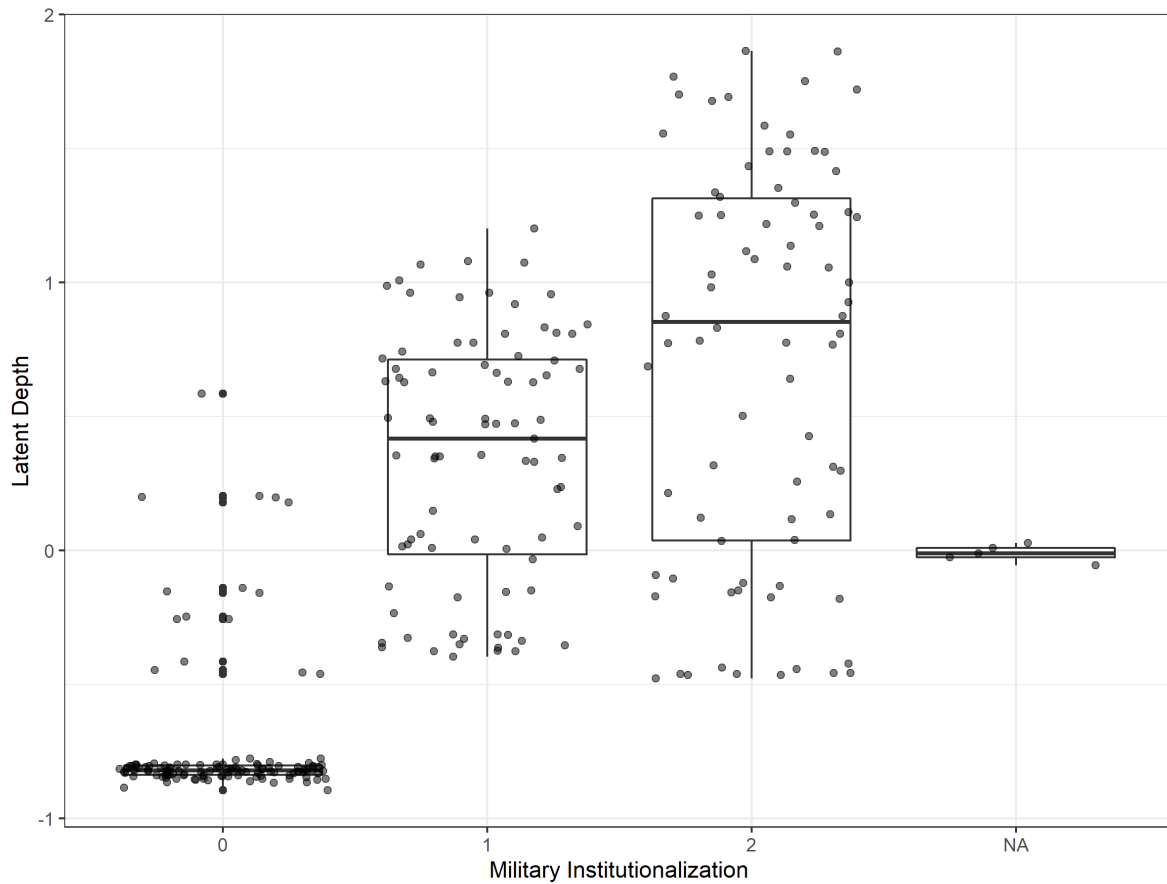


Figure 3: Scatter plot of latent treaty depth across the values of military institutionalization from Leeds and Anac (2005). The box plots summarize the distribution of latent treaty depth within each category of military institutionalization. Points are jittered within each level of the institutionalization score.

sources of depth.

Benson and Clinton (2016)'s measure addresses a different concept, covers fewer alliances and their measurement model relies on a set of potentially problematic distributional assumptions, so it produces more outliers that may impact inferences. Benson and Clinton's emphasis on the general costliness of the alliance means they include measures of issue linkages and secrecy, which are distinct from defense cooperation. Their measure is also based on version 3 of the ATOP data, so it has more limited temporal coverage. Last, their latent variable model makes a series of distributional assumptions that affect the estimated distribution of the latent depth measure. My latent variable model uses a semiparametric approach that accounts for dependencies between the factors and latent variables, so the distribution of the latent variable may be more accurate (Murray et al., 2013). For these three reasons, my latent measure better captures the concept of defense coordination and cooperation.<sup>1</sup> Despite these differences, I find a positive association between the lexical index of democracy in the most capable alliance member and Benson and Clinton's measure of depth.

I now discuss the results of analyses with both alternative dependent variables. I start with univariate models then estimate a bivariate model with each depth measure and unconditional military support. First, I fit an ordinal model of military institutionalization. Then, after rescaling Benson and Clinton's depth measure to range between 0 and 1, I fit a beta regression model. Figure 4 plots the substantive effect of the most capable alliance member's lexical index of electoral democracy score on both outcomes. Electoral democracy in the alliance leader increase the probability of high institutionalization and decrease the probability of no institutionalization. Alliance leader democracy is also positively associated with Benson and Clinton's measure of depth.

Figure 5 plots predictions from the two alternative depth measures based on bivariate models. For the military institutionalization measure, I could not fit an ordinal model in GJRM. Therefore, I

---

<sup>1</sup>If scholars want to measure the overall costliness of alliance obligations, they should consider Benson and Clinton's measure.

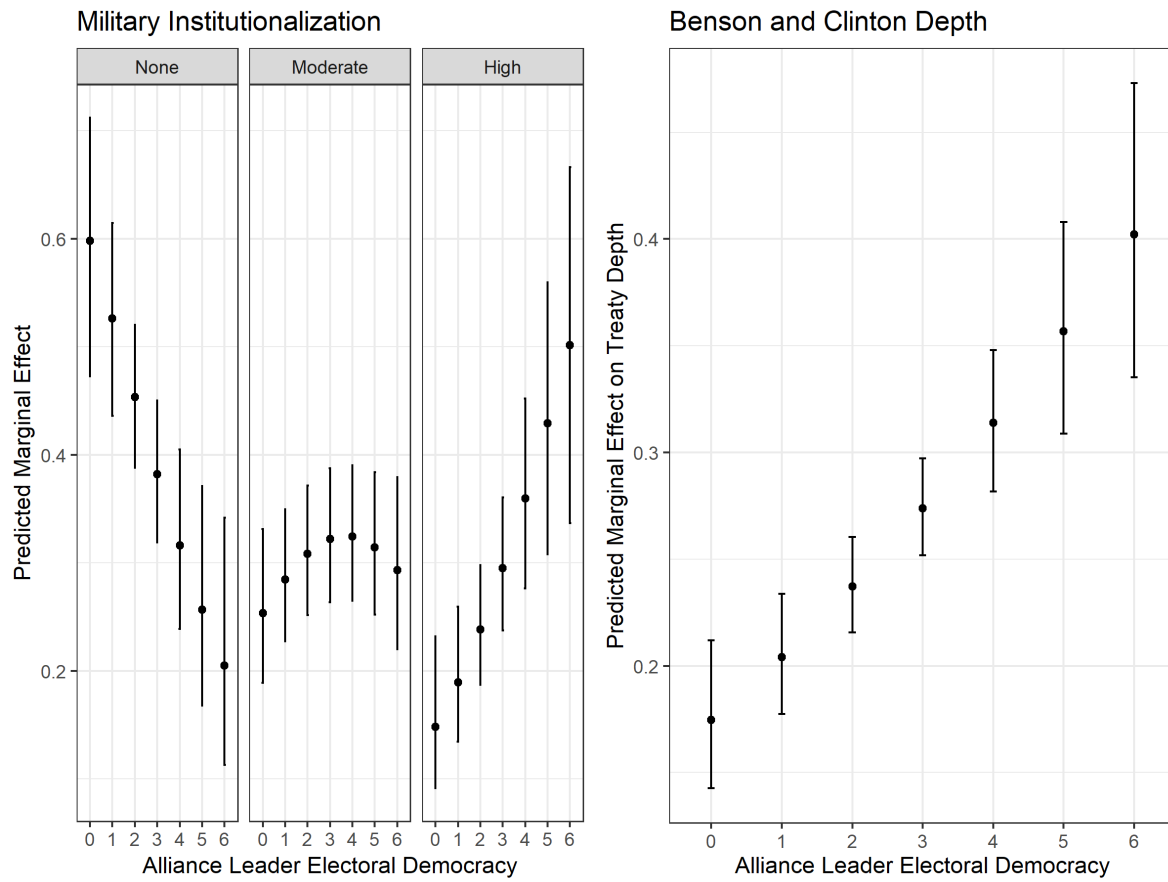


Figure 4: Predicted changes in treaty depth by the electoral democracy of the most capable alliance member. The line marks predicted values, and the shaded areas encapsulate the 95% confidence interval.

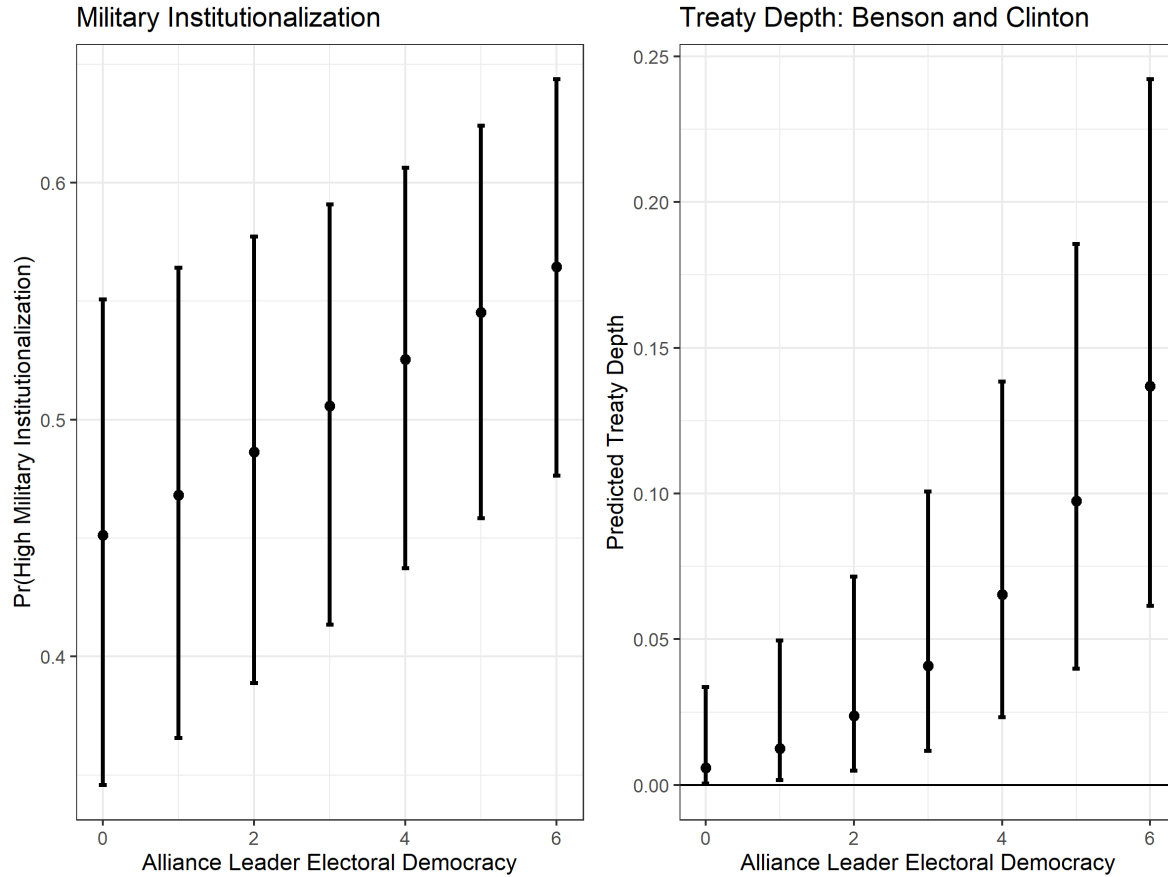


Figure 5: Predicted probability of military institutionalization and predicted treaty depth as a function of alliance leader electoral democracy. The line marks predicted values, and the error bars summarize the approximate 95% confidence interval. Predictions based on estimates from two joint generalized regression models, one for high military institutionalization and unconditional support, and another with Benson and Clinton's depth measure and unconditional military support. These estimates hold all other factors equal.

created a dummy variable and set it equal to one if the alliance had high institutionalization. I then used a probit estimator to model the high military institutionalization dummy. As the left panel of Figure 5 shows, the probability of high military institutionalization increases with alliance leader electoral democracy, although the predicted probabilities are fairly uncertain. This may reflect limited variation in the ordinal measure, which I further reduced by translating it into a dummy variable.

Findings about the association between treaty depth and Benson and Clinton's also match the

treaty depth hypothesis. Greater electoral democracy increases the overall costliness of the alliance obligations, probably due to the defense cooperation measures in Benson and Clinton's latent variable. In summary, I find similar results with four alternative measures of treaty depth, including two measures from previous scholarship.

### **3 Are Deep Alliances Less Reliable?**

Leeds and Anac (2005) find that institutionalized alliances are less reliable, which contradicts my claim that deep alliances increase treaty reliability. Besides potential bias from non-random selection into challenges of alliances (Smith, 1995), this empirical finding depends in part on their ordinal institutionalization measure. Replacing the ordinal measure with my continuous latent depth variable leads to different inferences about alliance treaty design and performance. Therefore, the surprising finding in this paper that better institutionalized alliances have worse performance may also reflect measurement decisions.

I focus my replication on Leeds and Anac's model of whether alliance members honor treaties with active military support. The replication focuses on this model because my argument examines defensive and offensive alliances. Their dataset includes 103 alliance performance opportunities, where states could honor or violate promises of military support. Using a logit model that adjusts for alliance formality, capability changes, changes in the policy process and whether the ally was the original target, Leeds and Anac find a negative relationship between institutionalization and performance. I replicate this estimate in the first column of Table 5.

I then replace Leeds and Anac's institutionalization measure with my latent treaty depth measure, which is in the second column of Table 5. The coefficient on depth, in contrast to military institutionalization, is positive, though neither estimate is statistically significant at conventional levels. Thus, the possible direction of this association shifts, depending on how alliance depth is measured.

Table 5:

	<i>Dependent variable:</i>		
	Leeds and Anac		Berkemeier and Fuhrmann
	(1)	(2)	(3)
Military Institutionalization	−0.543 (−1.306, 0.221)		
Latent Depth		0.258 (−0.484, 1.000)	0.347 (−0.372, 1.066)
Alliance Formality	−1.161 (−2.082, −0.240)	−1.538 (−2.483, −0.594)	
Capability Change	−1.841 (−3.135, −0.547)	−1.946 (−3.277, −0.614)	
Process Change	−1.802 (−3.336, −0.269)	−1.459 (−2.886, −0.032)	
Original Target	−0.723 (−1.849, 0.403)	−0.836 (−2.001, 0.329)	
Asymmetric Capability			−2.536 (−4.904, −0.168)
Non-Major Only			−1.740 (−3.954, 0.473)
Post 1945			−2.315 (−3.766, −0.864)
Share Electoral Democracy			−0.217 (−0.505, 0.070)
Share Executive Constraints			−0.450 (−1.587, 0.687)
Number of Members			1.054 (−0.544, 2.653)
Economic Issue Linkage			−0.093 (−0.786, 0.600)
Unconditional Support			0.263 (−2.343, 2.868)
Foreign Policy Concessions			1.799 (−0.391, 3.989)
Constant	3.430 (1.972, 4.888)	3.201 (1.799, 4.602)	2.524 (0.182, 4.866)
Observations	93	93	108
Log Likelihood	−39.131	−39.906	−48.298

*Note:*

95% Confidence Intervals in Parentheses.

To further assess this finding, I utilized an updated alliance performance dataset from Berke-meier and Fuhrmann (2018). In this analysis, I specified a logit model that uses latent depth, unconditional support, dummy indicators of asymmetric capability and symmetric non-major power pacts, a post 1945 dummy, the two indicators of democratic membership, the number of members, economic issue linkages, and foreign policy concessions to predict whether an alliance was honored in war. All of these factors are potential sources of reliability, per different arguments in the literature. In these models, I find a similar result— a positive relationship between depth and honoring military support that cannot be distinguished from zero.

In summary, I do not find that highly institutionalized alliances are less reliable. The change in the direction of the coefficient is suggestive, given the limited number of alliance performance opportunities in these datasets. Again, given non-random selection into alliance crises and challenges, placing excessive weight on the above results is unwise. They do suggest, however, that deep alliances are not necessarily less reliable when invoked.

## **4 Adjusting for Alliance Formation**

Observed alliances are the result of negotiations between potential members. Therefore the set of observed alliances is not a random sample. Rather, alliance members overcame the many obstacles to alliance formation (Poast, 2019). In this section, I account for non-random alliance formation using a hurdle model of observed alliances after combining a stratified random sample of non-allied groups of states with observed alliances. This research design produces partially similar inferences: electoral democracy in the most capable alliance member increases treaty depth but has an unclear relationship with conditions on military support.

My research design for this robustness check follows established procedures for dealing with selection into alliances and treaty design. I followed the suggestions of Poast (2010) for using k-adic data, because some alliances have more than two members. First I constructed a random

sample of groups of states that could have formed an alliance in each year, but did not (Fordham and Poast, 2014). Then I took a stratified sample of the non-allied k-ads to include five times as many non-allied observations as alliance observations for each observed value of alliance size. For example, there are 215 bilateral alliances, so I sampled 1075 non-allied bilateral groups. There is only one alliance with 34 members, so I added five non-allied k-ads with 34 members. Finally, I summarized key characteristics of these non-allied k-ads, including average Polity score, the electoral democracy of the most capable member, the number of members, mean threat, asymmetric capability, and wartime, and merged the non-allied data with the observed alliance data.

To model alliance treaty alliance design, I emulated the research design of Chiba, Johnson and Leeds (2015), who estimated a hurdle model to assess whether democracies were more likely to offer conditional obligations. Hurdle models have two stages or parts, which can account for non-random selection into alliances. A first stage model predicts which observations clear the hurdle to a second stage with non-zero outcome values. Observed alliances are the set of observations that cleared the hurdle and have an alliance treaty. Unlike a sample selection model, the second stage in a hurdle model is logically undefined, which is how alliance treaty design works.<sup>2</sup> Unless states overcome the barriers to alliance formation, no agreed treaty content exists.

Unlike Chiba, Johnson and Leeds (2015), however, I examine two aspects of alliance treaty design. Therefore, I estimate a bivariate hurdle model- one for the depth and the other for unconditional military support. To have zero values for the outcome at the hurdle stage, I adjusted the outcomes in these models. First, I shifted the scale of latent treaty depth by adding one, which shifted the distribution onto uniformly positive values. I then applied a gamma hurdle model to this rescaled depth outcome. For unconditional military support, I replaced the dummy variable in the manuscript with an ordinal indicator. This ordinal variable captures the relative strength of the military support promise in the alliance. ATOP alliances can be conditional on four things: adversaries, locations, a particular conflict, the number of opponents, a specific demand or non-

---

<sup>2</sup>Hurdle models also do not require an exclusion restriction for identification.



provocation. The maximum number of observed conditions in an alliance is four, which is the most conditional and limited alliance in the data. I therefore invert the number of conditions and assign unconditional alliances a value of four. Essentially, the measure then captures the number of circumstances where an alliance could have applied conditions but did not. To avoid including zeros in observed alliances, because 0 is the value for groups with no alliance, I add one to the count of conditions. Adding one also reflects how promising military support adds a base of commitment strength to any alliance (Morrow, 2000). Thus, the final measure of military support commitment strength ranges between 1 and 5 for observed alliances, but is zero for non-alliances.

In both models, I use average democracy, mean threat, asymmetric capability, wartime, and year varying intercepts to predict whether the group of states clears the alliance formation hurdle. In the second stage, I use electoral competition and executive constraints in the most capable alliance member, the number of members, mean threat, dummy indicators of asymmetric capability or non-major power only membership, wartime and year varying intercepts to model treaty design. I then estimate the two hurdle models simultaneously using the BRMS package for R, which employs STAN for fully Bayesian inference (Bürkner, 2017). Unlike the GJRM estimators in the manuscripts, I cannot directly estimate residual correlations. Instead, I connect the two hurdle models with a correlated set of year varying intercepts.<sup>3</sup>

After accounting for non-random alliance formation with the joint hurdle model, I find partially similar results to the manuscript. Figure 6 plots the results of the joint hurdle model. This figure shows the predicted marginal effect of open elections in the alliance leader on commitment strength and shifted treaty depth, along with 95% credible intervals. Electoral competition in the most capable alliance member increases treaty depth and has no clear relationship with the strength of conditions on military support.

---

<sup>3</sup>The complexity of this joint hurdle model and limited sample size in the second stage means results should be treated with caution. I find similar results with separate hurdle models for depth and conditions on military support. Results on file with the author.

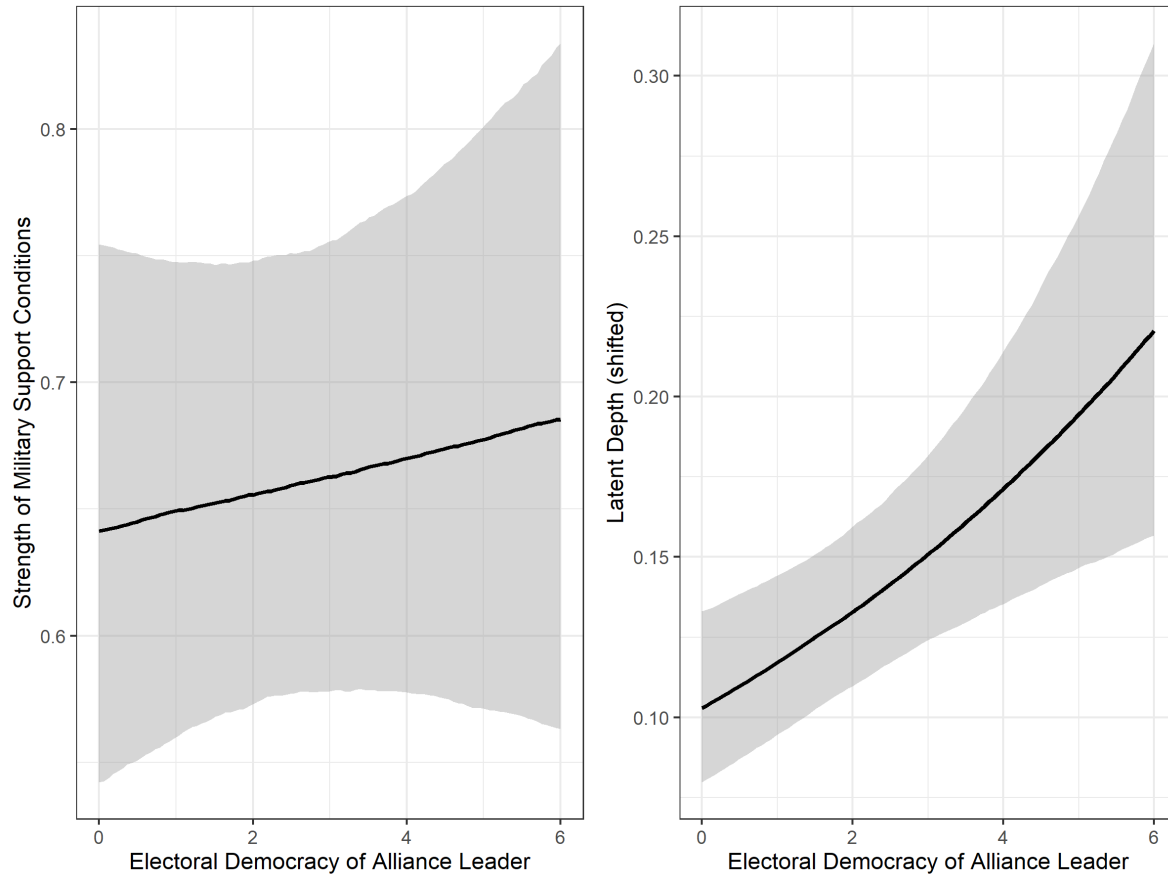


Figure 6: Predicted commitment strength and changes in treaty depth by electoral competition in the most capable alliance member when the treaty formed. Estimates are based on a joint hurdle model of alliances. The line marks predicted values, and the shaded areas encapsulate the 90% credible interval. The rug plot on the x-axis marks observed values of allied democracy. Predictions hold all other variables constant.

## 5 Uncertainty in Latent Treaty Depth

In a further robustness check, I consider how measurement uncertainty shapes inferences about the connection between non-major power alliances and treaty depth. Although there are perceptible differences in treaty depth, especially once states add substantial depth to the treaty, the latent measure of treaty depth varies, as the model produces a distribution of plausible depth values for each alliances. This uncertainty in the latent measure is common to all measurement models and offers a reasonable approximation of alliance politics, because the implications of alliance treaty design are uncertain.

To incorporate uncertainty over treaty depth, I fit a modification of the joint model. First, I created 1,000 alliance datasets, one for each draw of the posterior distribution of the latent measure. All other variables remained the same, but the treaty depth values are unique to each dataset. Then I fit the treaty depth model on 400 randomly sampled datasets from those 1,000 using Bayesian estimation through BRMS (Bürkner, 2017). Joint Bayesian estimation has the flexibility to incorporate the probit and beta models and can be easily extended to account for uncertainty in the depth measure, but it does not allow correlated errors. Fitting the model sequentially to each dataset produces 400 separate models, which I combine into a single model by aggregating the posterior draws into a unified posterior that accounts for uncertainty in the treaty depth measure.<sup>4</sup> This approach is analogous to common techniques for analyzing missing data, where multiple imputation generates uncertainty about the missing values (Hollenbach et al., 2018). After multiple imputation, researchers fit a separate model to each imputed dataset and then combine the results.

Even after accounting for uncertainty in the treaty depth measure, I find a similar pattern, albeit with more uncertainty in the unconditional military support coefficient. I plot the relevant posterior distributions in Figure 7. The 95% credible interval of the association between electoral democracy is uniformly positive and 99% of the posterior mass is positive. For unconditional military support,

---

<sup>4</sup>Standard convergence diagnostics indicate convergence in all 500 models. Diagnostics like  $\hat{r}$  are less useful for the full posterior, because some of the chains in the submodels do not overlap.

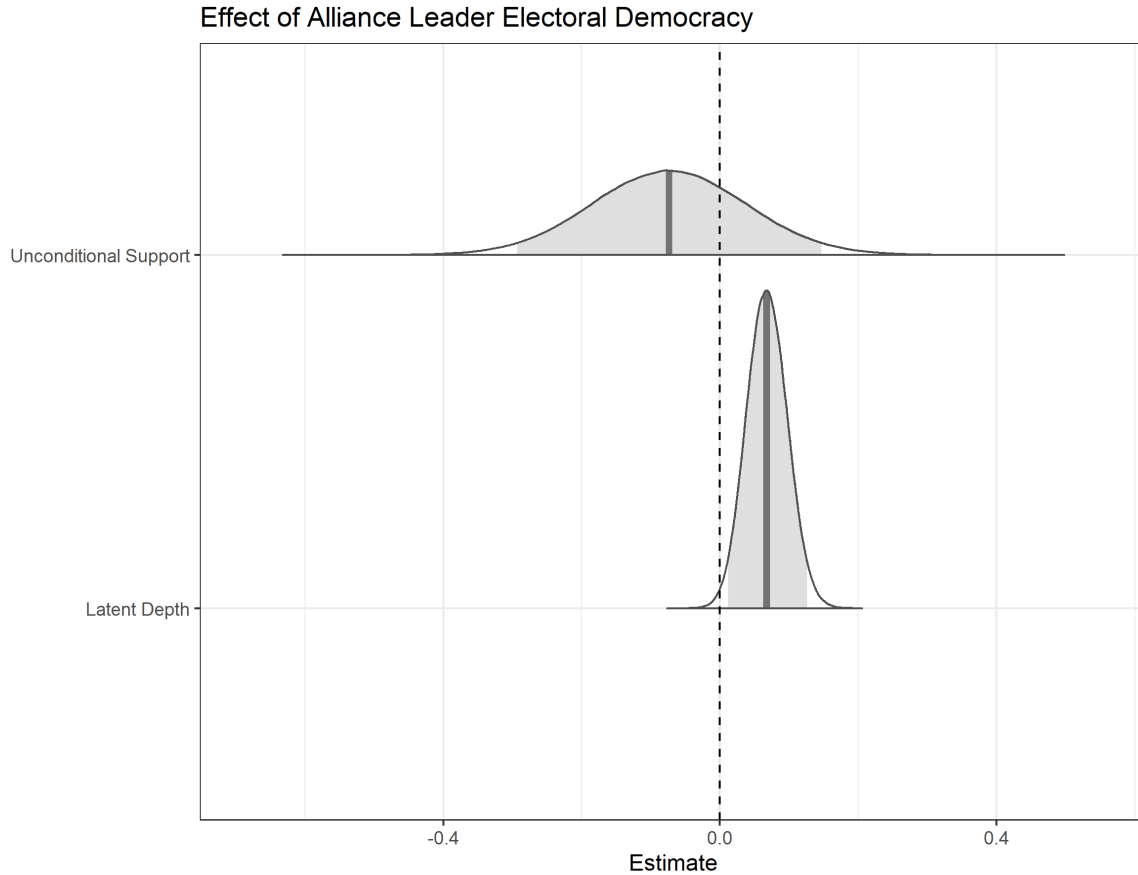


Figure 7: Estimated association between the democracy of the most capable alliance member and two sources of treaty depth from a Bayesian model that accounts for uncertainty in the latent depth measure. The density plot shows the posterior distributions, and the shaded area summarizes the 95% credible interval.

the 95% credible interval has substantial uncertainty, and the alliance leader electoral democracy has 78% negative posterior probability. Thus the preponderance of evidence from these estimates matches the hypotheses, but the effect on unconditional support is weaker than expected.

## 6 Issue Linkages and Treaty Credibility

(Poast, 2013) observes that issue linkages increase alliance credibility in buffer states. Buffer states face significant threats because they are located between competing major powers. which

makes generating credible commitments difficult. Because issue linkages make alliance commitments more credible in this “hard case,” Poast concludes that issue linkages are generally a source of credibility. In the manuscript, I use issue linkages as a control variable and do not discuss them in the theory. I offer minimal discussion of issue linkages to simplify the argument and empirical tests. Examining treaty depth, unconditional military support and issue linkages together is a logical next step in theory development, however. For this paper, I also find a positive association between alliance leader electoral democracy and treaty depth in a trivariate model of treaty depth, unconditional military support and issue linkages.

The trivariate model uses three dummy variables to measure the outcomes, as trivariate GJRM models can only include dummy variables. The depth dummy is equal to one if an alliance has greater than the median value of latent treaty depth. Unconditional military support is equal to one if the alliance places no conditions on military intervention. Last, I set the issue linkages dummy equal to one if the alliance promises economic cooperation.

Even after accounting for correlated errors between models of treaty depth, unconditional military support and issue linkages, I find a positive relationship between the electoral democracy of the most capable alliance member and treaty depth. I also find no clear association between the political institutions of the alliance leader and the probability of unconditional military support. Figure 8 summarizes the predicted probabilities of each credibility source for the values of the lexical index of electoral democracy. Estimates hold all variables besides the electoral democracy of the most capable alliance member at their mode or median.

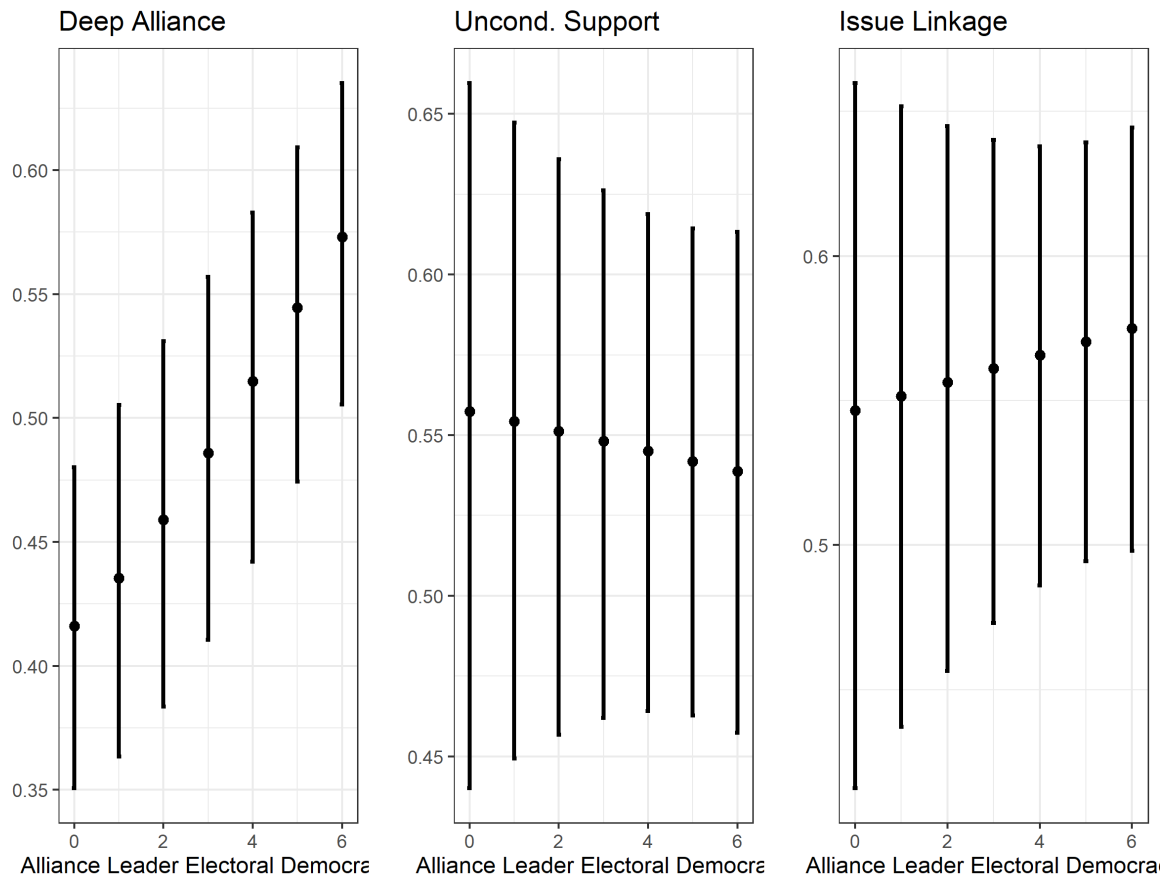


Figure 8: Predicted probabilities an alliance has greater than median treaty depth, unconditional military support or issue linkages for each value of alliance leader electoral democracy. The three outcomes are dummy variables for a deep alliance, unconditional military support, and issue linkages. Predictions based on estimates from a trivariate GJRM with a normal copula.

## References

- Benson, Brett V and Joshua D Clinton. 2016. "Assessing the Variation of Formal Military Alliances." *Journal of Conflict Resolution* 60(5):866–898.
- Berkemeier, Molly and Matthew Fuhrmann. 2018. "Reassessing the fulfillment of alliance commitments in war." *Research & Politics* April-June:1–5.
- Bürkner, Paul-Christian. 2017. "brms: An R package for Bayesian multilevel models using Stan." *Journal of Statistical Software* 80(1):1–28.
- Chiba, Daina, Jesse C Johnson and Brett Ashley Leeds. 2015. "Careful Commitments: Democratic States and Alliance Design." *The Journal of Politics* 77(4):968–982.
- Fordham, Benjamin and Paul Poast. 2014. "All Alliances Are Multilateral Rethinking Alliance Formation." *Journal of Conflict Resolution* 58(27):1–26.
- Hollenbach, Florian, Iavor Bojinov, Shahryar Minhas, Nils W. Metternich, Michael D. Ward and Alexander Volfovsky. 2018. "Principled Imputation Made Simple: Multiple Imputation Using Gaussian Copulas." *Sociological Methods & Research* pp. 1–29.
- Leeds, Brett Ashley and Sezi Anac. 2005. "Alliance Institutionalization and Alliance Performance." *International Interactions* 31(3):183–202.
- Morrow, James D. 2000. "Alliances: Why Write Them Down?" *Annual Review of Political Science* 3:63–83.
- Murray, Jared S, David B Dunson, Lawrence Carin and Joseph E Lucas. 2013. "Bayesian Gaussian Copula Factor Models for Mixed Data." *Journal of the American Statistical Association* 108(502):656–665.
- Poast, Paul. 2010. "(Mis) Using dyadic data to analyze multilateral events." *Political Analysis* 18(4):403–425.
- Poast, Paul. 2013. "Can Issue Linkage Improve Treaty Credibility? Buffer State Alliances as a 'Hard Case'." *Journal of Conflict Resolution* 57(5):739–764.
- Poast, Paul. 2019. *Arguing about Alliances: The Art of Agreement in Military-Pact Negotiations*. Cornell University Press.
- Smith, Alastair. 1995. "Alliance Formation and War." *International Studies Quarterly* 39(4):405–425.