Supplemental Material for *Complementarity in Alliances: How strategic compatibility and hierarchy promote efficient cooperation in international security*

J Andrés Gannon

Table of contents

A. Descriptive statistics

Table A1: Summary statistics of model variables. Year polynomials omitted from table.

	Unique	Missing Pct.	Mean	SD	Min	Median	Max
Year	45	0	1991.27	12.99	1970.00	1990.00	2014.00
Division of labor	2131	4	0.33	0.26	0.00	0.29	1.00
Strategic compatibility	815	0	0.34	0.27	0.00	0.29	1.00
Hierarchy	645	0	0.23	0.24	0.00	0.15	1.00
Peacetime coordination	4	4	0.84	0.84	0.00	1.00	2.00
Democracy ratio	70	2	0.39	0.41	0.00	0.30	1.00
Proportion contiguous	60	0	0.37	0.44	0.00	0.12	1.00
Maximum distance (log)	87	0	5.76	3.92	0.00	7.66	9.85
Number of rivals (log)	38	0	2.06	0.73	0.00	2.20	3.61
Alliance members (log)	31	0	1.17	0.83	0.69	0.69	3.50
Alliance age (avg)	335	0	19.41	14.68	0.00	16.00	64.24

Most of the missing data comes from the division of labor variable and is primarily due to two alliances for which there are many missing values – ATOP ID 3830 (bilateral defense pact between France and Comoros signed in 1978) is missing 37 values and ATOP ID 3900 (treaty establishing the Organization of Eastern Caribbean States between Antigua, Dominica, Grenada, Montserrat, St. Kitts/Nevis, Saint Lucia, and St. Vincent and the Grenadines) is missing 34.

B. Manuscript model supplementary info

Ordered beta results table

Table A2: Log odds coefficient estimates. Time dependencies modeled as year cubic splines given computational constraints for year fixed effects.

	(1)
Strategic Compatibility	2.022
	[1.613, 2.532]
Hierarchy	4.423
	[3.072, 6.470]
Peacetime Coordination	1.109
	[1.046, 1.176]
Democracy Ratio	1.586
	[1.360, 1.855]
Contiguity Ratio	1.161
	[0.760, 1.739]
Maximum Distance (log)	0.985
	[0.935, 1.035]
Number of Rivals (log)	1.409
	[1.282, 1.551]
Number of Members (log)	0.809
	[0.750, 0.870]
Alliance Age (avg)	0.996
	[0.991, 1.000]
Num.Obs.	2280
RMSE	0.22

C. Alternate explanatory variables

Strategic Compatibility

I define strategic compatibility as "consistency of states' security interests and agreement on the nature of the international threat environment". To provide more detail, including exact definitions from the original data sources, state B is considered a part of state A's threat environment if it meets any of the following conditions:

1) Strategic rivalry - states A and B have an un-directed strategic rivalry if they "regard each other as (a) competitors, (b) the source of actual or latent threats that pose some possibility of becoming militarized, and (c) enemies" (Thompson 2001, 560). The data comes from Thompson, Sakuwa, and Suhas (2021, 34–46) and codings are used without modification.

2) Peace scale rivalry - states A and B have an un-directed peace scale relationship coded as "severe rivalry", meaning "the states see one another as enemies and competitors…which encourage rivals to handle their contested issues via frequent and intense uses of violence." or states A and B are non-allied and have an un-directed peace scale relationship coded as "lesser rivalry", meaning that "The sentiments of threat, enmity and competition that remain—along with the persistence of unresolved issues—mean that lesser rivalries still experience isolated violent episodes (e.g. militarized disputes or MIDs), diplomatic hostility, and non-violent crises" (Diehl, Goertz, and Gallegos 2021, 610). I use version 3.1.

3) Politically relevant threat environment - state B is contiguous or a great power *and* it is non-allied with a kappa chance-corrected alliance similarity score below the population median. Contiguity and great power status are a commonly used scope condition for politically relevant dyads (Maoz 1996, 168) and the decision to subset that to dyads with inconsistent foreign policy orientations comes from Leeds and Savun (2007). The modifications I make relative to Leeds and Savun (2007) concern expanding the definition of contiguous to include less than 25 miles of water (M. A. Benson 2005; Bennett 2006; Braumoeller and Carson 2011) and instead of using the population's median s-score as a threshold for similar foreign policy orientation, I use the kappa chance-corrected measure to minimize statistical bias in s-score comparisons (Cohen 1960; Häge 2011; Chiba, Johnson, and Leeds 2015).

There are 1,358,924 directed-dyad years from 1970 to 2014 of which 40,428 categorize state B as being a member of state A's threat environment. Figure A1 shows the overlap among the 3 criteria used to determine each state's threat environment.

Criteria 1 is most similar to Poast (2019, 52–57) and Criteria 3 is most similar to Leeds and Savun (2007, 1127) with the following modifications:

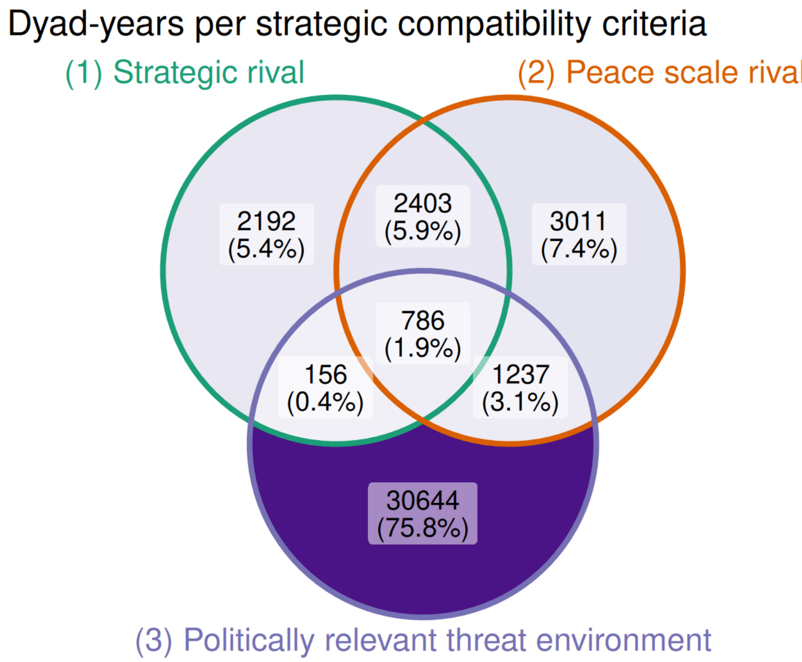


Figure A1: Overlap among strategic compatibility criteria. Counts refer to number of dyad-years.

- Poast (2019) uses a count of states as threats, while I weight each threat by their CINC score. If all NATO states view Russia as a threat and only the US views Cuba as a threat, the relative military salience of Russia and Cuba should be considered.
- Poast (2019, 93) addresses the absence of asymmetric rivalries in Thompson (2001) by adding contiguous great powers as a strategic rival. I address that through Criteria 3 in my measure.
- Poast (2019) makes strategic compatibility binary with high (low) strategic compatibility being above (below) the sample median, I keep it continuous and scale it by the highest observed value in the full population.
- Poast (2019) combines strategic compatibility (what I measure) with operational compatibility which addresses agreement about how to use military force against a threat based on offensive or defense war time doctrine (Bennett and Stam 1996; Stam 1996; Reiter and Stam 1998). I do not include a measure of operational compatibility given its similarity to the dependent variable, lack of vary within alliances, and absence of data for my time period.
- Leeds and Savun (2007) use un-weighted ATOP s-score, but I use the kappa-corrected s-score measure as suggested by (Häge 2011; Cohen 1960).

- Leeds and Savun (2007) exclude all alliance partners from a state's threat environment. Since allies do sometimes have serious rivalries and conflicts (Greece-Turkey in NATO, India-Pakistan in SCO, Russia-Georgia, Russia-Ukraine, Armenia-Azerbaijan in CIS, etc) I allow an ally to be in your threat environment if it is coded as a strategic rivalry or a peace level "serious rivalry" (Weitsman 1997; Bearce, Flanagan, and Floros 2006).

I find consistent results using weighted and un-weighted s-scores measuring alliance portfolio similarity using ATOP which has previously been used to measure variation in strength of alliance ties (Gibler and Rider 2004; Chiba, Johnson, and Leeds 2015; Fordham and Poast 2016). I opt against using UN voting similarity as a measure of strategic compatibility because of recognized problems with chance agreement (Häge 2011) the high frequency of consensus decisions (Häge and Hug 2016), and agenda effects bias year-over-year change (Bailey, Strezhnev, and Voeten 2017).

Table A3: Coefficient estimates with alternate variables for strategic compatibility.

	(1)	(2)
S-score (weighted)	0.320**	
	(0.099)	
S-score (unweighted)		0.422*
		(0.193)
Hierarchy	0.367**	0.372**
	(0.113)	(0.124)
Democracy Ratio	0.093	0.106*
	(0.048)	(0.053)
Contiguity Ratio	0.101	0.213
	(0.113)	(0.130)
Maximum Distance (log)	0.009	0.026
	(0.014)	(0.017)
Number of Members (log)	-0.045	-0.050
	(0.024)	(0.029)
Alliance Age (avg)	0.002	0.001
	(0.001)	(0.002)
Num.Obs.	2381	2381
R2	0.211	0.197
R2 Adj.	0.207	0.194
AIC	-240.3	-199.9
BIC	-176.8	-136.3
RMSE	0.23	0.23

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
All models include alliance-clustered standard errors.

Hierarchy

Alliance power asymmetry has been measured using a variety of similar measures. While none are identical to the new measure of hierarchy created, one would still expect the coefficient sign to be consistent. These measures are defined in the original texts as follows:

- potential military capacity (B. V. Benson and Clinton 2016, 873–75): log of the summed CINC score of all alliance members, presence of a major power, mean geographic distance between all alliance pairs, log of the number of alliance members, mean s-score of all pairs of alliance members, and the average Polity IV score of all alliance members.
- alliance depth (B. V. Benson and Clinton 2016, 872–73): "military contact, common defense policy, integrated command, military aid, military basing, specific contribution, organization, economic aid, and secret."
- peacetime coordination (Leeds and Anac 2005, 188–91): ordinal variable coded *high* if alliances have an integrated military command during peacetime and wartime, have a common defense policy obligation, and have joint troop placement agreements. Coded *moderate* if peacetime official military contact is required, there is a formal military organization for coordination, one party is required to provide training or technology to another, and there are specific plans to subordinate one military to another during conflict. Coded *low* otherwise.
- latent alliance depth (Alley 2021, 932): ordinal factor analysis measuring "additional policy coordination and military cooperation beyond a promise of military support. Defense cooperation in a deep alliance can take many forms, including an integrated military command, military aid, common defense policies, basing rights, international organizations, specific contribution requirements or companion military agreements."
- CINC dispersion (Gibler and Rider 2004, 316): "captures the dispersal of capabilities in an alliance; it ranges from 0 (when the capabilities in the alliance are equally distributed) to 1 (when one state holds all of the capabilities in the alliance). In this formula, $\sqrt{\sum_{i=1}^{n}(Si)^2 - \frac{1}{n}/(1-\frac{1}{n})}$, Si denotes i's share of the alliance's total capabilities, and n is the number of allies in the alliance."
- CINC asymmetry (B. V. Benson and Clinton 2016, 888): "measured by subtracting the CINC scores of the strongest and weakest pair of alliance members."
- great power participation (Gibler and Rider 2004; Leeds and Savun 2007, 188): Ratio variable indicating how many alliance members are great powers.

Table A4: Coefficient estimates with alternate variables for hierarchy.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Strategic Compatibility	0.243***	0.066	0.062	0.054	0.130*	0.114*	0.151***
	(0.059)	(0.046)	(0.044)	(0.045)	(0.051)	(0.044)	(0.040)
Potential Military Capacity	0.095***						
	(0.020)						
Alliance Depth		0.059***					
		(0.012)					
Peacetime Coord			0.049**				
			(0.018)				
Latent Depth				0.032			
				(0.020)			
CINC Dispersal					0.081**		
					(0.028)		
CINC Asymmetry						0.574*	
						(0.286)	
Great power ratio							0.260**
							(0.088)
Democracy Ratio	0.118**	0.159***	0.173***	0.136**	0.162***	0.141**	0.122**
	(0.044)	(0.044)	(0.039)	(0.045)	(0.045)	(0.043)	(0.043)
Contiguity Ratio	-0.167	-0.154	-0.193	-0.139	-0.041	-0.086	-0.095
	(0.112)	(0.120)	(0.112)	(0.122)	(0.121)	(0.115)	(0.113)
Maximum Distance (log)	-0.023	-0.027	-0.030*	-0.020	-0.007	-0.011	-0.015
	(0.015)	(0.016)	(0.015)	(0.016)	(0.017)	(0.016)	(0.016)
Number of Rivals (log)	0.030	0.097***	0.092***	0.092***	0.060*	0.063*	0.042
	(0.031)	(0.022)	(0.022)	(0.023)	(0.027)	(0.025)	(0.028)
Number of Members (log)	0.000	-0.010	-0.005	-0.022	-0.049	-0.012	0.033
	(0.018)	(0.019)	(0.015)	(0.020)	(0.026)	(0.020)	(0.028)
Alliance Age (avg)	0.000	0.003*	0.002	0.002	0.001	0.000	0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Num.Obs.	2007	2007	2280	2339	2379	2379	2379
R2	0.279	0.280	0.263	0.222	0.239	0.227	0.244
R2 Adj.	0.275	0.276	0.260	0.218	0.235	0.224	0.240
AIC	-326.6	-330.5	-417.7	-251.9	-324.3	-288.2	-339.7
BIC	-259.4	-263.2	-348.9	-182.8	-255.0	-218.9	-270.5
RMSE	0.22	0.22	0.22	0.23	0.22	0.23	0.22

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
All models include alliance-clustered standard errors.

Interaction term

Table A5: Coefficient estimates with interaction term.

	Year Polynomials		Year FE		Multilevel Model	
	(1)	(2)	(3)	(4)	(5)	(6)
Strategic Compatibility	0.219***	0.183*	0.220***	0.191*	0.221***	0.189***
	(0.051)	(0.090)	(0.053)	(0.092)	(0.023)	(0.030)
Hierarchy	0.479***	0.385*	0.462***	0.372*	0.452***	0.373***
	(0.072)	(0.167)	(0.073)	(0.170)	(0.037)	(0.057)
Strat Comp*Hierarchy	-0.239	-0.155	-0.205	-0.139	-0.175	-0.110
	(0.198)	(0.200)	(0.197)	(0.203)	(0.096)	(0.099)
Democracy Ratio		0.116*		0.115*		0.115***
		(0.046)		(0.047)		(0.017)
Contiguity Ratio		-0.131		-0.144		-0.146**
		(0.106)		(0.105)		(0.045)
Maximum Distance (log)		-0.025		-0.027		-0.026***
		(0.015)		(0.015)		(0.006)
Number of Rivals (log)		0.061*		0.057		0.064***
		(0.029)		(0.029)		(0.010)
Number of Members (log)		-0.029		-0.028		-0.032***
		(0.019)		(0.019)		(0.008)
Alliance Age (avg)		0.000		0.000		0.000
		(0.001)		(0.001)		(0.000)
Num.Obs.	2398	2280	2398	2280	2398	2280
R2	0.208	0.285	0.234	0.307		
R2 Adj.	0.206	0.281	0.219	0.290		
AIC	-230.6	-482.7	-228.8	-472.2	-156.8	-338.4
BIC	-190.1	-402.4	48.7	-156.9	-122.2	-263.9
RMSE	0.23	0.22	0.23	0.21	0.23	0.21

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
All models include alliance clustered standard errors.

D. Alternate and additional controls

Alternate controls for democracy

The models in the manuscript operationalize democracy as the proportion of alliance members that have a Polity score greater than 6. The results are robust to other operationalizations of regime type commonly used in the literature like average Polity score (B. V. Benson and Clinton 2016), lowest Polity score based on the "weakest link" principle (Oneal and Russett 1997; Fordham and Poast 2016), largest difference in Polity scores (Gibler and Rider 2004; Fordham and Poast 2016), and whether all members are of the same regime type. To address missingness in Polity data, I also includes controls using the Marquez (2016) extension of the UDS latent variable measure (Pemstein, Meserve, and Melton 2017).

Table A6: Coefficient estimates with alternate democracy variables.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Strategic Compatibility	0.163**	0.160*	0.178**	0.166*	0.164**	0.158*	0.180**
	(0.059)	(0.063)	(0.062)	(0.064)	(0.057)	(0.064)	(0.058)
Hierarchy	0.308*	0.397***	0.433***	0.421***	0.333**	0.413***	0.396***
	(0.120)	(0.109)	(0.111)	(0.113)	(0.123)	(0.114)	(0.107)
Polity (avg)	0.008**						
	(0.003)						
Polity (lowest)		0.003					
		(0.002)					
Polity (difference)			0.005				
			(0.003)				
Same regime type				-0.045			
				(0.032)			
UDS (avg)					0.067*		
					(0.026)		
UDS (lowest)						0.025	
						(0.019)	
UDS (difference)							0.088**
							(0.029)
Peacetime Coordination	0.038*	0.040*	0.034*	0.036*	0.034*	0.038*	0.028
	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)	(0.017)	(0.016)
Contiguity Ratio	-0.145	-0.091	-0.086	-0.069	-0.070	-0.037	-0.030
	(0.109)	(0.103)	(0.109)	(0.107)	(0.117)	(0.107)	(0.099)
Maximum Distance (log)	-0.027	-0.019	-0.018	-0.016	-0.019	-0.013	-0.012
	(0.015)	(0.014)	(0.015)	(0.015)	(0.015)	(0.014)	(0.014)
Number of Rivals (log)	0.065*	0.058*	0.038	0.045	0.067*	0.057*	0.033
	(0.027)	(0.027)	(0.024)	(0.024)	(0.028)	(0.028)	(0.023)
Number of Members (log)	-0.027	-0.030	-0.037	-0.034	-0.041	-0.037	-0.063**
	(0.020)	(0.019)	(0.020)	(0.021)	(0.021)	(0.020)	(0.020)
Alliance Age (avg)	0.000	0.000	0.001	0.001	0.000	0.000	0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Num.Obs.	2280	2280	2280	2280	2299	2299	2299
R2	0.288	0.273	0.278	0.274	0.295	0.277	0.307
R2 Adj.	0.284	0.269	0.274	0.270	0.291	0.273	0.303
AIC	-493.7	-445.5	-462.1	-449.2	-510.6	-454.0	-550.8
BIC	-419.2	-371.0	-387.6	-374.7	-436.0	-379.4	-476.2
RMSE	0.22	0.22	0.22	0.22	0.22	0.22	0.21

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
All models include alliance-clustered standard errors.

Alternate controls for geography

Instead of the maximum alliance-dyad distance, Bak (2018) measures the logged mean capital-to-capital distance between all alliance-dyads, treating contiguous states as distance = 0. Model (1) uses Bak (2018)' distance measure exclusively and Model (2) uses it as a replacement for maximum distance.

Table A7: Coefficient estimates with alternate geography variables.

	(1)	(2)
Strategic Compatibility	0.154*	0.155*
	(0.062)	(0.061)
Hierarchy	0.344**	0.333**
	(0.116)	(0.117)
Peacetime Coordination	0.035*	0.038*
	(0.017)	(0.017)
Democracy Ratio	0.108*	0.115*
	(0.044)	(0.045)
Contiguity Ratio		-0.130
		(0.119)
Mean Distance (log)	-0.009	-0.025
	(0.005)	(0.016)
Number of Rivals (log)	0.056*	0.063*
	(0.025)	(0.027)
Number of Members (log)	-0.044*	-0.042*
	(0.017)	(0.017)
Alliance Age (avg)	0.000	0.000
	(0.001)	(0.001)
Num.Obs.	2280	2280
R2	0.281	0.283
R2 Adj.	0.278	0.279
AIC	-474.0	-478.7
BIC	-405.2	-404.2
RMSE	0.22	0.22

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
All models include alliance-clustered standard errors.

Additional controls for NATO, US, and Cold War

Table A8: Coefficient estimates with additional control variables.

	(1)	(2)	(3)
Strategic Compatibility	0.140*	0.166**	0.161*
	(0.066)	(0.059)	(0.062)
Hierarchy	0.311*	0.426***	0.323**
	(0.121)	(0.114)	(0.119)
Peacetime Coordination	0.042*	0.029	0.039*
	(0.018)	(0.016)	(0.017)
Democracy Ratio	0.121*	0.140**	0.113*
	(0.046)	(0.050)	(0.046)
Contiguity Ratio	-0.128	-0.089	-0.133
	(0.107)	(0.111)	(0.105)
Maximum Distance (log)	-0.025	-0.018	-0.025
	(0.015)	(0.015)	(0.014)
Number of Rivals (log)	0.066*	0.074**	0.064*
	(0.027)	(0.026)	(0.027)
Number of Members (log)	-0.022	-0.046*	-0.029
	(0.021)	(0.019)	(0.019)
Alliance Age (avg)	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)
NATO/Warsaw Pact	-0.055		
	(0.042)		
United States member		-0.096*	
		(0.046)	
Cold War			-0.070*
			(0.032)
Num.Obs.	2280	2280	2280
R2	0.286	0.292	0.287
R2 Adj.	0.282	0.288	0.283
AIC	-485.1	-505.3	-488.6
BIC	-404.9	-425.1	-408.3
RMSE	0.22	0.22	0.22

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
All models include alliance-clustered standard errors.

E. Other model specifications

Panel-corrected standard errors

Table A9: Coefficient estimates using panel-corrected standard errors.

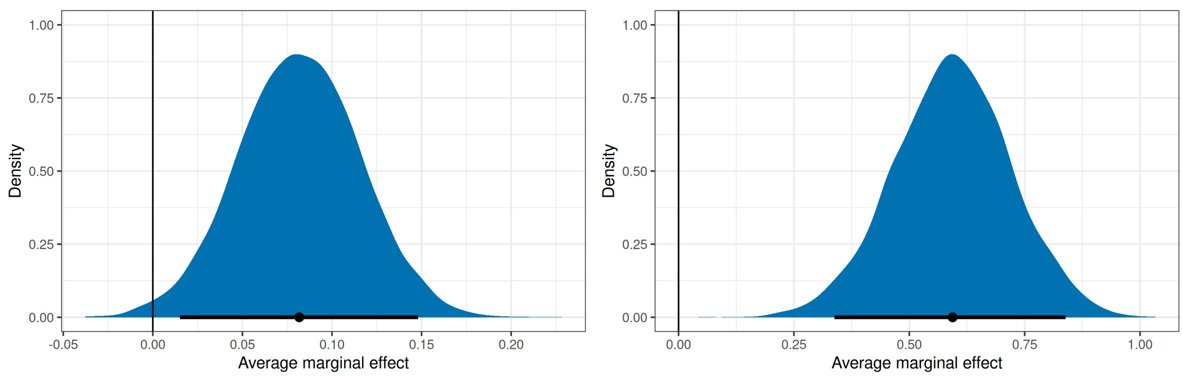
	Year Polynomials		Year Fixed Effects	
	(1)	(2)	(3)	(4)
Strategic Compatibility	0.186***	0.153*	0.192***	0.165*
	(0.049)	(0.065)	(0.050)	(0.066)
Hierarchy	0.403***	0.320*	0.397***	0.313*
	(0.073)	(0.122)	(0.072)	(0.123)
Peacetime Coordination		0.039*		0.039*
		(0.017)		(0.017)
Democracy Ratio		0.118*		0.116*
		(0.046)		(0.047)
Contiguity Ratio		-0.127		-0.140
		(0.117)		(0.117)
Maximum Distance (log)		-0.025		-0.026
		(0.016)		(0.016)
Number of Rivals (log)		0.064*		0.060*
		(0.028)		(0.028)
Number of Members (log)		-0.029		-0.027
		(0.023)		(0.023)
Alliance Age (avg)		0.000		0.000
		(0.001)		(0.001)
Num.Obs.	2398	2280	2398	2280
R2	0.206	0.284	0.232	0.307
R2 Adj.	0.204	0.281	0.217	0.290
AIC	-226.4	-482.2	-226.2	-472.2
BIC	-191.8	-407.7	45.5	-162.7
RMSE	0.23	0.22	0.23	0.21

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
All models cluster standard errors by alliance and year.

Bayesian multi-level model

A Bayesian multi-level model produces posterior distributions for the expected division of labor conditional on strategic compatibility and hierarchy (Bürkner 2017). Since alliances may differ in both the baseline expectation for the dependent variable as well as how they change over time, a Bayesian multi-level model accounts for alliance-level heterogeneity by allowing a random intercept and slope for each alliance nested within each year (Shor et al. 2007). As a result, alliance-specific deviations from the population average caused by things like alliance design or treaty depth are accounted for.

Figure A2 shows the results of the model run on 4 MCMC chains with 4,000 iterations per chain. The marginal trend is calculated for each independent variable conditional on the random effects of each alliance. In these results, strategic compatibility has a strong positive association with division of labor where the model results suggest greater than a 95% probability that the true marginal effect is greater than 0. The same is true for hierarchy.



(a) Strategic Compatibility

(b) Hierarchy

Figure A2: Average marginal effect on division of labor of a one standard deviation increase in the independent variables. Point estimates centered at the median value with 0.95 credible intervals. Note x-axes are not fixed.

Double/debiased Machine-learning

Double machine-learning is designed to solve the problem of omitted variable bias by using the residuals of the control variables and independent variables to separately predict the dependent variable then splitting the sample by cross-fitting (Bach et al. 2024).[1] Table A10 show the results, which are consistent with the original estimates.

First, the partialling out score function computes the residuals from a regression of the dependent variable on the control variables then a regression of the dependent variable on the independent variables. I then run an OLS of predicting the first stage from the second which creates valid post-selection inference even in cases of imperfect model selection by creating a "second chance" for omitted variables (Belloni, Chernozhukov, and Hansen 2014).

This is performed on samples generated using 5-fold re-sampling with one repeated sample split. The models are cross-fit, meaning I first split the sample into an auxiliary and a main and estimate the model on the auxiliary sample then perform the partialling out estimation by OLS on the main sample. I then reverse the two samples and take the average of the results, which prevents overfitting (Chernozhukov et al. 2018).

Table A10: Double/debiased machine learning model.

	(1)
Strategic compatibility	0.076*
	(0.036)
Hierarchy	0.614***
	(0.103)
score	partialling out
n.folds	5
cross.fitting	TRUE

* p <0.05, ** p <0.01, *** p <0.001

[1]Because this model can only operate on complete cases, I use the UDS democracy score (Marquez 2016; Pemstein, Meserve, and Melton 2017) instead of Polity to minimize missing values for the regime type control variable.

Works Cited

Alley, Joshua. 2021. "Alliance Participation, Treaty Depth, and Military Spending." *International Studies Quarterly* 65 (4): 929–43. https://doi.org/10.1093/isq/sqab077.

Bach, Philipp, Malte S. Kurz, Victor Chernozhukov, Martin Spindler, and Sven Klaassen. 2024. "DoubleML: An Object-Oriented Implementation of Double Machine Learning in R." *Journal of Statistical Software* 108 (3): 1–56. https://doi.org/gt3cht.

Bailey, Michael A., Anton Strezhnev, and Erik Voeten. 2017. "Estimating Dynamic State Preferences from United Nations Voting Data." *Journal of Conflict Resolution* 61 (2): 430–56. https://doi.org/10.1177/0022002715595700.

Bak, Daehee. 2018. "Alliance Proximity and Effectiveness of Extended Deterrence." *International Interactions* 44 (1): 107–31. https://doi.org/10.1080/03050629.2017.1320995.

Bearce, David H., Kristen M. Flanagan, and Katharine M. Floros. 2006. "Alliances, Internal Information, and Military Conflict Among Member-States." *International Organization* 60 (3): 595–625. https://doi.org/10.1017/S0020818306060188.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on Treatment Effects After Selection Among High-Dimensional Controls." *The Review of Economic Studies* 81 (2): 608–50. https://doi.org/f53wkb.

Bennett, D. Scott. 2006. "Exploring Operationalizations of Political Relevance." *Conflict Management and Peace Science* 23 (3): 245–61. https://doi.org/10.1080/07388940600837748.

Bennett, D. Scott, and Allan C. Stam. 1996. "The Duration of Interstate Wars, 1816–1985." *American Political Science Review* 90 (2): 239–57. https://doi.org/b6zbbd.

Benson, Brett V., and Joshua D. Clinton. 2016. "Assessing the Variation of Formal Military Alliances." *Journal of Conflict Resolution* 60 (5): 866–98. https://doi.org/10.1177/0022002714560348.

Benson, Michelle A. 2005. "The Relevance of Politically Relevant Dyads in the Study of Interdependence and Dyadic Disputes." *Conflict Management and Peace Science* 22 (2): 113–33. https://doi.org/10.1080/07388940590948556.

Braumoeller, Bear F., and Austin Carson. 2011. "Political Irrelevance, Democracy, and the Limits of Militarized Conflict." *Journal of Conflict Resolution* 55 (2): 292–320. https://doi.org/10.1177/0022002710393918.

Bürkner, Paul-Christian. 2017. "Brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80 (1): 1–28. https://doi.org/gddxwp.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *The Econometrics Journal* 21 (1): 1–68. https://doi.org/gc2snx.

Chiba, Daina, Jesse C. Johnson, and Brett Ashley Leeds. 2015. "Careful Commitments: Democratic States and Alliance Design." *The Journal of Politics* 77 (4): 968–82. https://doi.org/10.1086/682074.

Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and*

Psychological Measurement 20 (1): 37–46. https://doi.org/dghsrr.
Diehl, Paul F., Gary Goertz, and Yahve Gallegos. 2021. "Peace Data: Concept, Measurement, Patterns, and Research Agenda." *Conflict Management and Peace Science* 38 (5): 605–24. https://doi.org/10.1177/0738894219870288.
Fordham, Benjamin O., and Paul Poast. 2016. "All Alliances Are Multilateral: Rethinking Alliance Formation." *Journal of Conflict Resolution* 60 (5): 840–65. https://doi.org/10.1177/0022002714553108.
Gibler, Douglas M., and Toby J. Rider. 2004. "Prior Commitments: Compatible Interests Versus Capabilities in Alliance Behavior." *International Interactions* 30 (4): 309–29. https://doi.org/10.1080/03050620490883985.
Häge, Frank M. 2011. "Choice or Circumstance? Adjusting Measures of Foreign Policy Similarity for Chance Agreement." *Political Analysis* 19 (3): 287–305. https://www.jstor.org/stable/23011439.
Häge, Frank M., and Simon Hug. 2016. "Consensus Decisions and Similarity Measures in International Organizations." *International Interactions* 42 (3): 503–29. https://doi.org/gtv3zr.
Leeds, Brett Ashley, and Sezi Anac. 2005. "Alliance Institutionalization and Alliance Performance." *International Interactions* 31 (3): 183–202. https://doi.org/10.1080/03050620500294135.
Leeds, Brett Ashley, and Burcu Savun. 2007. "Terminating Alliances: Why Do States Abrogate Agreements?" *The Journal of Politics* 69 (4): 1118–32. https://doi.org/10.1111/j.1468-2508.2007.00612.x.
Maoz, Zeev. 1996. *Domestic Sources of Global Change*. Ann Arbor, MI: University of Michigan Press. https://doi.org/10.3998/mpub.14663.
Marquez, Xavier. 2016. "A Quick Method for Extending the Unified Democracy Scores." {{SSRN Scholarly Paper}}. Rochester, NY. https://doi.org/10.2139/ssrn.2753830.
Oneal, John R., and Bruce M. Russett. 1997. "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985." *International Studies Quarterly* 41 (2): 267–93. https://doi.org/cnd4m3.
Pemstein, Daniel, Stephen A. Meserve, and James Melton. 2017. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18 (4): 426–49. https://doi.org/bq365m.
Poast, Paul. 2019. *Arguing about Alliances: The Art of Agreement in Military-Pact Negotiations*. Ithaca: Cornell University Press.
Reiter, Dan, and Allan C. Stam. 1998. "Democracy, War Initiation, and Victory." *The American Political Science Review* 92 (2): 377–89. https://doi.org/10.2307/2585670.
Shor, Boris, Joseph Bafumi, Luke Keele, and David Park. 2007. "A Bayesian Multilevel Modeling Approach to Time-Series Cross-Sectional Data." *Political Analysis* 15 (2): 165–81. https://doi.org/bq4db5.
Stam, Allan C. 1996. *Win, Lose, Or Draw: Domestic Politics and the Crucible of War*. University of Michigan Press.
Thompson, William R. 2001. "Identifying Rivals and Rivalries in World Politics." *International Studies Quarterly* 45 (4): 557–86. https://www.jstor.org/stable/3096060.

Thompson, William R., Kentaro Sakuwa, and Prashant Hosur Suhas. 2021. *Analyzing Strategic Rivalries in World Politics: Types of Rivalry, Regional Variation, and Escalation/De-escalation*. Springer Nature.

Weitsman, Patricia A. 1997. "Intimate Enemies: The Politics of Peacetime Alliances." *Security Studies* 7 (1): 156–93. https://doi.org/10.1080/09636419708429337.