

Revise & Resubmit Memo for AJPS-48033 "Complementarity in Alliances: How strategic compatibility and hierarchy promote efficient cooperation in international security"

Dear *American Journal of Political Science* Editorial Board and reviewers,

I am grateful to the editorial team and reviewers for the opportunity to revise and resubmit the manuscript and for their thoughtful feedback. Please find my memo indicating changes made to address editor and reviewer comments. I am pleased with how the paper has improved and hope you all feel the same way.

The memo is organized by each reviewer comment formatted in the following way:

1. *Reviewer suggestion: italicized and provided verbatim*
2. Author comment: summary of the changes I have made and an explanation for the reviewer
3. Revised text manuscript highlighted in gray and in paragraph justified. Where appropriate, I have duplicated figures, tables, or new verbatim manuscript/appendix text. Note that the page numbers are approximate, formatting inconsistent due to pdf-to-word pasting, and footnotes provided in-text and shrunk to keep numbering consistent with the manuscript.

Every reviewer comment is on a new page, comments are duplicated, and new manuscript text provided here verbatim. Although it increases the memo page count beyond the suggested 5 pages, I hope this eases the re-reviewing process by minimizing back and forth between the memo and the revised manuscript and makes the memo easier to read through.

Thank you,

The author

Editorial Board

Author comments on editorial board suggestions:¹

New sections have been added to the literature review and conclusion clarifying the relationship to existing literature. The main clarification has been explaining that this paper is about the conditions under which alliances are successful in specializing and also that it identifies when specialization is complementary (division of labor) as that requires intra-alliance bargaining to coordinate that. The importance of specialization is now explained in more detail on p 8 and in the conclusion. I refer to empirical work on those topics and provide some specific examples that are illustrative of the ramifications for military efficacy, resource allocation, and domestic politics

Revised paragraph in literature review (p 6-7):

Conflicting empirical findings stem from inattention to the composition of arms aggregated across alliances. Most studies homogenize alliances (Becker et al. 2024; Gannon 2024) or aggregate defense spending (Alley 2021). As a result, the state of the art insufficiently explains why states facing a common threat vary in how they aggregate distinct capabilities against that threat. Gannon (2024) links specialization to alliances, assuming specialization occurs due to its economic and functional benefits. But the presence of alliances does not automatically induce specialization; it is instead contingent on characteristics of the alliance conducive to that specialization and a bargaining process that ensures individual specializations are collectively complementary. Since different capabilities have different public and private benefits, states have to actively ensure coordination so that everyone does not specialize in the same capabilities.

Understanding the conditions under which an alliance can reliably engage in collective defense as well as how that changes the strategic composition of arms across allied states is a worth-while endeavor, as how each state contributes to collective security has long been (Olson and Zeckhauser 1966; Christensen and Snyder 1990; Palmer 1990a; Snyder 1997), and remains, a central question in international politics (Fuhrmann 2020; Blankenship 2023; Kinne and Kang 2023).^{(fn 4 On alliance reliability generally, see Morrow (1994b); Leeds, Long, and Mitchell (2000); Leeds (2003a); Leeds (2003b); Leeds and Savun (2007); B. V. Benson (2012); Crescenzi et al. (2012); Mattes (2012a); Mattes (2012b); B. V. Benson, Meirowitz, and Ramsay (2014); Blankenship (2020); B. C. Smith (2021); B. V. Benson and Smith (2023).)}

Aside from the individual economic benefits of specialization, differences in how alliances bring military resources together has important implications for the outcome of conflict (Conybeare, Murdoch, and Sandler 1994; Bensahel 2007; Cappella Zielinski and Poast 2021; Lanoszka 2022, 118–23) as well as the political and economic costs of defense production (Hartley 1987; De Vestel 1995). Alliances with complementary military portfolios may have more effective warfighting forces because each member can focus on a smaller set of competencies, making them more successful at task execution. One reason the Western response the Iran-Iraq Tanker War was effective was because the Dutch and Belgian navies had reconfigured their navies to specialize in mine countermeasures while Norway and Denmark

¹ Exact text of editor's decision letter omitted at the request of the editorial board.

developed coastal water capabilities. Both of these decisions “took some of the pressure off the US Navy which could therefore afford to concentrate on high intensity operations on the open ocean” (Till 2005, 316). Through effective intra-alliance bargaining that avoids overlap capability, alliances can generate more military output per unit of spending, ultimately improving force readiness and combat effectiveness.

Because such a division of labor reduces costs via economies of scale, it also serves an important domestic political purpose. President Ronald Reagan’s Secretary of Defense Caspar Weinburger (1981) noted that the alternative to working side by side with NATO allies for naval responsibilities risked the US “marching alone”, which would mean the United States “could lose at home the critical public support for which we have labored long and hard.” Notably, such a division of labor can also be a political liability precisely because it involves working side by side. Canadian reliance on United Kingdom logistics and transport vehicles during the 2001 Afghanistan war is exemplary of this phenomenon (Auerswald and Saideman 2014, 232), as were European concerns that dependence on the United States for advanced munitions and radar during the Kosovo War exposed one-sided reliance, prompting the former German ambassador to the US von Kyaw (1999) to note “if we are not careful, we Europeans will become the Hessians of the Americans...we never have a guarantee against neoisolationist developments. Nobody knows how Congress will decide. Nobody knows what sort of U.S. president we might have.” Although the focus of this article is on the causes of a division of labor across allies rather than its consequences, the enduring political significance of this phenomenon motivates its study.

New paragraph in conclusion (p 32-33):

Successful defense cooperation involves an intra-alliance bargaining process that distributes the gains of coordinating complementary military capabilities (Snidal 1985; Fearon 1998). This study provides insights into how alliances overcome obstacles to coordinate defense portfolios. States can specialize in military capabilities, but must negotiate terms of cooperation, determining specialization. By examining the role of strategic compatibility and hierarchy, I explain why some alliances can achieve a division of labor, while others remain trapped in redundancy. Alliances that exhibit high levels of strategic compatibility are more likely to coordinate effectively. Additionally, alliances with hierarchical decision-making structures are better equipped to manage coordination problems and the risk of defection.

Reviewer 1

R1 Major comment #1: I think the framing around collective action in the introduction is a major weakness. The paper is strong enough to stand on a more focused alliance politics framing, and the current framing is both disconnected from the argument and overstates the applicability of collective action models to alliances.

1. The first is that the argument and emphasis on collective action in the introduction are somewhat at odds. Collective action is often understood in terms of a problem where hierarchy is absent, so there is no means of enforcing agreements. Ostrom's work on this is particularly relevant, as it addresses solutions to collective action in the absence of a centralized authority (covenants without the sword). So given greater hierarchy, it's almost tautological to say that collective action problems diminish.

2. It's also not clear to me that alliances are actually a good domain for testing the extent of collective action problems. The Olson and Zeckhauser paper everyone cites on this has huge empirical problems (see Alley 2021 R&P), and extensions outside of NATO yield inconsistent evidence for it's predictions at best. Rather than collective action, it's more fruitful to conceptualize alliance relations and defense spending in terms of bargaining, which is what the author does anyways. Strong collective action arguments like Olson and Zeckhauser's rule out bargaining almost entirely.

3. Taken together, I think a more compelling frame would be the dilemma the author mentions briefly in the introduction and throughout the argument- members gain from specialization, but specializing is risky if your allies abandon you. The big gain from specialization is reducing the opportunity costs of spending on defense, a "regrettable necessity" to quote Nordhaus and Tobin. There's also a clear connection to make to Fearon's 2018 IO on the costs of anarchy. In addressing this problem, alliance members bargain over the distribution of defense capabilities (pg 8). This would also match the argument.

Author comment:

The first two paragraphs of the introduction have been entirely rewritten. The emphasis is no longer on collective action and is instead framed around intra-alliance bargaining and the puzzle of specialization having individual gains, but being risky. The two pieces the reviewer suggested referencing have been added, as has a reference to the canonical Powell piece on guns, butter, and anarchy as it also explains the costs of anarchy in a similar way to Fearon 2018

Revised text manuscript (pg 2):

A fundamental challenge for military alliances is distributing defense capabilities efficiently among member states. Specializing in capabilities reduces the opportunity costs of defense spending, termed a “regrettable necessity” by Nordhaus and Tobin (1973, 515–17).^(fn 1) On the societal costs of defense spending and international security, see also Powell (1993) and Fearon (2018). However, specialization risks arise if allies fail to fulfill their commitments, leaving one vulnerable (Snyder 1984; Palmer 1990b) and arming with allies in mind may also conflict with domestic political priorities (Kurth 1973; Fordham 2019). This creates a dilemma: specialization yields economic and operational efficiencies, but requires deciding each partner's role and

managing coordination costs and risks (Niou and Ordeshook 1994; Papayoanou 1997). What factors shape whether an alliance can reliably engage in collective defense, and how does that change the strategic composition of arms across allied states?

This study identifies the conditions under which alliances can overcome obstacles to coordinating the specialization of their military capabilities to avoid duplication, and second, explains the alliance characteristics that play a role in shaping the composition of defense portfolios. The division of defense capabilities in military alliances provides evidence of how states manage this specialization dilemma. Alliance politics should be understood as a bargaining process in which members negotiate over the distribution of defense responsibilities. Rather than simply aggregating capabilities, alliances vary in how they divide military capabilities among themselves.

R1 Minor comment #1: As for controls, I was curious as to alliance design indicators, like Leeds and Anac's peacetime cooperation measure of Alley's measure of treaty depth. These seem especially important for alliance-year data.

Author comment:

Those variables were initially not included because of concern they were somewhat similar to the hierarchy independent variable. After returning to the original texts, I am less concerned about that and consequently include the Leeds and Anac measure as a control in the main text. The Alley measure, which is derived from the Leeds and Anac measure, is included in the appendix as a robustness check.

There are no substantial changes to the model results based on the new specifications.

Revised text manuscript (pg 26-27):

For each model specification, I first include only the explanatory variables to avoid post-treatment bias and over-adjustment, especially given the panel data has unspecified temporal dynamics (Montgomery, Nyhan, and Torres 2018; Clarke, Kenkel, and Rueda 2018; Dworschak 2023). I add control variables existing research suggests could influence military portfolio similarity or security cooperation.^(fn 25) Many variables were compiled using peacesciencer (Miller 2022). To account for other features of alliance design, the models control for peacetime military coordination indicating the presence of peacetime integrated military command and a common defense policy (Leeds and Anac 2005). Democratic allies may differ in trust and expectations of cooperation so, I control for the proportion of alliance members with a Polity score greater or equal to 6 (Chiba, Johnson, and Leeds 2015; DiGiuseppe and Poast 2016; Warren 2016; Fjelstul and Reiter 2019). Because geographic proximity may better enable allied states to cooperate (Bak 2018) and increases the chance two states fight alongside each other (Gartzke and Gleditsch 2022), the models control for the proportion of alliances members that are contiguous and the maximum capital to capital distance across members (Fordham and Poast 2016). As discussed in Section 3.2, more diverse threat environments make it harder to coordinate a division of labor, so the models account for the logged number of total threats faced by all alliance members. A control for the logged number of states in each alliance year accounts for the fact that more partners increases the potential for capability overlap (Chiba, Johnson, and Leeds 2015) and because the benefits of hierarchy decline with group size (Beek et al. 2024). Established alliances with deeper roots have an easier time cooperating, so I include the average number of years each member has been in the alliance (Palmer 1990a; B. V. Benson and Clinton 2016).

Table 1: Coefficient estimates for regression models.

	Year Polynomials		Year FE		Multilevel Model	
	(1)	(2)	(3)	(4)	(5)	(6)
Strategic Compatibility	0.186*** (0.046)	0.153* (0.061)	0.192*** (0.048)	0.165* (0.064)	0.197*** (0.019)	0.168*** (0.024)
Hierarchy	0.403*** (0.067)	0.320** (0.119)	0.397*** (0.067)	0.313* (0.121)	0.396*** (0.021)	0.327*** (0.039)
Alliance Institutionalization		0.039* (0.017)		0.039* (0.017)		0.035*** (0.006)
Democracy Ratio		0.118* (0.046)		0.116* (0.047)		0.116*** (0.017)
Contiguity Ratio		-0.127 (0.106)		-0.140 (0.106)		-0.142** (0.045)
Maximum Distance (log)		-0.025 (0.015)		-0.026 (0.015)		-0.026*** (0.006)
Number of Rivals (log)		0.064* (0.027)		0.060* (0.027)		0.066*** (0.009)
Number of Members (log)		-0.029 (0.019)		-0.027 (0.019)		-0.031*** (0.008)
Alliance Age (avg)		0.000 (0.001)		0.000 (0.001)		0.000 (0.000)
Num.Obs.	2398	2280	2398	2280	2398	2280
R2	0.206	0.284	0.232	0.307		
R2 Adj.	0.204	0.281	0.217	0.290		
R2 Marg.					0.127	0.201
R2 Cond.					0.217	0.289
AIC	-226.4	-482.2	-226.2	-472.2	-158.4	-341.9
BIC	-191.8	-407.7	45.5	-162.7	-129.4	-273.1
RMSE	0.23	0.22	0.23	0.21	0.23	0.21

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

R1 Minor comment #2: For a bounded outcome variable, a robustness check with ordered beta regression (Kubinec PA), would be helpful.

Author comment:

The results of an ordered beta regression are provided in Appendix Section 2.1 (Table A2). That model, its justification, and the suggested citation are in the main text as the substantive interpretation with a conditional effects plot. Since the coefficients of an ordered beta regression cannot be compared to those of the OLS results, I clarify that point rather than add the ordered beta regression to Table 1 in the main text

Revised text manuscript (pg 29):

The results are also substantively significant. Figure 4 illustrates the conditional effects of the two explanatory variables using an ordered beta regression, which appropriately accounts for the bounded nature of the dependent variable (Kubinec 2022).^(fn 22: Ordered beta regression results are provided in the appendix, as the log odds scale coefficients are not comparable to the other models.)

2 Manuscript model supplementary info

2.1 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	2.908 [2.379, 3.536]
Hierarchy	6.718 [4.654, 9.716]
Democracy Ratio	1.264 [1.083, 1.486]
Contiguity Ratio	2.550 [1.666, 3.784]
Maximum Distance (log)	1.105 [1.050, 1.159]
Number of Members (log)	0.763 [0.703, 0.829]
Alliance Age (avg)	0.998 [0.994, 1.003]
Num.Obs.	2206
RMSE	0.23

R1 Minor comment #3: A quick description of the multilevel model in appendix section 5.2 would be helpful.

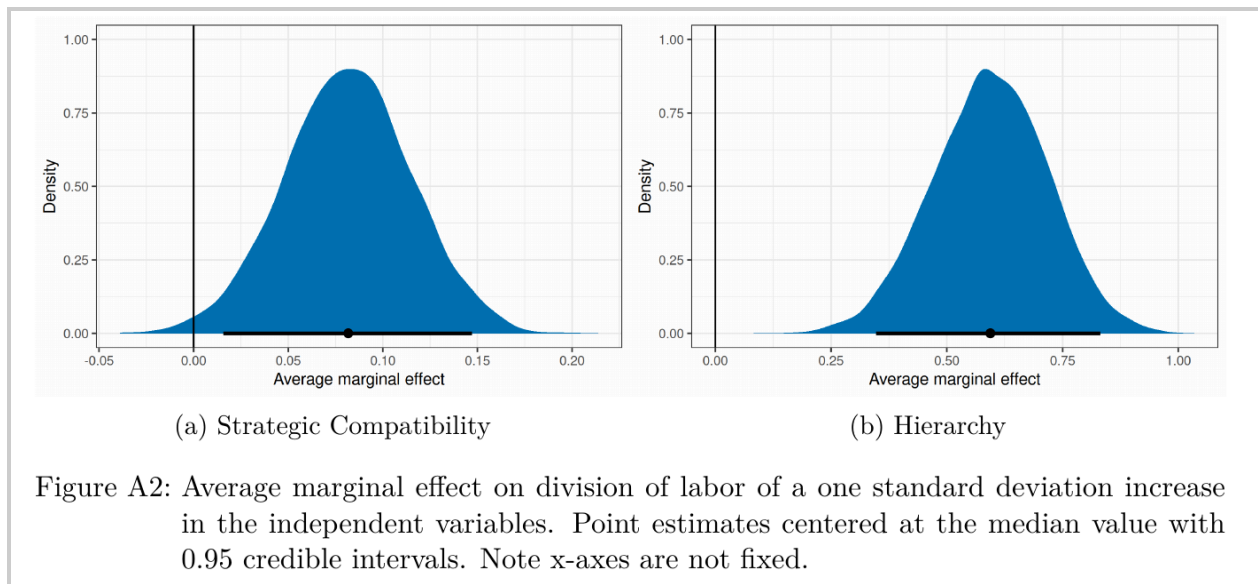
Author comment:

A description of the multilevel model in the appendix is now provided, including its purpose and interpretation. As an aside, Reviewer 1's suggestion to add the control for alliance peacetime coordination changes the results of the multilevel model to be more consistent with the theory. Initially, the credibility interval for strategic compatibility was positive at the 0.1 level and it is now positive at the 0.05 level.

Revised text manuscript (appendix pg 19):

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.



Reviewer 2

R2 major comment #1 (Information asymmetry mechanism): My first comment is regarding the theoretical argument. In arguing for the benefits of hierarchy, the author emphasizes how a leading "alliance manager" can solve the information asymmetry between allies regarding who is contributing what. I am not quite convinced as to how hierarchy can solve information asymmetry. To make this point clear, the author may want to describe what kind of information is held privately by which actor. If this refers to "uncertainty about future tasks" (page 14), then it seems there is no actor who holds better information about this or has any incentive to withhold such information to gain more benefit. When the author talks about asymmetric information, I was thinking more about actors' willingness to rely on their allies for security (= their willingness to make policy concessions in exchange for security benefits). A protégé may have an incentive to appear "self-sufficient" in the eyes of the defender state in hopes of negotiating a lower price for a given level of protection. Likewise, the defender state could have private information about its willingness to protect the protégé (by the way, this would be highly correlated with strategic compatibility) without a policy concession from the protégé and has an incentive to make it look lower to extract more concessions from the protégé. If the author has either of these types of information asymmetry in mind, I don't see how hierarchy can solve this problem. If, alternatively, the author has a different type of information in mind, please make it more explicit. Please also make it clear why and how hierarchy can solve such information asymmetry.

Author comment:

The language of that section has been rewritten to focus more on the leading alliance manager solving coordination problems that result from incomplete information rather than information asymmetry itself. By framing it as a coordination problem, the role of a central actor is more clear and the theory is able to make a weaker assumption about the intentional or unintentional withholding of information by bargaining actors which creates a more intuitive mechanism by which hierarchy helps

Revised text manuscript (pg 13-15):

Hierarchy reduces opportunism and coordination costs by structuring task division (Pondy 1977; Oye 1985) and creating mutual interdependence (Gulati, Wohlgezogen, and Zhelyazkov 2012, 533). Task division represents a coordination problem because even if they both prefer cooperation to non-cooperation, actors may disagree on the form cooperation should take and how to share its surplus gains. In a simplified context where we consider a division over the military domains of land, air, and sea, two actors may agree that they should each retain whatever ground forces are needed to defend their own territory, but they may disagree about who should provide air and naval forces that are involved in collective defense. Both actors may prefer to provide the naval forces due to domestic political considerations or expected production costs, but as in a coordination game their strongest preference is to coordinate their actions and complement the other party's decision.

Hierarchy means there is someone who is expected to be the leader (Galbraith 1977; Morrow 1994a, 408), and that leader can solve the coordination problem benevolently by just shouldering the burden of creating an ordering regime knowing that all alliance members will

benefit (Kindleberger 1986, 288–305) or it could do so in a self-interested way because hierarchy allows that actor to establish the regime that it prefers (Krasner 1976). Because such coordination involves high transaction costs, weaker states in particular benefit from joining a hierarchical as opposed to non-hierarchical alliance regime because another actor shoulders the coordination costs that a weaker state is unable to overcome even if doing so reaps the benefits of coordinated defense efforts (Weber 1997, 329; Gulati and Singh 1998, 304).

The dominant state delegates nodes of responsibility to the subordinate state either because those tasks are less important niche capabilities or because the subordinate state can perform those tasks at a lower cost given comparative advantage offered by geography or industrial capacity (Sugiyama and Sugawara 2017). By transferring a purely exchange relationship into a power relationship, hierarchy ensures unified command and standard operating procedures that create the type of task coordination needed for a division of responsibility (March and Simon 1958; Galbraith 1977; Gulati and Singh 1998; Koremenos, Lipson, and Snidal 2001). Under either mechanism, hierarchy serves as a regime that makes coordination easier by centralizing the flow of authoritative communication which ensures that “information can spread among the entire group via a minimum number of connections” (Zafeiris and Vicsek 2018, 55) and by minimizing information failures through “strong and clear rules government behavior...between networks of states linked under a common leader” (Lake 2009, 101).^{(fn 14 “Authoritative communication” is consistent with experimental evidence suggesting stag’s hunt coordination games more often result in efficient equilibrium under one-way than two-way communication regardless of payoff symmetry (Agranov 2024).)}

R2 comment #2 (selection bias): My second comment is regarding a potential selection bias. The two explanatory factors, strategic compatibility and hierarchy, could potentially influence the likelihood that states form an alliance in the first place. For example, if we think that higher strategic compatibility encourages potential allies to form an alliance in the first place (i.e., they form an alliance because they find it easier to implement division of labor), then this could potentially attenuate the impact of compatibility. If so, by focusing only on observed alliances, the author underestimates the effect of compatibility on division of labor. However, if we think that alliances are needed when "raw" strategic compatibility (i.e., compatibility in the absence of an alliance) is low, then strategic compatibility may be negatively correlated with alliance formation. If so, focusing only on observed alliances may overestimate the effect of compatibility on division of labor. A similar argument could be made about hierarchy and alliance formation. The author might not be able to "solve" this problem completely empirically, but I'd encourage the author to address it at least theoretically.

Author comment:

A new paragraph has been added to the theory section acknowledging the selection bias in the endogenous relationship between alliance formation itself and the independent variables. As suggested by the editor I have added a reference to Poast (2019), which deals with this issue explicitly, and identifies ways it can be thought about theoretically in biasing the results in either direction

While it cannot be fully addressed empirically, the editor correctly identifies that the double/debiased machine learning model in the appendix is designed to deal with precisely this kind of selection bias. That was not made clear in the original manuscript, so a brief note on that has been added to the main text and a more thorough description of that model provided in the appendix.

Revised text manuscript (pg 16):

Importantly, both strategic compatibility and hierarchy could influence alliance formation itself, potentially impacting the observed relationship of these factors on division of labor (Poast 2019a, 67–70). Like theories skeptical about compliance with international institutions (Mearsheimer 1994; Downs, Rocke, and Barsoom 1996), alliances may form among states with high strategic compatibility and when hierarchical structures are accepted because they find it easier to coordinate when they already have much to gain, thus attenuating the observed relationship between strategic compatibility and hierarchy on the division of labor. Alternatively, alliances may form in cases of low compatibility or the absence of hierarchy because those conditions make informal ad-hoc cooperation uncertain or abandonment likely, which could overestimate the relationship of these variables with a division of labor (Snyder 1984; Gannon and Kent 2021). While difficult to entangle both theoretically and with observational panel data, these dynamics suggest that alliance formation and internal cooperation are intertwined, which future research could explore further.

Revised text manuscript (pg 30):

Despite the robustness of the results across a wide variety of model specifications, the limitations of quantitative models using observational panel data warrant caution in interpreting this as evidence of a causal relationship. There is a plausibly endogenous relationship that the models cannot account for – a state’s relationship with other states influences the capabilities each state produces but the capabilities each state has at their disposal also impacts the decision to ally with another state (DiGiuseppe and Poast 2016). Furthermore, threat environments could be endogenous to alliance formation if alliances create threats rather than protect against them. And, as discussed in Section 3.3, strategic compatibility and hierarchy could influence selection into alliances itself.

Some of these endogeneity concerns are addressed with the double/de-biased machine learning model outlined in the appendix which mitigates bias through post-selection inference both into the sample and by separating the independent variables from the controls in a two-stage residual model, though challenges related to unobserved non-alliances remain. But even with that, these problems cannot be fully overcome with the data and methods used here. However, the goal of this study is to identify a robust empirical pattern consistent with the novel conception of alliances I have outlined with the hope that the theory and results presented here provide a justification for more causally-oriented studies. Future research could explore this issue further by, for example, examining the conditions under which alliances form and how these conditions influence the division of labor within alliances.

Revised text manuscript (Appendix pg 19):

5.3 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 shows 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	

R2 comment #3 (typo): My third comment is rather minor. On pages 4-5, the author describes the manuscript's three contributions using First, ..., Second, ..., and Third, ... Between "Second" and "Third", I believe the author mistakenly inserted a roadmap of the paper, writing "The argument is organized as follows. ..." I think this section should come after "Third, ...".

Author comment:

The typo has been fixed and the third contribution is now properly placed before the roadmap.

Revised text manuscript (pg 3-4):

This study advances our understanding of several key issues in international politics. First, it clarifies how alliances and armaments interact (Morrow 1993; Diehl 1994), showing that armament distribution within alliances is shaped by compatible security interests and centralized decision-making. The efficacy of military spending can only be understood in the context of what that money is spent on, and this is particularly true of allies whose spending could be high but superfluous or low but effective. Second, it addresses the debate over the efficacy of alliances in improving security and stability (Kimball 2010; Blankenship 2023), offering a novel mechanism – complementarity – by which allies can pool distinctly specialized capabilities to improve the efficiency of defense spending (Gannon 2024). Third, it challenges the foundational view that states behave as like-units due to the ability to use force and the absence of central authority (Stein 1982; Posen 1984), demonstrating instead that functional differentiation occurs when aligned states adopt different roles and responsibilities through cooperation (Bukovansky et al. 2012, 6–7).

The argument is organized as follows. Section 2 reviews existing explanations for how states improve their security through arming and allying. Section 3 details the theory of a shared production model of security, explaining how strategic compatibility and hierarchy can foster effective intra-alliance bargaining that promotes a division of labor in defense capabilities. Section 4 presents a quantitative empirical test of this theory using cross-national data on disaggregated military capabilities from 1970 to 2014. Section 5 concludes with a discussion of the implications of these findings for theories about intra-alliance bargaining and conflict and offers suggestions for future research.

Reviewer 3

R3 comment #1: Research question. In the introduction the author proposes various research questions. I encourage the author to centralize the research question on page 7 - "What factors shape whether an alliance can reliably engage in collective action and how does that change the strategic composition of arms across allied states?" This question best reflects to two prongs of the paper and how they interact - Under what conditions do alliance members engage in a division of labor vis-à-vis member capabilities? And, related, how does the nature of the alliance affect member defense portfolios?

Author comment:

The introduction has been rewritten and now focuses around the two questions suggested by the reviewer. I agree with the reviewer's assessment that that is the core puzzle in this paper

Revised text manuscript (pg 2):

A fundamental challenge for military alliances is distributing defense capabilities efficiently among member states. Specializing in capabilities reduces the opportunity costs of defense spending, termed a "regrettable necessity" by Nordhaus and Tobin (1973, 515–17).^(fn 1) On the societal costs of defense spending and international security, see also Powell (1993) and Fearon (2018).) However, specialization risks arise if allies fail to fulfill their commitments, leaving one vulnerable (Snyder 1984; Palmer 1990b) and arming with allies in mind may also conflict with domestic political priorities (Kurth 1973; Fordham 2019). This creates a dilemma: specialization yields economic and operational efficiencies, but requires deciding each partner's role and managing coordination costs and risks (Niou and Ordeshook 1994; Papayoanou 1997). What factors shape whether an alliance can reliably engage in collective defense, and how does that change the strategic composition of arms across allied states?

This study identifies the conditions under which alliances can overcome obstacles to coordinating the specialization of their military capabilities to avoid duplication, and second, explains the alliance characteristics that play a role in shaping the composition of defense portfolios. The division of defense capabilities in military alliances provides evidence of how states manage this specialization dilemma. Alliance politics should be understood as a bargaining process in which members negotiate over the distribution of defense responsibilities. Rather than simply aggregating capabilities, alliances vary in how they divide military capabilities among themselves.

R3 comment #2: Theory. I appreciate the author's discussion of strategic compatibility, but I found that it appeared to conflate the nature of threat and the level of threat. Alliance member A and B may both find State Y threatening but it could also be that alliance member A and B do not find it very threatening, or not enough to coordinate defense portfolios. Thus, one outcome could be, high agreement on nature of threat (strategic compatibility) but low degree of shared threat salience means lower division of labor alternatively, high agreement on nature of threat and high agreement on threat salience equals high division of labor. I would also like to see the author better address the diversity of the threat environment. They do so suggestively when providing examples of how they created an index for strategic compatibility but not theoretically. A more singular threat environment (say all alliance members only finding State Y threatening) would probably yield the highest division of labor (following the logic the author laid out). Finally, regardless of nature of threat and level of threat (say those are both agreed upon) there is also must be agreement in how to meet the threat (say policy compatibility).

Author comment:

The first paragraph of the theory section on strategic compatibility (section 3.2) has been rewritten to be more clear about the two dimensions of threat the reviewer suggested. The point about the diversity of the threat environment is especially well taken, as singular versus distributed threat environments should matter significantly for how states collectively arm against those threats. As policy compatibility is one way of thinking about the dependent variable, I include a footnote clarifying that in relation to Poast 2019 who uses similar language and from whom this variable is heavily inspired

In addition to clarifying that theoretical point, I constructed a new variable identifying the total number of threats faced by an alliance and re-ran the models with that variable. All models in the main text and the appendix now include that variable, and the results for the independent variables of interest remain consistent with theoretical expectations. Although I am hesitant to interpret the coefficient for the new control variable given the reasons described in the paper, it is positive, statistically significant in all models, and makes the “alliance size” variable insignificant in 2 of the 3 models, which poses an interesting question for future research about the relationship between the warring coalition of different sizes. It could be that as the number of threats increases, the incentive to “divide and conquer” increases.

Revised text manuscript (pg 11-12):

I define strategic compatibility as the consistency of states' security interests and degree to which they agree on the nature of the possible international threat environment (Poast 2019a, 5). This involves two dimensions: consistency about who constitutes a security threat and the salience of those shared threats.(fn 7 Poast (2019a) separates ideal war plan compatibility into strategic and operational compatibility. My definition is consistent with how he thinks about strategic compatibility, while operational compatibility – defined as similarity in military doctrine concerning agreement on how to meet the threat – is most similar to my dependent variable. One reading of my argument is that successful operational compatibility – agreement about how to jointly meet a threat – is a function of strategic compatibility.) When security interests among states are consistent, an adversary that poses a threat to one state's security

interests also poses a threat to the other states' security interests (Yarhi-Milo, Lanoszka, and Cooper 2016; Markowitz and Fariss 2018). Saliency of those shared threats means that the threat is a high priority for the states in question due to the military power of the adversary. During the Cold War, for example, not all Warsaw Pact states got equal military attention from NATO strategic planners. Bulgaria, in particular, was noted as militarily weak and thus little was done to coordinate Western defense portfolios regarding that specific threat (Nelson 1989). Naturally, the more diverse the threat environment the harder it will be to achieve an efficient division of labor. In a singular threat environment, where all alliance members focus on a single adversary, a division of labor is easier to achieve. By contrast, a diverse threat environment complicates coordination, as members must balance multiple security concerns. Even where there is strong alignment on both the nature and saliency of the threats, more possible scenarios for future conflict opponents makes complementarity harder to achieve.

Revised text manuscript (pg 25-26):

For each model specification, I first include only the explanatory variables to avoid post-treatment bias and over-adjustment, especially given the panel data has unspecified temporal dynamics (Montgomery, Nyhan, and Torres 2018; Clarke, Kenkel, and Rueda 2018; Dworschak 2023). I add control variables existing research suggests could influence military portfolio similarity or security cooperation.²⁵ To account for other features of alliance design, the models control for peacetime military coordination indicating the presence of peacetime integrated military command and a common defense policy (Leeds and Anac 2005). Democratic allies may differ in trust and expectations of cooperation so, I control for the proportion of alliance members with a Polity score greater or equal to 6 (Chiba, Johnson, and Leeds 2015; DiGiuseppe and Poast 2016; Warren 2016; Fjelstul and Reiter 2019). Because geographic proximity may better enable allied states to cooperate (Bak 2018) and increases the chance two states fight alongside each other (Gartzke and Gleditsch 2022), the models control for the proportion of alliances members that are contiguous and the maximum capital to capital distance across members (Fordham and Poast 2016). As discussed in Section 3.2, more diverse threat environments make it harder to coordinate a division of labor, so the models account for the logged number of total threats faced by all alliance members. A control for the logged number of states in each alliance year accounts for the fact that more partners increases the potential for capability overlap (Chiba, Johnson, and Leeds 2015) and because the benefits of hierarchy decline with group size (Beek et al. 2024). Established alliances with deeper roots have an easier time cooperating, so I include the average number of years each member has been in the alliance (Palmer 1990a; B. V. Benson and Clinton 2016).

Table 1: Coefficient estimates for regression models.

	Year Polynomials		Year FE		Multilevel Model	
	(1)	(2)	(3)	(4)	(5)	(6)
Strategic Compatibility	0.186*** (0.046)	0.153* (0.061)	0.192*** (0.048)	0.165* (0.064)	0.197*** (0.019)	0.168*** (0.024)
Hierarchy	0.403*** (0.067)	0.320** (0.119)	0.397*** (0.067)	0.313* (0.121)	0.396*** (0.021)	0.327*** (0.039)
Alliance Institutionalization		0.039* (0.017)		0.039* (0.017)		0.035*** (0.006)
Democracy Ratio		0.118* (0.046)		0.116* (0.047)		0.116*** (0.017)
Contiguity Ratio		-0.127 (0.106)		-0.140 (0.106)		-0.142** (0.045)
Maximum Distance (log)		-0.025 (0.015)		-0.026 (0.015)		-0.026*** (0.006)
Number of Rivals (log)		0.064* (0.027)		0.060* (0.027)		0.066*** (0.009)
Number of Members (log)		-0.029 (0.019)		-0.027 (0.019)		-0.031*** (0.008)
Alliance Age (avg)		0.000 (0.001)		0.000 (0.001)		0.000 (0.000)
Num.Obs.	2398	2280	2398	2280	2398	2280
R2	0.206	0.284	0.232	0.307		
R2 Adj.	0.204	0.281	0.217	0.290		
R2 Marg.					0.127	0.201
R2 Cond.					0.217	0.289
AIC	-226.4	-482.2	-226.2	-472.2	-158.4	-341.9
BIC	-191.8	-407.7	45.5	-162.7	-129.4	-273.1
RMSE	0.23	0.22	0.23	0.21	0.23	0.21

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

R3 comment #3: Alternative arguments and alternative framing suggestion. Defense portfolios - capabilities - are highly path dependent and subject to domestic pressures (defense industry, parochial interests, electoral politics to name a few). None of this is acknowledged in the piece. It may be that regardless of strategic compatibility or hierarchy that states do not coordinate their defense portfolios (division of labor) in such idealized ways. An alternative way to frame the piece (versus the standard collective action literature) is to say, there are all these variables working against defense portfolio coordination / division of labor ranging from domestic to the collective, yet some alliances can overcome these barriers and coordinate. When and how does this happen? Here it would be even more compelling if you can provide some descriptive statistics from your data and say which alliances during specific periods have the highest levels of division of labor and which have the lowest.

Author comment:

Domestic politics are now discussed briefly in the introduction (see previous comment) and in more detail in the literature review. The change to the framing in the introduction now makes more clear that the contribution of the theory is explaining how, despite domestic and international barriers to doing so, there are conditions under which states manage to coordinate defense portfolios with their allies.

More descriptive statistics have been added for the dependent variable, including changing figure 1 to include actual numeric values for two of the examples now provided in the text. I feel this is a nice (and necessary) complement to the descriptive figures that were already provided for the two independent variables

Revised text manuscript (pg 2):

A fundamental challenge for military alliances is distributing defense capabilities efficiently among member states. Specializing in capabilities reduces the opportunity costs of defense spending, termed a “regrettable necessity” by Nordhaus and Tobin (1973, 515–17).^(fn 1) On the societal costs of defense spending and international security, see also Powell (1993) and Fearon (2018). However, specialization risks arise if allies fail to fulfill their commitments, leaving one vulnerable (Snyder 1984; Palmer 1990b) and arming with allies in mind may also conflict with domestic political priorities (Kurth 1973; Fordham 2019). This creates a dilemma: specialization yields economic and operational efficiencies, but requires deciding each partner’s role and managing coordination costs and risks (Niou and Ordeshook 1994; Papayouanou 1997). What factors shape whether an alliance can reliably engage in collective defense, and how does that change the strategic composition of arms across allied states?

Revised text manuscript (pg 5-6):

This contrasts with international trade models, where specialization yields efficient and interdependent divisions of labor. Those pessimistic about the efficacy of alliances make a more common assumption explicit; because the high politics of international security concern survival while the low politics of economics do not, true dependence on others for security is prohibitively risky and states avoid it whenever possible. When alliances do form, states join forces as “temporary marriages of convenience” (Mearsheimer 2001, 33) against today’s common enemy, but they do so by aggregating capabilities rather than dividing responsibilities

(Olson and Zeckhauser 1966). Proponents argue this distinguishes domestic and international politics (Waltz 1979, 97–99); while there is “functional specialization” within states, the same is not true in international politics, where power is more evenly distributed and lacks functional specialization (Posen 1984, 36–37). As with economic protectionism, domestic politics can also bias against efficient forms of international security coordination, as defense industries (Kurth 1972), parochial interests (Cappella Zielinski and Schilde 2019), partisan ideologies (Fordham 2002; Whitten and Williams 2011), and bureaucracies (Halperin, Clapp, and Kanter 1974) create path dependence that resists change.

Revised text manuscript (pg 18-19):

The distribution of this index across the entire universe of alliance-years is shown in Figure 1. In 2002, ANZUS had a higher division of labor than average, while the Collective Security Treaty Organization (CSTO) had a lower one. The high division of labor for ANZUS reflects the alliance’s ability to coordinate military capabilities across its members, with the Australia and the United States developing a strong expectation of mutual support and military commitments through Australia’s development of expeditionary forces and naval surveillance capabilities while the United States contributes nuclear extended deterrent capabilities (Von Hlatky 2013, 113–37). Contrastly, the CSTO’s inability to even establish a functioning rapid reaction force is emblematic of a “propensity to defect from cooperation” where each states prefers to “establish separate regional armed groups designated to meet the defense needs of specific members of CSTO” (Davidzon 2022, 172).

