Political Science Research and Methods (2026), 1–17
doi:10.1017/psrm.2026.10132

ORIGINAL ARTICLE

Is alliance credibility interdependent?

Joshua Alley[1] and Brian Blankenship[2]

[1]Department of Political Science, Baylor University, Waco, TX, USA and [2]Department of Political Science, University of Miami, Coral Gables, FL, USA
Corresponding author: Joshua Alley; Email: joshua_alley@baylor.edu

(Received 8 January 2026 UTC; revised 8 April 2026 UTC; accepted 8 May 2026 UTC)

Abstract
How might changes in the credibility of one alliance relate to the credibility of other alliance commitments? Debate between competing views on this question spans decades in part because systematic tests are difficult. We contribute systematic evidence with a new measure of credibility in U.S. alliances and novel technique to estimate interdependence. Our measure draws on expert ratings of allied beliefs over U.S. credibility and our analysis employs Gaussian process regression to analyze credibility. We find strong correlations in allied credibility perceptions within regions and little correlation across regions. Supplementary analyses suggest that these correlations reflect common unobservables and a small impact of shared perception of past U.S. behavior.

Keywords: international relations; international security; measurement; spatial methods; survey methodology

1. Introduction

Credibility is essential to military alliances. Whether partners will honor their commitments determines if an alliance treaty aggregates capability or is a "scrap of paper." As a result, leaders and scholars pay close attention to potential determinants of credibility.

Perhaps the most debated predictor of credibility is the credibility of a state's *other* alliance commitments. Scholars regularly dispute whether alliance credibility is interdependent. Some argue that credibility is actor-specific, whether due to intrinsic characteristics such as domestic political institutions (Gaubatz, 1996; Gartzke and Gleditsch, 2004) or the past behavior of governments and leaders (Schelling, 1966; Harvey and Mitton, 2016; Lupton, 2020). In this view, an actor's perceived credibility in one alliance is likely to be correlated with its perceived credibility in other alliances. Actions toward one partner therefore reverberate across commitments. Others argue that credibility connections across alliances are overstated (Hopf, 1994; Mercer, 1996), or require similar situations (Gibler, 2008; Weisiger and Yarhi-Milo, 2015; Henry, 2022).

These arguments have distinct implications. If lower credibility in one alliance predicts lower credibility elsewhere, then weakening or withdrawing from one commitment might endanger others. Schelling (1966, pg. 124) therefore argued that upholding a strong reputation may be "one of the few things worth fighting over." But if alliance credibility is not particularly interdependent, then states face fewer costs from weakening commitments to one ally. This gap is especially important in contemporary American foreign policy.

© The Author(s), 2026. Published by Cambridge University Press on behalf of EPS Academic Ltd. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psrm.2026.10132

The interdependence debate continues in part because systematic evidence is rare. Broad measures in existing work are poorly suited to address the debate either because they are indirect proxies, poorly suited to test interdependence, or both. Quantitative scholarship generally relies on observable proxies for alliance credibility such as the design of alliance treaties (Leeds and Anac, 2005; Benson, 2012; Alley, 2021) and foreign-deployed military forces (Fuhrmann and Sechser, 2014). Unfortunately, these indicators are not well-suited to assessing credibility interdependence. Not only are they often fixed or slow to change over time, but they are the tools that actors use to manipulate partner's beliefs about their credibility. Using these measures therefore risks conflating credibility with how states produce it. Others use alliance violations to infer credibility (Leeds and Anac, 2005; Crescenzi *et al.*, 2012; Nieman and Gibler, 2023). Alliance treaty violations are rare, however, and inferring credibility from treaty violations is subject to non-random selection.

Other scholars analyze interdependence with methods that provide rich snapshots but may not generalize. Individual case studies (Snyder, 1997; Kim and Simón, 2021) and surveys (Kim *et al.*, 2024) have useful internal validity, but might not apply elsewhere. Our measure and analysis complement case studies and experiments with global results across a long time.

We contribute to interdependent credibility debates by improving on how scholars measure alliance credibility and estimating interdependence between alliances. First, we develop a new measure of alliance credibility and use it to analyze interdependence across U.S. alliances from 1945 to 2022. Our credibility measure draws on expert ratings of alliance commitment that explicitly incorporate subjective perceptions of the likelihood that partners would honor their obligations. To correct for differences in expert perceptions, regularize the measure and aggregate responses, we use an item response theory (IRT) model to estimate U.S. credibility for each alliance-year observation.

We then estimate the interdependence of credibility using spatiotemporal Gaussian process regression.[1] We use this modeling strategy because it captures correlations in time and space under a flexible structure that does not hard-code proximity. Unlike many political science studies that adjust for interdependence to ensure unbiased estimates of some independent variable, interdependence is our quantity of interest, so flexible estimation is critical. Our model situates annual observations of U.S. credibility in time and geographic space and estimates credibility correlations within and between alliances. The model also estimates how spatial correlations between alliance observations decay over time.

We find that perceived credibility is correlated within regions, but weakly correlated across regions. We observe strong correlations in credibility when allies are close together, but weak or absent correlations as distance increases. That is, knowing how an ally in Europe, the Middle East, East Asia, or Oceania perceived American credibility can help predict other allies' beliefs within that region, but is less useful for predicting credibility elsewhere. Moreover, additional analyses in the appendix suggest that very little intra-regional correlation is driven by shared perceptions of U.S. behavior. In cases where the United States terminated or violated an alliance or backed down in a dispute, nearby allies' perceptions of U.S. credibility declined. But these effect sizes are small, and spatial regressions find little autoregression and a clear spatial error process. Spatial error means that factors that are not included among regression predictors, such as shared interests, create correlations between alliances. This implies that much of the within-region correlations stem from otherwise unobserved shared perceptions of U.S. interests and capabilities in a region, not past actions (Press, 2005).

Our findings thus support conditional claims about interdependent credibility and detail the geographic bounds of interdependence for the United States. Credibility is more likely to be correlated among close allies with similar concerns. For example, a decline in perceived credibility in one U.S. ally in East Asia would likely be associated with similar declines among other allies in the region, but would be unlikely to reverberate outside of the region.

[1] See Prati *et al.* (2023) and Carlson and Zeigler (2024) for political science introductions.

https://doi.org/10.1017/psrm.2026.10132

This article makes three contributions. First, we offer a novel measure that directly captures actors' perceptions of partners' credibility. Although the interdependence of commitments is among the most venerable debates in international politics, existing data do not directly measure perceived credibility and use indicators with theoretical and measurement shortcomings. Second, we use a novel methodology that makes detailed estimates of correlations in space and time. Other scholars could use our data to explore other causes and consequences of U.S. credibility or use our modeling strategy to examine similar forms of interdependence. Third, our findings reinforce an emerging body of findings which suggest conditional interdependence of commitments with more general evidence from alliance politics (Weisiger and Yarhi-Milo, 2015; Lupton, 2020; Henry, 2022).

2. The debate over alliance interdependence

We follow Press (2005, 10) in defining credibility as "the perceived likelihood that the threat will be carried out if the conditions that are supposed to trigger it are met." Since we focus on credibility within alliances, we consider the perceived likelihood that a country will intervene in the defense of its allies if they are attacked. Like other aspects of reputation, credibility is therefore in the eye of the beholder. We focus on beliefs about intervention because they are the essence of alliances that promise active military support.

In turn, we define interdependence as the degree to which the perceived credibility of a state's commitments to different alliances are correlated. This definition is agnostic as to the sources of interdependence, which vary.[2] We equate interdependence with correlations as a broad initial cut, and turn to mechanisms with other tests.

There are several reasons why credibility might be interdependent. The simplest is that allies might base credibility assessments on aggregate partner attributes. Allies might consider partners' internal characteristics, such as domestic politics and regime type (Gartzke and Gleditsch, 2004; Leeds *et al.*, 2009), or situational characteristics such as relative power.

Alternatively, a large body of scholarship argues that credibility is interdependent because state behavior reveals information about interests and willingness to fight. In this telling, defending one ally will lead other allies to view the partner's commitment as credible, while failing to support one's partner or backing down in a crisis will negatively affect observers' credibility perceptions (Crescenzi *et al.*, 2012; Harvey and Mitton, 2016; Dafoe *et al.*, 2021). Schelling (1966, pg. 55) argued explicitly that the U.S. made many Cold War commitments and had to stick to them because "our threats are interdependent." The strongest version of this claim has few bounds, especially in policy and journalism. For example, some observers argued that U.S. withdrawal from Afghanistan would undermine U.S. credibility around the world (Schmitt and Cooper, 2021).

Other scholars advance a more modest claim: that credibility is conditionally interdependent. Allies may be more likely to draw inferences from how their partner behaves toward allies in similar situations. This similarity may stem from a variety of sources, including the stakes involved, the identity and strength of the adversary, and geography (Snyder, 1991; Gibler, 2008; Weisiger and Yarhi-Milo, 2015; Lupton, 2020). Henry (2022) likewise argues that allies look for evidence that their partner shares their interests, rather than unconditional displays of loyalty to other allies.

A related body of research suggests that security interests vary systematically by region. For example, Buzan and Wæver's (2003) "regional security complex theory" suggests that "security interdependence is normally patterned into regionally based clusters" (p. 4) because geographic proximity leads to repeated interactions and relational patterns. Danilovic (2002) argues that the inherent credibility of great powers' deterrent threats varies by region due to distinct regional interests. A state's capability to project power may vary by region as well, whether due to the composition and location of its military forces or adversaries' capabilities (Bak, 2018).

[2]The definition is thus slightly broader than that recently offered by Henry (2020, 45): "that what happens in one alliance affects the expectations of other allies."

https://doi.org/10.1017/psrm.2026.10132

Table 1. Summary of arguments about interdependence between alliance commitments

Argument	Correlated alliances	Across regions?	Causal?
Strong interdependence	Many	Yes	Yes
Conditional interdependence	Some	Weak	Varies
No interdependence	Few or none	No	No

Similarly, perceptions of credibility may be correlated within regions because allies with similar regimes and cultures cluster and shared democracy and cultural similarity predict alliance reliability (Leeds, 2003; Lee *et al.*, 2025). This may be particularly relevant for the perceived credibility of a democratic patron like the United States, given that democracies are likely to defend their democratic allies, especially given high stakes (Clare, 2013). One might thus expect credibility to be highly interdependent within regions, but not necessarily across regions, owing to shared environments, shared allied characteristics and common overarching interests in each region.

A third strand of scholarship advances multiple reasons why credibility might not be interdependent at all. Situations may be too different to draw inferences about a partner's inherent credibility from its behavior toward another actor (Hopf, 1994; Schmidt, 2025). In this telling, observers are instead likely to rely on their assessments of a state's interests in whatever is at stake, rather than its past behavior in other circumstances (Press, 2005). Moreover, a partner's efforts to defend one ally may even undermine other allies' perceptions of that partner's credibility, insofar as those efforts expend resources that could defend them instead (Krebs and Spindel, 2018; Kim and Simón, 2021; Kim *et al.*, 2024). Finally, Mercer (1996) argues that psychological biases from in-group/out-group dynamics mean allies will rarely see a partner's displays of resolve as evidence of its intrinsic resolve, while adversaries will rarely see displays of irresolution as evidence of intrinsic irresolution. Rather, those observers will explain these events in terms of the situation.

These arguments make distinct predictions about the prevalence of correlations in alliance credibility, the distance at which credibility is correlated, and whether changes in the credibility of one alliance cause changes elsewhere. The strongest interdependence arguments suggest that most of a state's alliance commitments should have correlated credibility, even across substantial geographic and political distances. Conditional interdependence arguments predict that some alliances with similar characteristics are correlated, especially when they are closer. No interdependence arguments claim that few alliances have correlated credibility, and credibility will be uncorrelated across and even within regions. Table 1 summarizes these implications.

Debate between these perspectives continues in part because systematic evidence is lacking. There is a variety of rich case-study and survey research on these questions, but generalizing from such evidence is challenging. Furthermore, many existing general measures of credibility are ill-suited for addressing interdependence.

3. Measuring credibility

Scholars have measured alliance credibility in six ways. Each measure has uses, but also pertinent shortcomings for assessing whether credibility is interdependent.

Some early work used treaty duration as a proxy for credibility and commitment (Gaubatz, 1996; Reed, 1997). Alliance treaties can remain active but lack credibility, however. Some scholars argue that the Rio Pact is hardly an alliance at all (Rapp-Hooper, 2020), despite being an active defense treaty according to the ATOP database (Leeds *et al.*, 2002).

Due to these concerns, later work used alliance treaty violation to infer credibility (Leeds and Anac, 2005; Nieman and Gibler, 2023). Violation demonstrates ex post that a commitment was not credible in a specific instance, but using it to measure credibility beforehand presents significant difficulties. The most important concern is that, like audience costs (Schultz, 2001), alliance violations are subject

https://doi.org/10.1017/psrm.2026.10132

to partial observability and strategic selection. States often design alliance obligations to minimize the risk of violation (Benson, 2012; Chiba *et al.*, 2015), and opportunities to violate a treaty in response to a challenge may only arise when credibility is weak (Smith, 1996). Furthermore, treaty violation is rare. Some statistical models attempt to correct for these issues (Nieman and Gibler, 2023), but selection models often employ instrumental variable specifications that are vulnerable to commonplace violations of the exclusion restriction (Lal *et al.*, 2024) or model misspecification (Rainey and Jackson, 2018).

The emphasis on treaty violation and comprehensive data from the Alliance Treaty Obligations and Provisions (ATOP) project led scholars to focus on treaty design as a source of credibility. Researchers sought to capture commitment by considering conditions on intervention (Benson, 2012; Chiba *et al.*, 2015), issue linkages (Poast, 2013), and peacetime cooperation (Leeds and Anac, 2005; Alley, 2021). Treaty content is usually static, however, while credibility is dynamic. Moreover, what is written down may not occur in practice.

Recently, scholars have turned to signals of support (Fuhrmann and Sechser, 2014; McManus and Nieman, 2019; Blankenship, 2020). These works argue that alliance partners reveal commitment by either sinking costs or tying their hands. Taking costly actions such as deploying troops or making public statements of reassurance reveals commitment because a leader who was less willing to honor their alliance commitments would not do such things. Signals are closer, but are still an indirect and potentially noisy measure of credibility, given potential misperceptions (Jervis *et al.*, 2021). The signals one state sends may miss their intended audience or have unintended consequences. Measuring credibility with signals themselves may therefore introduce substantial measurement error. Most importantly, the signals states send are not credibility itself, so signal measures risk conflating credibility with its predictors.

Finally, while not focused on alliance credibility in particular, some scholars treat dispute behavior as a proxy for credibility. One might expect actors that are perceived as credible to be more successful in deterring challengers from initiating disputes and more likely to win their disputes (Schultz, 2001; Weisiger and Yarhi-Milo, 2015). Thus, credibility might be interdependent to the extent that a state and its allies face repeated challenges over time and/or simultaneous challenges across space. However, like other behavioral measures, dispute behavior is at best an indirect proxy for credibility. Moreover, it is difficult to say whether dispute behavior is a cause or effect of credibility. States and their allies may be targeted for disputes when challengers expect less resistance, raising questions about whether displays of irresolution reduce perceived credibility or low perceived credibility contributes to more challenges.

Measuring credibility is difficult, and all of these approaches are worthy efforts to proxy for a critical concept. Unfortunately, none measures credibility in a way that matches the concept directly or facilitates tests of interdependence.

4. Our credibility measure: expert ratings

We address measurement challenges by drawing on expert knowledge of how foreign policy elites perceived their alliance partners. Expert surveys have a long history in political science and are particularly useful for measuring concepts like credibility that require deep domain and contextual knowledge. For instance, Braumoeller (2013) used expert surveys to classify the foreign policy orientation and ideology of great powers. In our case, experts are well-suited to capture the nuances of what credibility means to different allies and thus capture beliefs in a more granular and varied way than prior measures.

Our survey asked experts to classify the credibility of defensive alliances in the ATOP data (Leeds *et al.*, 2002). Credibility refers to the extent to which states expected support in war from their alliance partners, or, for wartime coalitions, expectations that partners would continue to fight according to

https://doi.org/10.1017/psrm.2026.10132

their obligations. We focus on defensive alliances because these treaties contain crucial obligations of support in war, unlike neutrality or non-aggression pacts.

We programmed 289 unique surveys for ATOP defense pacts in Shiny apps, and experts received up to three links for alliances on which they had expertise. Supplementary Appendix section 1 includes sample links and screenshots of several surveys. With help from research assistants, we matched a wide range of experts to specific alliances based on their publications and background. Experts either had meaningful background on a specific alliance or on the foreign policy of key members. Our expert pool included diplomatic historians, political scientists with case-specific expertise, and think tank members. To complement the general search and find experts who had written on specific alliances, we also examined the editorial boards of major history journals, book reviews featured in the American Historical Review dating back to 2000, and back issues of both major history journals and major security studies journals dating back to 2000. We likewise distributed a general version of our survey that indexed links to all alliance surveys to members of the American Historical Association, the Society for Historians of American Foreign Relations, the British International Studies Association, and H-Diplo via their discussion boards.[3]

This initial wave of the survey generated 54 total submissions, and 28 submissions on U.S. alliances. Of the experts surveyed on U.S. alliances, 17% responded. Among our U.S. submissions, we received eight for NATO, four for Japan, three for South Korea and Israel, two for the Philippines and Taiwan, and one for all other alliances, including ANZUS, Spain, Pakistan, Thailand, the Rio Pact, and Iran. While experts could rate multiple alliances, only one did in our U.S. sample.

Our criteria for including U.S. alliances in the survey were to include all of those listed as defense pacts in the ATOP data for the duration they were in force. Those countries with defense pacts include four allies–Israel, Iran, New Zealand, and Taiwan–whose defense pacts with the United States were abrogated before the end of our sample period. Of these four, Israel, New Zealand, and Taiwan have continued to have some expectation of wartime support from the United States–New Zealand because it remains under Australia's protection through ANZUS, Taiwan because of the Taiwan Relations Act, and Israel because of U.S. practice. Thus, we extend the timeframe for inclusion into the sample for Israel, New Zealand, and Taiwan. We exclude informal allies that never appear in the ATOP defense pact data (e.g., Saudi Arabia).

Due to the small pool of potential respondents for any alliance, we did not collect detailed demographic data on respondents. Given the cognitive burden of rating an alliance's perceived credibility in each year, we streamlined each survey to minimize attrition. Moreover, even simple data on gender, occupation, and age could compromise respondent anonymity in small expert pools. Email correspondence, however, indicates that respondents came from all over the world, including the United States, allies, and other countries. This reduces potential American bias in the ratings (Colgan, 2019).

Our smaller expert pool is an unavoidable consequence of analyzing alliances. Similar efforts such as the Varieties of Democracy project recommend many raters per case, but start with a much deeper pool of country experts. Many alliances had less than five living experts, the minimum recommended by V-Dem. Given response rates, using this rule would restrict our data to only a few U.S. alliances.

We asked respondents to focus on the prevailing beliefs among the relevant foreign policy elite. There are often differences of opinion among a state's foreign policy elites, but we emphasized the dominant consensus, should one exist. Crucially, this highlights the perceived credibility of the alliances by the actors themselves, rather than perceptions of the experts.

While our survey has responses for many alliances, our analysis of interdependent credibility focuses on U.S. defense pacts. We do this because we have ratings for every U.S. alliance after 1945. Furthermore, much of the policy and scholarly debate around interdependent credibility examines the vast U.S. alliance network (Kim and Simón, 2021; Henry, 2022; Kim *et al.*, 2024).

[3]The general link allowed recipients to search for alliances by region and time period; see: https://alliancecredibility.shinyapps.io/alliance-cred-survey/.

https://doi.org/10.1017/psrm.2026.10132

The aggregation in our credibility measure varies by alliance size and type. For bilateral and trilateral alliances (e.g., ANZUS), we asked respondents to measure each member's perception that the other member or members would honor their obligations of support in war. Multilateral alliances–those with four or more members–were more challenging, because asking experts to rate credibility perceptions for every treaty member would have created enormous cognitive burdens and produced inconsistent coverage within treaties. For multilateral asymmetric alliances which had a clear alliance leader in the form of one state that was much more capable than the others, we asked experts to rate the leader's commitment and whether members could expect support from at least one alliance member. In our U.S. data, this applies to NATO and the Rio Pact. For multilateral alliances without a leader, which are absent in the U.S. data, expert respondents assessed whether at least one alliance partner would provide promised support in war. This accounts for the critical distinction between multilateral and bilateral alliances (Poast, 2010), because while group and single-country perceptions differ, we measure partner-level perceptions in both cases. It is also important because many existing studies of alliance credibility only analyze bilateral treaties (Leeds *et al.*, 2009; Nieman and Gibler, 2023). The downside is that we only ask for aggregate ratings of multilateral alliances because it would not be possible to have comprehensive expert coverage for all of their members.

The key survey measure rates an alliance actor's credibility on a scale from one to four. A score of one indicates "an alliance in name only," where partners had no expectation of support. Two indicates a partner was "unsure" of support. Three marks years when actors were "fairly sure" that their partner(s) would support them in wartime, and four when alliance actors were "very sure." We asked respondents to mark the initial level and then only mark years when that credibility changed, using the four-point scale to ensure that responses were roughly consistent across respondents.[4]

There are at least three conceptual objections to this approach. First, while we ask respondents for the prevailing view in the foreign policy elite of an alliance actor at that time, there may be some retrospective element to the ratings. Respondents' knowledge of what happened in the near or distant future may influence their views of how policymakers viewed the alliance at the time. But while hindsight brings some potential risks, it also allows more information to come to light about the inner workings of the alliance and how policymakers viewed credibility, which experts with deep historical knowledge are ideally suited to evaluate.

Second, we collected ratings of perceived credibility for partners, not adversaries. Credibility applies to both partners and adversaries, however. We did not ask about adversaries because identifying an adversary for every alliance is difficult and adversaries can shift over time. To the extent an adversary's assessment diverges from those of the partners, this limits the data's utility for exploring research questions about adversary behavior. Nevertheless, the focus on ally perceptions makes the data well-suited to assessing alliance interdependence, especially in terms of reassurance.

Third, while our scale is relatively simple, experts will perceive it differently. Individual respondents may place distinct thresholds between the different outcome levels.[5] These commonplace perceptual differences are sometimes called "differential item functioning" (Marquardt, 2020). Differences across alliances, including multilateral and bilateral treaties, further increase this concern.

We therefore address distinct expert perceptions and combine responses onto a common scale, using a hierarchical IRT model. In IRT models, observed data reflect an underlying latent concept, and the model reduces noise and estimates different weights to measure that concept. In our application, expert ratings are the observed data, and high ratings reflect high credibility, just as many correct test answers reflect higher ability. Using an IRT model to measure credibility accounts for expert and alliance-specific differences in expert perceptions. Other popular measures of political ideology

[4]Respondents only provided these numeric data in their responses. Qualitative descriptions of why credibility shifted would have been helpful, but would have added significant time burdens.

[5]Expert ratings can also be biased by respondents' political concerns or context (Little and Meng, 2024; Weidmann, 2024).

https://doi.org/10.1017/psrm.2026.10132

(Clinton *et al.*, 2004) and democracy (Marquardt and Pemstein, 2018) that face similar challenges in harmonizing perceptions of latent concepts also use IRT models to aggregate expert ratings.

In our ordered logistic IRT model, we estimate latent U.S. credibility in each alliance-year observation based on the expert ratings.[6] The credibility estimate in each year depends on three inputs. First, the expert ratings themselves. Second, an expert-specific discrimination parameter lets the difference between levels of our scale vary for each expert submission across all the years and partners they rated. Crucially, there are no regional patterns in these parameters, which means that our findings are not driven by common expert approaches within regions. Third, we adjust for potential shifts in the level or meaning of credibility across different alliances with alliance-specific intercepts. We model dynamics in the model with a Student-t prior on latent credibility, which allows sharp changes over time (Reuning *et al.*, 2019). This approach fits the observed data much better than a dynamic IRT model with a random walk prior.

The hierarchical model has two critical properties. First, it accounts for differences in how experts perceive the scale and shifts in the differences in levels across alliances, including potential distinctions between multilateral and bilateral treaties. Second, the model pools information across alliances by giving key parameters a common distribution, which makes alliances with fewer expert ratings less likely to be outliers. Alliance-year observations with fewer expert ratings shift closer to the overall mean. This regularization, which is visible in comparing the raw ratings data and the model estimates for partners such as Iran, is important because we have uneven depth in our geographic coverage.

Our final measure is the posterior median of the latent credibility parameter in the IRT models, rescaled to range between 0 and 1.[7] The resulting measure of credibility has several useful properties. First, they directly match our definition of credibility. Second, these measures have some face validity in changes over time and how they correlate with other indicators. Crucially, we also find equivalent results with median credibility ratings.

5. Expert ratings data

First, we present the expert ratings of U.S. credibility. In Figure 1, faint lines mark the individual submissions, and the black line summarizes the median rating. We use the median because it is often a more accurate summary than the average when expert coders give different responses (Lindstädt *et al.*, 2020). For alliances with only one submission, the median and submission lines overlap.

These ratings have some face validity. Across several ratings, U.S. credibility toward Japan is high and steady over time. Perceived U.S. credibility in NATO, which is bundled across all members of the multilateral alliance, is generally high during the Cold War. It then drops slightly at the end of the Cold War and plunges during the Trump administration, recovering under Biden. U.S. credibility also falls dramatically in South Korea during the Carter administration, which publicly considered withdrawing American forces. U.S. credibility to Taiwan bottoms out after the U.S. abrogated a formal treaty commitment and recognized the People's Republic of China, and has slowly recovered since. Iranian perceptions of U.S. credibility drop around the fall of the Shah. Pakistani expectations of U.S. support are volatile and highly contingent on geopolitical context.

Due to potential differences in perceptions of the scale across experts, IRT estimates are likely more reliable than the raw data. We present the IRT ratings in Figure 2. The ratings are generally similar in when credibility shifts, but the levels depend somewhat on the expert discrimination parameters. For instance, the average rating for Spain moves toward the median value because there is little data. NATO also has greater uncertainty due to wider expert disagreement.

[6]Credibility corresponds to the ability parameter in an IRT model, as higher ability means higher scores on the credibility scale. See the appendix for a formal presentation and parameter estimates and checks.

[7]We account for measurement uncertainty in the appendix by fitting models at different quantiles of the posterior distribution.

https://doi.org/10.1017/psrm.2026.10132

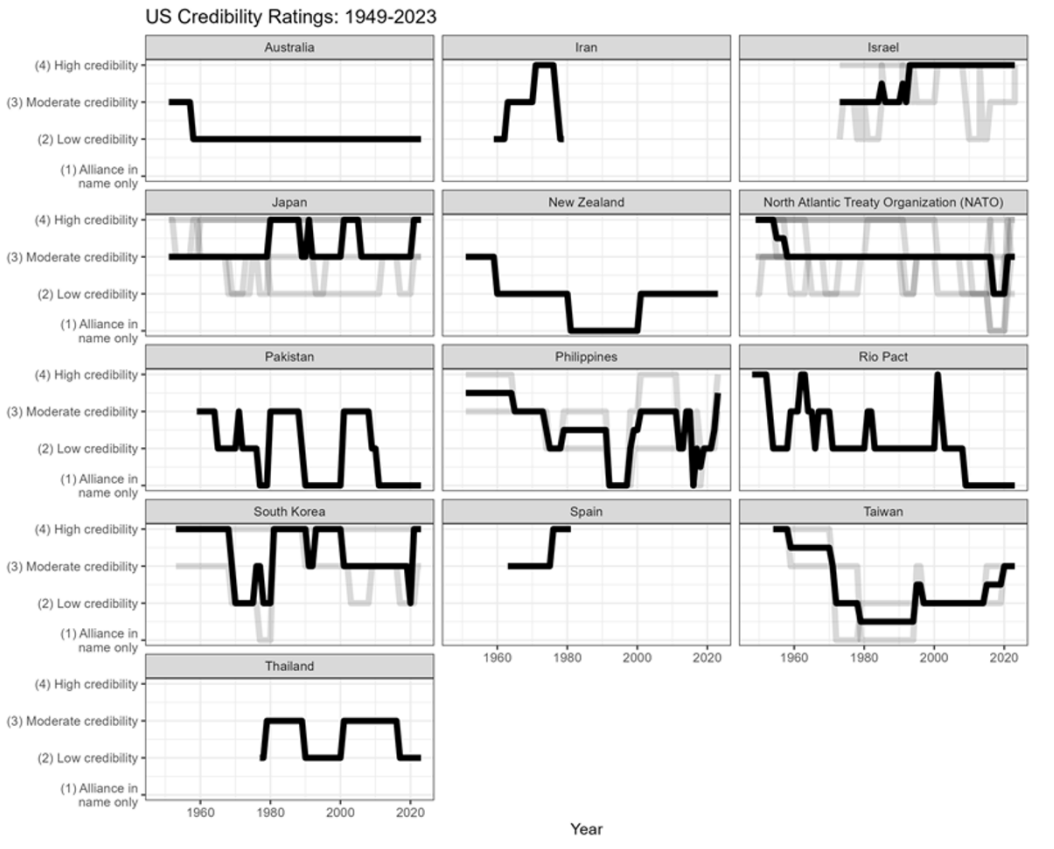


Figure 1. Ratings and median credibility rating for U.S. alliance partners. The faint lines show individual expert ratings, and the black line marks the median.

Our expert ratings also correlate with other measures in sensible ways, as Figure 3 shows. Figure 3 presents scatterplots measuring the correlation between credibility ratings and prior measures. High scores on the latent measure of McManus and Nieman (2019), along with more U.S. troops (Allen *et al.*, 2022), leader visits, statements of support (Blankenship, 2020), and military exercises (D'Orazio, 2013) in a year, are all correlated with greater credibility. UN voting distance is also correlated with credibility, as greater UN distance overlaps with low credibility (Bailey *et al.*, 2017). Presidential visits, statements, and S-scores are less correlated with credibility, however, S-scores are essentially uncorrelated.

Although our measure correlates with existing measures, it adds value in three ways. First, some large shifts in credibility have little to do with the indicators in prior measures, and are often the result of specific events or changes in leadership. U.S. credibility toward New Zealand collapsed due to disagreements over port access for nuclear-armed vessels, for example. Second, all these measures have substantial unexplained variation in credibility. The direction of the correlations is as we should expect, but there is ample variation around the general trend, as observations with similar values on objective indicators can have dramatically different credibility values in our measure. Third, most of these variables outside of S-scores and UN voting distance are factors that might influence credibility, but are not themselves measures of credibility. Moreover, S-scores are poorly correlated with perceived credibility, while UN voting distance is a measure of foreign policy similarity rather than credibility.

https://doi.org/10.1017/psrm.2026.10132

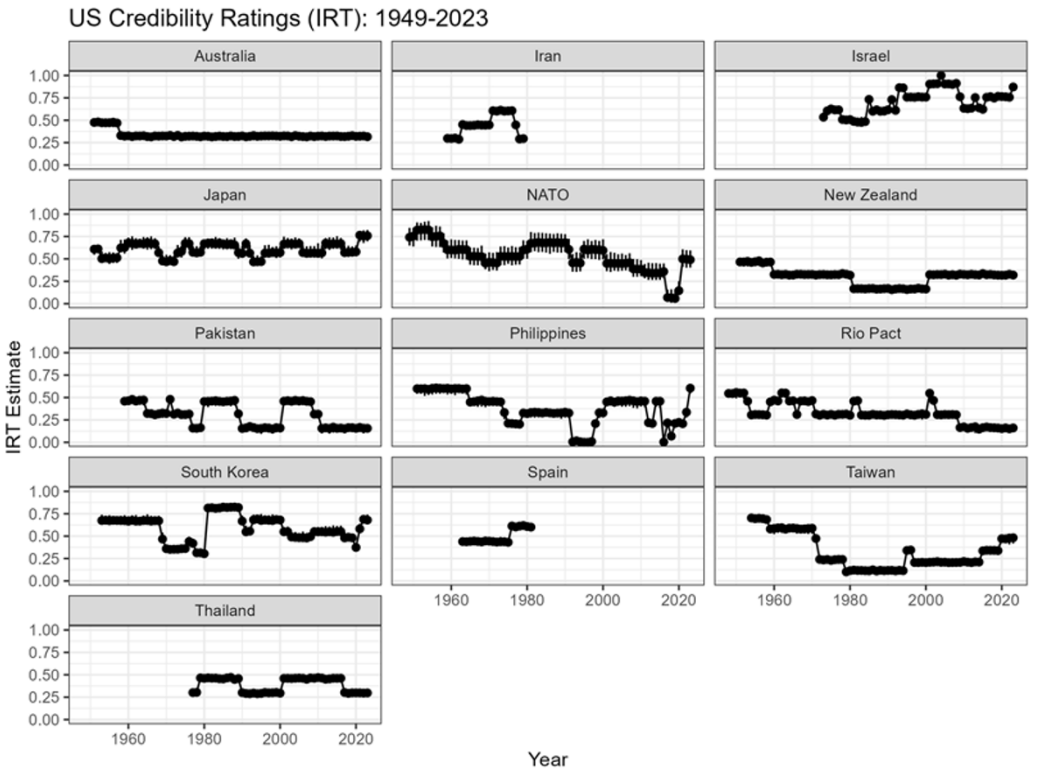


Figure 2. IRT estimates of credibility for U.S. alliance partners. Error bars mark the 90% credible interval.

In addition to the conceptual advantages, our measure has more complete coverage of alliances and years for the United States. The McManus and Nieman (2019) latent measure ends in 2012 and the statement data in 2010 (Blankenship, 2020). Furthermore, Taiwan is missing in some data because it is not a COW system member and South Korea did not join the UN General Assembly until 1991, which limits coverage of some interest-based measures.

6. Is credibility interdependent?

We now use our credibility measure to examine interdependence across U.S. alliances. Unlike some prior work, we focus on inferring interdependence, not adjusting for autocorrelation to make unbiased inferences about a covariate. To do so, we must account for three forms of correlation. These include contemporaneous spatial correlations of credibility between different alliances, persistent spatial correlations over time, and correlations over time within alliances. We estimate all three correlations but focus on spatial correlations in the same year because the three arguments make different predictions about how much interdependence persists as distance increases. Strong interdependence arguments suggest meaningful correlations across long distances, while conditional and no interdependence theories expect weak or absent correlations at great distances.

In testing, we are especially wary of inadvertently biasing results by fixing temporal and spatial lags in a model specification or hard-coding decay functions (Crescenzi *et al.*, 2012; Weisiger and Yarhi-Milo, 2015). We therefore use Gaussian process regression to estimate correlations in U.S. alliance credibility. These models are a flexible mix of traditional linear models and machine learning (Prati

https://doi.org/10.1017/psrm.2026.10132

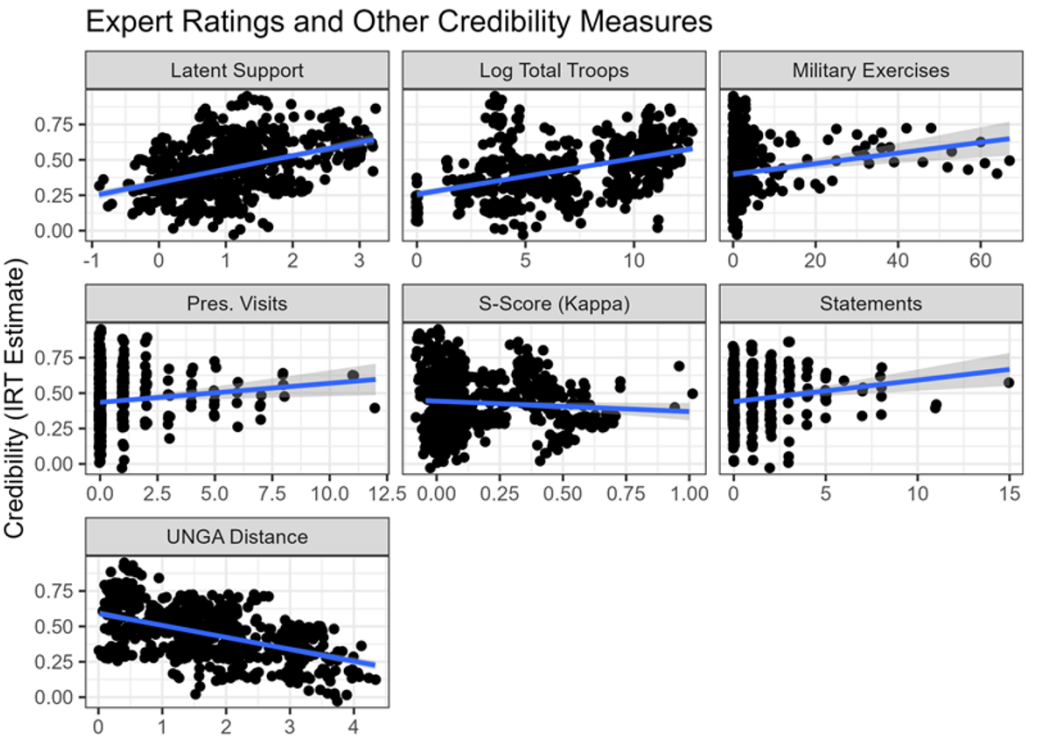


Figure 3. Correlations between expert credibility measure and other proxies for alliance commitment and credibility. Points mark alliance-year observations, and a linear smooth shows the correlation. Latent support, total troops, and UN voting distance are averaged for multilateral alliances, while we sum exercises, statements, and presidential visits.

et al., 2023),[8] and scholars often use them to examine spatial interdependence in continuous distances (McElreath, 2016; Gill, 2021; Aksoy and Carlson, 2022; Chen *et al.*, 2023). We estimate correlations between alliance observations in time and physical space because failing to account for one type of interdependence will impact our estimates of the other (Cook *et al.*, 2023).

The key intuition behind our approach is that alliance-year observations of credibility become more dissimilar and perhaps uncorrelated as geographic and temporal distance increases. Observations within an alliance are perfectly similar in geography, but differ in time. Observations of different alliances have different locations and become even less similar as the gap in years increases. Because the rate at which correlations decay is uncertain, we estimate it instead of fixing proximity.

The model estimates correlations between all the alliance-year observations in our data based on their year, latitude and longitude. Each observation has a unique intercept based on its geographic position and year, and the model estimates correlations between these alliance-year intercepts. The most important choice is how we group these latent parameters of the Gaussian process.

We report results from a model wherein all observations could be correlated depending on their distance from one another in both space and time. A second model we report in the appendix groups responses by year, estimating correlations between allies within years. The model we report here has the most complicated possible structure but fits far better than simpler models because it encapsulates the full correlation structure. Alternative structures lead to pathological behavior in cross-validation comparisons of model fit.

[8]Gaussian process regression uses a series of multivariate normal distributions over input functions to encode flexible representations of different data-generating processes.

https://doi.org/10.1017/psrm.2026.10132

Initial correlations between the closest possible data points, whatever their level, can decay to zero or remain as distance increases. This decay depends on two parameters that control the covariance function of the Gaussian process. One parameter sets the overall variance and another controls the scale of the process—the distance at which correlations matter. A key choice is the kernel encoding some prior beliefs about the shape of input functions. We use a squared exponential kernel because it encodes a smooth prior that is otherwise very flexible (Prati *et al.*, 2023). This allows correlations in credibility to persist over long distance in time and space or drop off very quickly.

Our outcome variable is the median credibility rating from the IRT model, rescaled between 0 and 1. This bounded outcome has a lumpy distribution, so we use ordered beta regression for the likelihood (Kubinec, 2023). We then fit models that use a dummy indicator of whether the alliance is bilateral, the log of U.S. troops present in the partner, U.S. changes in presidential party, partner democracy, and s-scores to predict mean credibility. We use these measures because they have the best temporal and cross-sectional coverage. In addition to these baseline covariates, we added the aforementioned Gaussian process term or kernel based on an observation's latitude, longitude, and year.[9] In the appendix, we provide a formal presentation of this Bayesian ordered beta regression with a constructed kernel that allows for non-independent errors based on time and space alongside some covariates outside the kernel.

7. Results

We find conditional correlations in credibility. Alliance credibility is correlated within regions and weakly or uncorrelated across regions. This is inconsistent with both strong interdependence and no interdependence arguments.

First, the Gaussian process regression produces some informative parameter estimates, including the maximum covariance and scale that controls the rate of decay. The maximum covariance gives the highest correlations between observations, and its 90% credible interval ranges between 0.23 and 1.2, so there is ample uncertainty in the Gaussian process. The latent scale parameter is very close to zero, which suggests that the strongest correlations occur at very short distances. This makes sense because the Gaussian process includes temporal autocorrelation within alliances, and these observations differ only in time.

The covariance and scale parameters inform the correlations between each pair of alliances in space and time. To begin, we present the maximum spatial correlation between every pair of alliance observations in Figure 4. This figure plots a matrix of correlations between alliances in the same year, leaving uncorrelated alliances and the diagonal blank. These correlations in the same year are the largest possible spatial correlations, as observations in different years have additional distance. These spatial correlations weaken as the temporal gap between observations increases, and we document this decay in the appendix.

Alliances in the same region have strong spatial correlations, while alliances in disparate regions are weakly correlated or uncorrelated. The Rio Pact is uncorrelated with all other U.S. alliances, and is therefore absent from Figure 4. Credibility in NATO and the bilateral U.S.-Spain pact (1963–1980) is strongly correlated, but NATO credibility with Israel, the next nearest U.S. ally, is less than one-fourth as strong.

The three U.S. alliances in the Middle East exhibit meaningful correlations. When the United States had an alliance with the Shah of Iran, this treaty was correlated with credibility toward Israel and Pakistan. Pakistan and Israeli perceptions of U.S. credibility are less correlated due to greater distance.

The bilateral alliances of the U.S. "hub and spokes" system in East Asia show strong credibility correlations. The Philippines, Taiwan, Japan, and South Korea are all highly correlated with immediate

[9] We average capital latitude and longitude for multilateral alliances. All models fit are with brms (Bürkner, 2017). To facilitate model estimation, we fit an approximate Gaussian process using Hilbert spaces, with 20 basis functions and the recommended default multiplicative constant of 5/4 (Riutort-Mayol *et al.*, 2023).

https://doi.org/10.1017/psrm.2026.10132

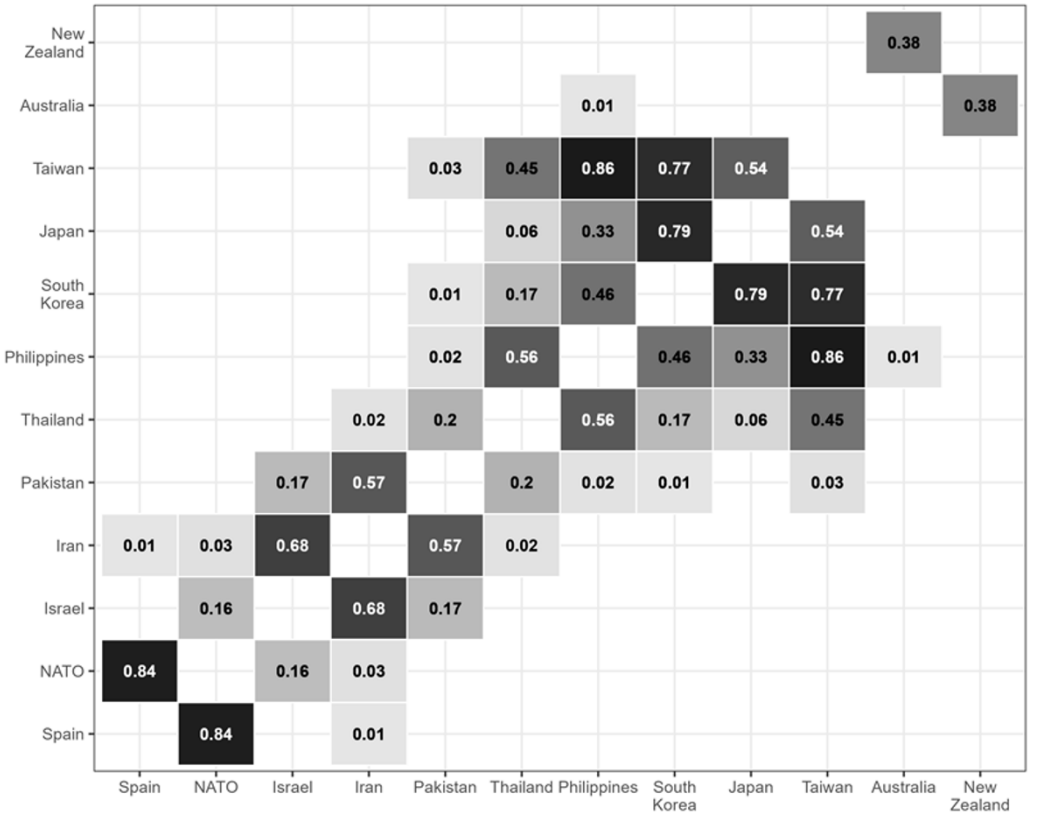


Figure 4. Matrix of the estimated correlation between U.S. credibility with alliance partners from 1950 to 2018. Each estimate gives the maximum spatial correlation between alliance partners, which then decays over time. Correlations of zero or one are left blank, so the Rio Pact is omitted entirely.

neighbors. Taiwan and the Philippines have the strongest spatial correlation of any alliance. Thailand, which is further away, is modestly correlated with the Philippines and Taiwan.

U.S. credibility to the ANZUS partners of Australia and New Zealand is also correlated. Otherwise, these alliances are uncorrelated with other Asian alliances because of the vast distances between them and other partners. This mirrors a general pattern of weak or zero correlations across regions.

Substantively, stronger correlations suggest that allied perceptions have a more similar underlying process. Smaller correlations imply much smaller credibility shifts because allied perceptions come from distinct sources. A correlation of .04, the median correlation between NATO and Iran, would mean that a .1 increase in U.S. credibility toward NATO would produce an increase of 0.004 in Iranian perceptions of U.S. credibility. Conversely, the 0.87 correlation between the Philippines and Taiwan suggests that if U.S. credibility toward the Philippines is 0.1 higher, U.S. credibility toward Taiwan would rise by 0.087, because the underlying processes are more similar.[10]

The spatial decline in credibility correlations is consistent with conditional credibility arguments. While there are correlations across regions, they are far weaker than correlations within regions. As a result, the Rio Pact is uncorrelated with other U.S. alliances. NATO exhibits a strong correlation with the U.S.-Spain defense pact and a modest correlation with Israel, but otherwise has weak ties to other U.S. commitments. Similarly, ANZUS members have correlated credibility perceptions and a weak link to the Philippines.

[10]This assumes equal variances in the two countries.

https://doi.org/10.1017/psrm.2026.10132

These results are robust to multiple robustness checks, which are available in the supplementary appendix. The changes include: (1) dropping predictors besides the Gaussian process terms and bilateral dummy, (2) dropping the Rio Pact because it is a geographic outlier that some observers do not consider an alliance, (3) fitting the model at different quantiles of the posterior, and (4) using median credibility from the expert ratings instead of the IRT results. We also check for regional clusters in expert discrimination and find little evidence of this.

8. Why is credibility correlated within regions?

While the preceding results suggest that alliance credibility is correlated within regions, they provide limited insights into why this pattern exists. The correlations could reflect changes in credibility itself, the diffusion of U.S. actions, or common factors. We unpack these possibilities with two additional tests, results for which are in sections five and six of the Supplementary appendix.

First, we test whether these correlations might follow from allies reacting to U.S. behavior in the same way. We do this with three measures of U.S. behavior: (1) whether the United States had backed down or won in recent militarized interstate disputes; (2) U.S. performance during alliance performance opportunities; and (3) U.S. alliance treaty terminations that were in violation of the alliance's terms. The results suggest that past actions have a small role in credibility correlations. However, effect sizes are small and sensitive to model specification.

Second, we use spatial regression models to test direct relationships between credibility values (autocorrelation) and common unobservables (spatial error). The spatial tests are critical because different spatial models reflect different processes. Contrary to strong interdependence arguments, which argue that credibility changes in one alliance impact credibility elsewhere, we find little autocorrelation. Instead, nonzero spatial error suggests that our Gaussian process models pick up correlated unobservables within regions. This and small effects of U.S. actions suggest that, following Press (2005), observed credibility correlations reflect region-specific assessments of U.S. capabilities, interests, and reliability given distinct threats.

9. Discussion and conclusion

Our findings support conditional credibility arguments. Alliance credibility is strongly correlated among close allies but weakly correlated beyond regions. This implies that alliance credibility is interdependent, but not on a global scale.

There is therefore little evidence for strong claims about interdependent or independent alliance credibility. We find little evidence of meaningful correlations in credibility across regions or spatial autocorrelation. As a result, upholding credibility in one region will have at best a marginal relationship with credibility in more distant alliances. But nearby allies do exhibit meaningful correlations due to spatial error.

The results have policy implications as well. Policymakers who are considering reducing alliance commitments can expect that lost credibility in one treaty matters most to proximate partners. But downgrading U.S. commitments in one region might not lead perceived credibility in other regions to decline. The results thus suggest some optimism that prioritizing alliance commitments in one region will not come at the expense of a tarnished reputation elsewhere. To the extent that support for strengthening or maintaining the status quo in U.S. alliances draws on fears that allies make inferences about American credibility across regions, our findings suggest this concern may be misplaced.

Our study has some limitations. For one, there is heterogeneity in how different partners within multilateral alliances like NATO perceive credibility, and our measure does not capture those differences. Moreover, we focus on U.S. allies, and other countries with fewer commitments may exhibit different patterns of interdependence and independence. In particular, one might object that two features may limit the generalizability of these findings beyond the United States. The first is that U.S.

https://doi.org/10.1017/psrm.2026.10132

alliances span the globe, which may make interdependence especially important. Yet given that our findings suggest that interdependence is strongest at the regional level, the results should apply to other countries with multiple allies in the same region. Second, U.S. alliances are highly asymmetric and many are far from the U.S. homeland. This could make allies particularly attuned to changes in U.S. credibility and how the United States behaves toward nearby partners.

Future research should therefore broaden the scope of this analysis beyond the United States, or consider interdependent credibility within multilateral U.S. alliances. Scholars could also use our credibility measure to predict other aspects of U.S. partners' foreign policy. Others could also tackle the causal mechanisms behind interdependence and clarify their relative weight. Researchers should expand and integrate work on the psychology of key actors as they do this.

Credibility is conditionally interdependent within regions. Upholding credibility matters most for proximate partners. Leaders need not fight for allies everywhere, but nor can they reduce commitment without undermining their credibility among nearby allies.

Acknowledgements. The authors thank Don Casler, Timothy Crawford, Florian Hollenbach, Tongfi Kim, Ashley Leeds, Paul Poast, Luis Simón, Jeff Zeigler, and participants at the 2024 Political Studies Association of Ireland Annual Meeting for helpful comments. Special thanks to the experts who answered our survey. We used Claude Code to check the code for errors or issues.

Declaration of interests. The authors declare no competing interests and acknowledge funding from the University of Miami and University College Dublin.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psrm.2026.10132.

References

Aksoy D and Carlson D (2022) Electoral support and Militants' targeting strategies. *Journal of Peace Research* **59**(2), 229–241.

Allen MA, Flynn ME and Martínez Machain C (2022) US global military deployments, 1950-2020. *Conflict Management and Peace Science* **39**(3), 351–370.

Alley J (2021) Alliance participation, treaty depth and military spending. *International Studies Quarterly* **65**(4), 929–943.

Bailey MA, Strezhnev A and Voeten E (2017) Estimating dynamic state preferences from United Nations voting data. *Journal of Conflict Resolution* **61**(2), 430–456.

Bak D (2018) Alliance proximity and effectiveness of extended deterrence. *International Interactions* **44**(1), 107–131.

Benson BV (2012) *Constructing International Security: Alliances, Deterrence, and Moral Hazard*. Cambridge University Press.

Blankenship B (2020) promises under pressure: Statements of reassurance in US alliances. *International Studies Quarterly* **64**, 1017–1030.

Braumoeller BF (2013) *The Great Powers and the International System: Systemic Theory in Empirical Perspective*. Cambridge University Press.

Bürkner P-C (2017) Brms: An r package for Bayesian multilevel models using stan. *Journal of Statistical Software* **80**(1), 1–28.

Buzan B and Wæver O (2003) *Regions and Powers: The Structure of International Security*. Cambridge University Press.

Carlson D and Zeigler J (2024) *Modeling Without Conditional Independence: Gaussian Process Regression for Time-Series Cross-Sectional Analyses*.

Chen Y, Garnett R and Montgomery JM (2023) Polls, context, and time: A dynamic hierarchical bayesian forecasting model for US senate elections. *Political Analysis* **31**(1), 113–133.

Chiba D, Johnson JC and Leeds BA (2015) Careful commitments: Democratic states and alliance design. *The Journal of Politics* **77**(4), 968–982.

Clare J (2013) The deterrent value of democratic allies. *International Studies Quarterly* **57**(3), 545–555.

Clinton J, Jackman S and Rivers D (2004) The statistical analysis of roll call data. *American Political Science Review* **98**(2), 355–370.

Colgan JD (2019) American bias in global security studies data. *Journal of Global Security Studies* **4**(3), 358–371.

Cook SJ, Hays JC and Franzese RJ (2023) STADL up! the spatiotemporal autoregressive distributed lag model for TSCS data analysis. *American Political Science Review* **117**(1), 59–79.

Crescenzi MJC, Kathman JD, Kleinberg KB and Wood RM (2012) Reliability, reputation, and alliance formation. *International Studies Quarterly* **56**(2), 259–274.

D'Orazio V (2013) International Military Cooperation: From Concepts to Constructs. Unpublished manuscript.

https://doi.org/10.1017/psrm.2026.10132

Dafoe A, Zwetsloot R and Cebul M (2021) Reputations for resolve and higher-order beliefs in crisis bargaining. *Journal of Conflict Resolution* **65**(7-8), 1378–1404.
Danilovic V (2002) *When the Stakes are High: Deterrence and Conflict among Major Powers*. University of Michigan Press.
Fuhrmann M and Sechser TS (2014) Signaling alliance commitments: Hand-tying and sunk costs in extended nuclear deterrence. *American Journal of Political Science* **58**(4), 919–935.
Gartzke E and Gleditsch KS (2004) Why democracies may actually be less reliable allies. *American Journal of Political Science* **48**(4), 775–795.
Gaubatz KT (1996) Democratic states and commitment in international relations. *International Organization* **50**(1), 109–139.
Gibler DM (2008) The costs of reneging: Reputation and alliance formation. *Journal of Conflict Resolution* **52**(3), 426–454.
Gill J (2021) Measuring constituency ideology using Bayesian Universal Kriging. *State Politics & Policy Quarterly* **21**(1), 80–107.
Harvey FP and Mitton J (2016) *Fighting for Credibility: US Reputation and International Politics*. University of Toronto Press.
Henry ID (2020) What allies want: Reconsidering loyalty, reliability, and alliance interdependence. *International Security* **44**(4), 45–83.
Henry ID (2022) *Reliability and Alliance Interdependence: The United States and Its Allies in Asia, 1949-1969*. Cornell University Press.
Hopf T (1994) *Peripheral Visions: Deterrence Theory and American Foreign Policy in the Third World, 1965-1990*. University of Michigan Press.
Jervis R, Yarhi-Milo K and Casler D (2021) Redefining the debate over reputation and credibility in international security: Promises and limits of new scholarship. *World Politics* **73**(1), 1–37.
Kim DG, Byun J and Ko J (2024) Remember Kabul? Reputation, strategic contexts, and American credibility after the Afghanistan withdrawal. *Contemporary Security Policy* **45**(2), 265–297.
Kim T and Simón L (2021) a reputation versus prioritization trade-off: Unpacking allied perceptions of US extended deterrence in distant regions. *Security Studies* **30**(5), 725–760.
Krebs RR and Spindel J (2018) Divided priorities: Why and when allies differ over military intervention. *Security Studies* **27**(4), 575–606.
Kubinec R (2023) Ordered beta regression: A parsimonious, well-fitting model for continuous data with lower and upper bounds. *Political Analysis* **31**(4), 519–536.
Lal A, Lockhart M, Xu Y and Zu Z (2024) How much should we trust instrumental variable estimates in political science? Practical advice based on 67 replicated studies. *Political Analysis* **32**(4), 521–540.
Lee S, Huygens A and Mitchell SM (2025) Do birds of a feather deter better? Cultural affinityand alliance deterrence. *International Interactions* **51**(1), 138–160.
Leeds B, Ritter J, Mitchell S and Long A (2002) Alliance treaty obligations and provisions, 1815-1944. *International Interactions* **28**(3), 237–260.
Leeds BA (2003) Alliance reliability in times of war: Explaining state decisions to violate treaties. *International Organization* **57**(4), 801–827.
Leeds BA and Anac S (2005) Alliance Institutionalization and Alliance Performance. *International Interactions* **31**(3), 183–202.
Leeds BA, Mattes M and Vogel JS (2009) Interests, institutions, and the reliability of international commitments. *American Journal of Political Science* **53**(2), 461–476.
Lindstädt R, Proksch S-O and Slapin JB (2020) When experts disagree: Response aggregation and its consequences in expert surveys. *Political Science Research and Methods* **8**(3), 580–588.
Little AT and Meng A (2024) Measuring democratic backsliding. *PS: Political Science & Politics* **57**(2), 149–161.
Lupton D (2020) *Reputation for Resolve: How Leaders Signal Determination in International Politics*. Cornell University Press.
Marquardt KL (2020) How and how much does expert error matter? Implications for quantitative peace research. *Journal of Peace Research* **57**(6), 692–700.
Marquardt KL and Pemstein D (2018) IRT models for expert-coded panel data. *Political Analysis* **26**(4), 431–456.
McElreath R (2016) *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press.
McManus RW and Nieman MD (2019) Identifying the level of major power support signaled for protégés: A latent measure approach. *Journal of Peace Research* **56**(3), 364–378.
Mercer J (1996) *Reputation and International Politics*. Cornell University Press.
Nieman MD and Gibler DM (2023) Peaceful neighborhoods and democratic differences. *The Journal of Politics* **85**(1), 92–106.
Poast P (2010) (Mis) using dyadic data to analyze multilateral events. *Political Analysis* **18**(4), 403–425.
Poast P (2013) Can issue linkage improve treaty credibility? buffer state alliances as a 'Hard Case'. *Journal of Conflict Resolution* **57**(5), 739–764.
Prati A, Chen Y, Montgomery J and Garnett R (2023) *A Gaussian Process Framework for Structured, Flexible, and Interpretable Machine-Learning Models*.
Press DG (2005) *Calculating Credibility: How Leaders Assess Military Threats*. Cornell University Press.
Rainey C and Jackson RA (2018) Unreliable inferences about unobserved processes: A critique of partial observability models. *Political Science Research and Methods* **6**(2), 381–391.

https://doi.org/10.1017/psrm.2026.10132

Rapp-Hooper M (2020) *Shields of the Republic: The Triumph and Peril of America's Alliances*. Harvard University Press.
Reed W (1997) Alliance duration and democracy: An extension and cross-validation of 'Democratic states and commitment in international relations'. *American Journal of Political Science* **41**(3), 1072–1078.
Reuning K, Kenwick MR and Fariss CJ (2019) Exploring the dynamics of latent variable models. *Political Analysis* **27**(4), 503–517.
Riutort-Mayol G, Bürkner P-C, Andersen MR, Solin A and Vehtari A (2023) Practical hilbert space approximate Bayesian gaussian processes for probabilistic programming. *Statistics and Computing* **33**(17), 1–28.
Schelling TC (1966) *Arms and Influence*. Yale University Press.
Schmidt A (2025) Damaged relations: How treaty withdrawal impacts international cooperation. *American Journal of Political Science* **69**(1), 223–239.
Schmitt E and Cooper H (2021) Top Defense Officials Acknowledge They Advised Against Withdrawing All Troops from Afghanistan. *New York Times*. https://www.nytimes.com/2021/09/28/us/politics/milley-austin-afghanistan-war.html.
Schultz KA (2001) Looking for audience costs. *Journal of Conflict Resolution* **45**(1), 32–60.
Smith A (1996) to intervene or not to intervene: A biased decision. *Journal of Conflict Resolution* **40**(1), 16–40.
Snyder GH (1997) *Alliance Politics*. Cornell University Press.
Snyder J (1991) Introduction. In Jervis R and Snyder J (eds), *Dominoes and Bandwagons: Strategic Beliefs and Great Power Competition in the Eurasian Rimland*. Oxford University Press.
Weidmann NB (2024) Recent events and the coding of cross-national indicators. *Comparative Political Studies* **57**(6), 921–937.
Weisiger A and Yarhi-Milo K (2015) Revisiting reputation: How past actions matter in international politics. *International Organization* **69**(2), 473–495.

Cite this article: Alley J and Blankenship B (2026) Is alliance credibility interdependent? *Political Science Research and Methods*, 1–17. https://doi.org/10.1017/psrm.2026.10132

https://doi.org/10.1017/psrm.2026.10132