International Studies Quarterly (1999) **43**, 115–144

Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions

Curtis S. Signorino

University of Rochester

and

Jeffrey M. Ritter

Harvard University

The pattern of alliances among states is commonly assumed to reflect the extent to which states have common or conflicting security interests. For the past twenty years, Kendall's τ_b has been used to measure the similarity of nations' "portfolios" of alliance commitments. Widely employed indicators of systemic polarity, state utility, and state risk propensity all rely on τ_b. We demonstrate that τ_b is inappropriate for measuring the similarity of states' alliance policies. We develop an alternative measure of policy portfolio similarity, S, which avoids many of the problems associated with τ_b, and we use data on alliances among European states to compare S to τ_b. Finally, we identify several problems with inferring state interests from alliances alone, and we provide a method to overcome those problems using S in combination with data on alliances, trade, UN votes, diplomatic missions, and other types of state interaction. We demonstrate this by comparing the calculated similarity of foreign policy positions based solely on alliance data to that based on alliance data supplemented with UN voting data.

1. Introduction

International relations scholars have devoted considerable effort to testing hypotheses derived from systemic and choice-theoretic theories of international politics. For each of these purposes researchers have attempted to measure and compare the similarity of states' foreign policies. Since Bueno de Mesquita (1975) it has become common practice to rely on Kendall's τ_b applied to alliance commitments as a measure of the similarity of two states' foreign policies.

Authors' note: An earlier version of this article appeared as Harvard Center for International Affairs Working Paper 97-7. Previous versions have also been presented at the 1997 Annual Meeting of the Midwest Political Science Association and the 1996 Summer Political Methodology Conference. We would like to extend special thanks to Bruce Bueno de Mesquita, D. Scott Bennett, and Erik Gartzke for providing us with data as well as helpful comments. We have also benefited from the comments of Vesna Danilovic, Chris Gelpi, Gary King, Lisa Martin, Eric Reinhardt, Jas Sekhon, Ken Shepsle, Richard Tucker, the participants in the Harvard Government Department's Rational Choice Discussion Group, and three anonymous reviewers. Gauss code for the S procedure can be found at http://www.rochester.edu/College/PSC/signorino/papers.htm. Support from the Harvard Center for International Affairs was provided during the writing of this article.

©1999 International Studies Association.
Published by Blackwell Publishers, 350 Main Street, Malden, MA 02148, USA, and 108 Cowley Road, Oxford OX4 1JF, UK.

Those interested in testing systemic theories of international politics use the τ_b measure of alliance portfolio similarity to identify alliance "clusters" and to measure the extent to which those clusters are discrete or overlapping (e.g., Bueno de Mesquita, 1975, 1978, 1981b; Bueno de Mesquita and Lalman, 1988; Ostrom and Aldrich, 1978; Organski and Kugler, 1980; Stoll, 1984; Stoll and Champion, 1985; Iusi-Scarborough, 1988; W. Kim, 1989, 1991; C. Kim, 1991). In a pioneering article Altfeld and Bueno de Mesquita (1979) proposed that alliance portfolios could be interpreted as revealed preferences over security issues. Since then, many scholars have employed the similarity of states' alliance portfolios as a useful indicator of the similarity of those states' security interests. These authors subject choice-theoretic models of international conflict to empirical tests, using τ_b as the basis for operational measures of states' willingness to take risks and of those states' expected utilities for challenging each other (e.g., Bueno de Mesquita, 1978, 1980, 1981a, 1985; Altfeld and Bueno de Mesquita, 1979; Berkowitz, 1983; Altfeld, 1984; Altfeld and Paik, 1986; Bueno de Mesquita and Lalman, 1986, 1988, 1992; Lalman, 1988; Iusi-Scarborough, 1988; C. Kim, 1991; W. Kim, 1991; Lalman and Newman, 1991; Kim and Morrow, 1992; Huth, Bennett, and Gelpi, 1993).

In this article we revisit the issue of how to measure the similarity of states' foreign policy positions. We begin with a simple question: how well does τ_b measure the similarity of two states' alliance policies? We argue that while Kendall's τ_b is a useful measure of association for ranked categorical data (e.g., see Levy, 1981), it is inappropriate to use Kendall's τ_b as an indicator of the similarity of states' alliance policy positions. Our analysis of the traditional measure of alliance portfolio similarity leads us to address broader concerns about the practice of interpreting the similarity of alliance portfolios as a measure of states' common interests and to develop a new measure better suited to this purpose than τ_b. Bueno de Mesquita's work was quite controversial in the early 1980s, but the controversy consisted primarily of denunciation and defense; few of Bueno de Mesquita's critics offered much constructive advice. Our argument does touch on the foundations of the empirical side of Bueno de Mesquita's research project, but with the aim of strengthening those foundations rather than simply chipping away at them. We provide a more appropriate measure of "similarity"—one that is also more consistent with the theoretical side of that research project—and we eliminate the strict reliance on alliance data to measure similarity of foreign policy positions.

This article proceeds as follows: in section 2, after defining terms like *similarity* and *alliance portfolio*, we outline the problems involved in using Kendall's τ_b as a measure of the similarity of states' alliance portfolios and the problems involved in inferring similarity of interests from data on military alliances. In section 3, we develop an alternative measure of similarity, S, which is generalizable to a larger foreign policy space. In section 4, we conduct a detailed comparison of τ_b and S, employing hypothetical examples to demonstrate how the measures work, individual empirical cases to illustrate the substantive basis for the differences between the measures, and a large-scale comparison to demonstrate that S paints a sufficiently different portrait of alliance portfolio similarity to warrant the attention of empirical researchers. Finally, we provide an example of how S can be used to assess the similarity of foreign policy positions by combining the information in the alliance data with information from other data sources. We conclude in section 5 by summarizing our main points and highlighting outstanding issues that should be addressed in future research.

2. Alliance Portfolios and Kendall's τ_b as a Measure of Alliance Policy Similarity

To set the stage, we should first explain exactly what we mean when we refer to a state's "alliance portfolio." The Correlates of War (COW) Alliances Data Set classifies alliances into four types, which in this article we code as follows: 0=no alliance, 1=entente, 2=neutrality or nonaggression pact, 3=mutual defense pact. We follow Bueno de Mesquita (1975:195) in assuming that these categories represent increasing degrees of formal alliance obligations between states and that it is therefore appropriate to treat the data as ordinal.[1] We also follow the convention of coding states as having implicit mutual defense pacts with themselves, since defense pacts lie at the high end of the ordinal scale and it seems reasonable to assume that states will defend themselves if attacked.

If the states in the system in a given year are indexed $k = 1 \ldots N$, then state i's alliance portfolio is an $N \times 1$ vector $A^i = [a_1^i\ a_2^i \ldots a_N^i]'$, in which each element $a_k^i \in \{0, 1, 2, 3\}$ represents i's alliance commitment to state k. A simple example may help clarify the notation. Table 1 displays the alliances between states identified by the Correlates of War as major powers in 1816 and 1905.[2] France's major-power alliance portfolio in 1905 is $A^{FRN} = [\,1\,3\,0\,0\,2\,3\,]'$.[3] A state's alliance portfolio, then, is simply the entire set of that state's alliance commitments in a given year.[4]

An alternative (equivalent) way of representing the data—which we will use throughout this article—is to treat two alliance portfolios as ordered cross-classifications of alliances and to represent this in a 4×4 contingency table. If A^i and A^j represent states i's and j's vectors of alliance policies toward N states indexed $k = 1 \ldots N$, then the elements of the contingency table are comprised of the joint rankings (a_k^i, a_k^j). As an example, Table 2(a) shows France's and Italy's alliance portfolios A^{FRN} and A^{ITA} for 1905, which are taken directly from Table 1(b). Based on the cross-classification of the rankings (a_k^{FRN}, a_k^{ITA}), we can form the contingency table shown in Table 2(b). Table 2(c) shows the corresponding contingency table of counts.

Bueno de Mesquita (1975:188–96) suggests that states' alliance portfolios reflect their security interests, and that states with similar alliance portfolios might be grouped together into "clusters," while states with dissimilar alliance portfolios are likely to have very different or even conflicting foreign policy goals. Comparing states' entire alliance portfolios allows us to take into account both their alliance commitment to each other and their alliance commitments to other states. These indirect ties are important indicators of common or conflicting interests: two alliance partners may very well have clashing obligations to other states, while two nonallied states may have very similar foreign policy goals reflected in their convergent obligations to other states.

Before we can begin to evaluate whether or not the similarity of states' alliance portfolios is in fact a good indicator of the similarity of their foreign policy positions we must first develop some sense of what it means for alliance policies to be "similar."

---

[1] This assumption is certainly not unproblematic, but we postpone further discussion of it until later in section 2.

[2] Several versions of the COW alliances data are in circulation; in order to assure replicability and to facilitate comparison with Bueno de Mesquita's earlier results, we rely on the version available from the ICPSR as I5602.

[3] Note that in the empirical analyses later in this article, each state's portfolio will include all the states in the European system, not just the major powers.

[4] Henceforth, we will refer to "alliance portfolio," "alliance policies," and "alliance commitments" interchangeably. The term *commitment* can be slightly confusing in this context, since it can refer both to a promise and to a true intention of carrying out that promise. Unless otherwise specified, we will use the former sense of the word: saying that state A has an alliance commitment with state B means only that A has promised to fulfill certain obligations; it says nothing about whether A really intends to fulfill those obligations. Finally, our "alliance portfolios" should not be confused with the investment portfolio models of alliances developed by scholars such as John Conybeare (1992).

TABLE 1. Major Power Alliances in 1816 and 1905

	FRN	*UK*	*GMY*	*AUH*	*RUS*
FRN	3	0	0	0	0
UK	0	3	3	3	3
GMY	0	3	3	3	3
AUH	0	3	3	3	3
RUS	0	3	3	3	3

(a) 1816

	UK	*FRN*	*GMY*	*AUH*	*ITA*	*RUS*
UK	3	1	0	0	0	0
FRN	1	3	0	0	2	3
GMY	0	0	3	3	3	0
AUH	0	0	3	3	3	1
ITA	0	2	3	3	3	0
RUS	0	3	0	1	0	3

(b) 1905

Each table element denotes the type of alliance the column nation has with the row nation. A state's alliance portfolio is the column vector of its alliances with each of the row nations (0=no alliance, 1=entente, 2=neutrality pact, 3=defense pact).

Essentially, two states' alliance portfolios are similar to the extent they share the same alliance commitments with each of the members of the international system. Table 1(a) displays a very straightforward case: in 1816, in the wake of the Napoleonic Wars, Britain and Germany had "perfectly similar" or identical major-power alliance portfolios, while those of Britain and France were completely dissimilar.

Because states are assumed to have mutual defense pacts with themselves, two states cannot have perfectly similar alliance portfolios unless they also have mutual defense pacts with each other—this restriction does not apply to their alliance ties to other states, however. Two states, i and j, may have a common interest in defending k from attack *or* in maintaining neutrality in the event that k is involved in hostilities *or* in consulting each other before taking military action relative to k *or* in offering k no pledge at all. Thus the alliance portfolios of Britain and France in 1816 would still have been "perfectly similar" if they had both had ententes with Russia and no alliance with Austria. Formally, two alliance portfolios, A^i and A^j, are perfectly similar when $a_k^i = a_k^j$ for all states k. Two states' alliance portfolios become less similar as their alliance commitments to the members of the international system diverge, and they become completely dissimilar when the states' commitments to each and every system member differ as much as possible, as in the case of Britain and France in 1816.

While the notion of "similarity" is not difficult to grasp intuitively, it is rare to find an example as clear-cut as we see in Table 1(a). Often, states' alliance ties look more like those in Table 1(b). How similar are the alliance portfolios of France and Italy in 1905? Are they more similar or less similar to each other than Britain's and Russia's are? By how much? To answer questions like these, we need an operational measure of alliance portfolio similarity. Bueno de Mesquita (1975:198) argues that "the degree of similarity in alliance commitments . . . can be summarized through the computation of an appropriate measure of association," and his use of Kendall's τ_b has become standard in the literature. We believe, however, that it is time to

TABLE 2. Contingency Tables Based on France's and Italy's Alliance Portfolios over Major Powers in 1905

(a)

	A^{FRN}	A^{ITA}
UK	1	0
FRN	3	2
GMY	0	3
AUH	0	3
ITA	2	3
RUS	3	0

(b)

ITA \ FRN	*0*	*1*	*2*	*3*
0		UK		RUS
1				
2				FRN
3	GMY AUH		ITA	

(c)

ITA \ FRN	*0*	*1*	*2*	*3*
0	0	1	0	1
1	0	0	0	0
2	0	0	0	1
3	2	0	1	0

(a) shows France's and Italy's alliance portfolios A^{FRN} and A^{ITA} in 1905; (b) displays the cross-classification of the alliance rankings (a_k^{FRN}, a_k^{ITA}); (c) shows the corresponding contingency table of counts. (The alliance categories along the top and left of the contingency tables are 0=no alliance, 1=entente, 2=neutrality pact, and 3=defense pact.)

reevaluate this standard approach: does τ_b—or any other measure of association—really capture "similarity" in the sense discussed above?

2.1. *Kendall's τ_b Measure of Association*

Kendall's τ_b is one among a host of measures that fall under the rubric of "measures of association." Chi-square and proportional reduction in error measures for nominal data, the Goodman-Kruskal γ for ordinal data, Spearman's ρ for interval data, and Pearson's product-moment correlation for continuous data are all measures of this type.[5] For the task at hand, Kendall's τ_b seems appealing because it is specifically designed to measure the association between two sets of ordinal rankings when "tied" rankings are permitted and because it is easily interpretable.

Assume two individuals i and j have ranked N items and denote those rankings by A^i and A^j, respectively. The calculation of τ_b is based on comparisons of pairs of

[5] For general references to these measures (and others cited in this article) see Bishop, Fienberg, and Holland, 1975, Kotz and Johnson, 1988, Kendall and Stuart, 1961, Kendall and Gibbons, 1990, and Liebetrau, 1983.

joint rankings (a_k^i, a_k^j) and (a_l^i, a_l^j) and whether those pairs of joint rankings are "concordant," "discordant," or tied. A pair of rankings (a_k^i, a_k^j) and (a_l^i, a_l^j) is considered concordant if $a_k^i > a_l^i$ and $a_k^j > a_l^j$ or if $a_k^i < a_l^i$ and $a_k^j < a_l^j$. They are considered discordant if $a_k^i > a_l^i$ and $a_k^j < a_l^j$ or if $a_k^i < a_l^i$ and $a_k^j > a_l^j$. If all pairings of joint rankings are concordant, then A^i and A^j are perfectly positively associated. If all are discordant, A^i and A^j are perfectly negatively associated.

To calculate τ_b for two rankings A^i and A^j of N items (see, e.g., Kendall and Stuart, 1961:562–63), first define a matrix X of all paired comparisons in A^i:

$$x_{kl} = \begin{cases} +1 & \text{if } a_k^i < a_l^i \\ 0 & \text{if } a_k^i = a_l^i \\ -1 & \text{if } a_k^i > a_l^i \end{cases}. \tag{1}$$

Similarly, define a matrix Y of all paired comparisons in A^j:

$$y_{kl} = \begin{cases} +1 & \text{if } a_k^j < a_l^j \\ 0 & \text{if } a_k^j = a_l^j \\ -1 & \text{if } a_k^j > a_l^j \end{cases}. \tag{2}$$

The measure for τ_b is then given by

$$\tau_b = \frac{\sum_{k,l} x_{kl} y_{kl}}{\sqrt{\sum_{k,l} x_{kl}^2 \sum_{k,l} y_{kl}^2}} \qquad k, l = 1, 2, \ldots N; k \neq l. \tag{3}$$

When A^i and A^j share the same number of ordinal levels, the contingency table is square and the τ_b measure of correlation takes on values in the interval [–1, 1], where $\tau_b = 1$ represents complete concordance in rankings, $\tau_b = -1$ represents complete discordance in rankings, and $\tau_b = 0$ represents independence in rankings.

The application of τ_b to measuring the similarity of alliance portfolios between states is straightforward. In a system of N states, the τ_b statistic compares the order in which state i ranks its alliance relationships with states $1, \ldots, N$ to the order in which state j ranks its alliance relationship with states $1, \ldots, N$. Consider states i's and j's pair of alliance rankings (a_k^i, a_k^j) and (a_l^i, a_l^j) for states k and l. When i has a stronger (weaker) alliance commitment to k than to l and j has a stronger (weaker) alliance commitment to k than to l, then the pair of rankings are concordant. When i has a stronger (weaker) alliance with k than with l, while j has a weaker (stronger) alliance with k than with l, their rankings are said to be discordant. With the two portfolios A^i and A^j of i's and j's alliance commitments to each nation $k = 1, \ldots, N$, we can then use equation (3) to calculate the association between i's and j's alliance commitments. A decade after Bueno de Mesquita first introduced this approach to measuring the similarity of states' alliance portfolios, Michael Wallace (1985:102–3) called it "a major advance in every respect" that represents "a notable improvement in sophistication" over previous measures.[6]

[6] Wallace nevertheless has strong reservations about using τ_b to measure systemic polarity and develops an alternate approach of his own.

2.2. "Not Tau-b": Problems with τ_b as a Measure of Policy Similarity

Unfortunately, while Kendall's τ_b is an elegant approach to measuring rank-order correlation, equation (3) is sufficient to show that τ_b does not measure "similarity." Kendall's τ_b reflects the extent to which states *i* and *j* rank their alliance commitments to paired members of the international system in the same order, whereas we would like to measure the extent to which states *i* and *j* have the same type of alliance commitments to each of the individual members of the international system. Table 3 displays comparisons of several hypothetical alliance portfolios in order to help clarify the difference between similarity and association.[7]

Perfectly Negative Association Does Not Imply Complete Dissimilarity of Alliance Policies

According to the conventional interpretation, a τ_b score of –1 should imply that the two alliance portfolios being compared are completely dissimilar. In fact, however, two alliance portfolios may be perfectly negatively associated without being completely dissimilar. $\tau_b = -1$ simply means that whenever state *i* ranks some state *k* higher than another state *l*, state *j* ranks *k* lower than *l*. In other words, *i*'s and *j*'s rankings of *k* and *l* do not have to be opposite in order to generate $\tau_b = -1$, only discordant. Tables 3(a)–(c) show three hypothetical pairs of alliance portfolios that are not completely dissimilar but that nevertheless produce a τ_b score of –1.

Consider Table 3(a), for example. Although it is true that *i* and *j* have opposing views about the ordinal relationship of their alliance commitments to the four system members, *i* and *j* do not have antithetical alliance policies: they are mutually allied to the states in the (1,2) and (2,1) cells, albeit at different levels of commitment. International relations researchers examining alliance patterns should not be pleased that τ_b takes on a value of –1 whenever all of the elements fall into the main negative diagonal of the contingency table, because truly opposite alliance policies occur only when all of the elements are concentrated in the (0,3) and (3,0) cells.

Tables 3(b) and (c) further emphasize this point. In both cases, with the exception of the ties, *i* and *j* have diametrically opposed rankings of their alliance commitments, so that $\tau_b = -1$. If we were using τ_b as a measure of alliance policy similarity, we would infer that *i* and *j* had completely different alliance policies. However, in Table 3(b), *i* and *j* actually agree that they should have some formal alliance commitment with all of the states in the system, including each other. Moreover, *i* and *j* have identical types of alliance commitments with two of their possible partners, represented by the observations in the (2,2) cell. Similarly, in Table 3(c), although *i* and *j* do not perfectly agree about the alliance types with the two other nations, they have closer alliance commitments to each other than in Table 3(b), and their commitments to the other members differ by only one category. It seems clear that the alliance portfolios compared in Tables 3(b) and (c) are more similar than those compared in Table 3(a), but all three cases generate a Kendall's τ_b of –1. As these three cases make clear, two portfolios may be perfectly negatively associated without being completely dissimilar.

Not All Identical Alliance Policies Can Be Measured with Association

In section 2.1 we noted that τ_b indicates perfect positive association ($\tau_b = 1$) of two portfolios when all the elements fall on the main positive diagonal of the

[7] We do not mean to imply that these specific patterns of alliance commitments necessarily appear commonly in the alliance data, or that the association measured by τ_b differs from similarity only in these cases. We submit these "ideal-type" examples to clarify an analytic argument, and we reserve consideration of the empirical significance of our observations until section 4.

TABLE 3. Problems with Using Association to Measure Policy Similarity

$A^i = [\,0\ 1\ 2\ 3\,]'$
$A^j = [\,3\ 2\ 1\ 0\,]'$

	0	1	2	3
0	0	0	0	1
1	0	0	1	0
2	0	1	0	0
3	1	0	0	0

$\tau_b = -1$
(a)

$A^i = [\,1\ 2\ 2\ 3\,]'$
$A^j = [\,3\ 2\ 2\ 1\,]'$

	0	1	2	3
0	0	0	0	0
1	0	0	0	1
2	0	0	2	0
3	0	1	0	0

$\tau_b = -1$
(b)

$A^i = [\,2\ 2\ 3\ 3\,]'$
$A^j = [\,3\ 3\ 2\ 2\,]'$

	0	1	2	3
0	0	0	0	0
1	0	0	0	0
2	0	0	0	2
3	0	0	2	0

$\tau_b = -1$
(c)

$A^i = [\,3\ 3\ 3\ 3\,]'$
$A^j = [\,3\ 3\ 3\ 3\,]'$

	0	1	2	3
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	4

τ_b = undefined
(d)

	0	1	2	3
0	10	0	0	1
1	0	0	0	0
2	0	0	0	0
3	1	0	0	0

$\tau_b = -.09$
(e)

	0	1	2	3
0	0	0	0	1
1	0	10	0	0
2	0	0	0	0
3	1	0	0	0

$\tau_b = -1$
(f)

	0	1	2	3
0	0	0	0	1
1	0	0	0	0
2	0	0	10	0
3	1	0	0	0

$\tau_b = -1$
(g)

	0	1	2	3
0	0	0	0	1
1	0	0	0	0
2	0	0	0	0
3	1	0	0	10

$\tau_b = -.09$
(h)

(a)–(c) illustrate that perfectly negative association does not necessarily imply complete dissimilarity of alliance policies. Note that in each case, even though the rankings are perfectly negatively associated, *i* and *j* have some similarity of alliance commitments. (d) illustrates that τ_b is undefined when there is no variation in either *i*'s or *j*'s rankings. As a measure of policy similarity, this seems problematic in the case shown, since *i* and *j* have identical alliance policies. (e)–(h) demonstrate how the association of two portfolios can change even when they maintain the same degree of agreement. In each case, *i* and *j* disagree on the same two alliances (i.e., with each other) and agree exactly on all ten other alliances. (The alliance categories along the top and left of the tables are 0=no alliance, 1=entente, 2=neutrality pact, and 3=defense pact.)

corresponding contingency table. Actually, the requirement is slightly stronger than this: the elements also cannot all fall within a single cell. In Table 3(d), the two states *i* and *j* have identical alliance commitments, which include defense pacts with each other and with the other two system members. It would be reasonable to expect a measure of alliance portfolio similarity to indicate that *i* and *j* have identical portfolios, but τ_b is undefined in this case. Recall from equations (1) and (2) that τ_b compares one state's ranking of elements to the other's ranking of the same elements. If either *i* or *j* ranks all of the elements in the same category there is no "order" to its ranking of the elements. As a result, one of the terms in the denominator of equation (3) is zero, and no meaningful value exists for τ_b.

Association May Change While Similarity Remains the Same

Tables 3(e)–(h) serve to illustrate another problem with using τ_b as a measure of alliance portfolio similarity: equally similar pairs of alliance portfolios may not be equally associated. These four tables differ from Tables 3(a)–(c) in that they posit an international system of twelve states rather than one of four states. In each of Tables 3(e)–(h), states *i* and *j* have mutual defense pacts with themselves, no alliance with each other, and identical alliance commitments to the other ten states. The cases differ only in the type of alliance commitments *i* and *j* have with the other ten states in the system: "no alliance" in case (e), ententes in case (f), neutrality/nonaggression pacts in case (g), and defense pacts in case (h).

Notice that in cases (e) and (h) the τ_b score is –.09, indicating mild dissimilarity, even though *i* and *j* are in perfect agreement about the types of alliance commitments they should have with ten of the twelve states in the system. In contrast, $\tau_b = -1$ in cases

(f) and (g), implying that i and j have completely dissimilar portfolios in these cases. This is rather surprising, considering that in cases (f) and (g) states i and j remain in complete agreement about the level of alliance commitment they should have with ten of the twelve states in the system and that the difference between their commitments to the remaining two partners is exactly the same as it is in (e) or (h). Although i's and j's alliance portfolios are equally similar across the four cases, the value of τ_b changes to reflect the linear association of the elements in the contingency table.[8]

In sum, then, Kendall's τ_b is a correlation coefficient designed to measure the association of two sets of rankings. For its intended purpose, τ_b is an appropriate measure. However, we suggest that the association of two alliance portfolios is not suitable as a measure of the similarity of alliance policy positions. Association does not necessarily imply similarity, and vice versa. Moreover, τ_b is undefined for certain types of identical policies. This critique is not limited to τ_b alone: *any* of the measures of association or correlation mentioned at the beginning of section 2.1 would be inappropriate as an indicator of the similarity of two states' alliance policies for essentially the same reasons.[9]

2.3. Alliance Commitments and Foreign Policy Interests

As we have noted, researchers have commonly used the τ_b measure of alliance portfolio similarity as an approximation of the extent to which pairs of states have common or conflicting security interests. Unfortunately, it is possible for states to have very similar alliance portfolios when it is not clear that they have any common security interests at all, and it is possible for states with very strong common security interests to have very dissimilar alliance portfolios. This problem deserves special attention because it arises not from the use of τ_b to measure portfolio similarity, but from the limitations of the available data on military alliances themselves.

It seems fair to assume that a "defense pact" represents the highest level of obligation between states and that a promise of mutual consultation embodied in an entente represents a lower level of obligation. Neutrality and nonaggression pacts are usually treated as representing an intermediate commitment, requiring more specific action than ententes but less support than defense pacts. Several scholars have raised concerns about whether or not this rank ordering of alliance commitments is appropriate.[10] While we share some of these concerns, we focus here on the often overlooked problems that arise from the "no alliance" category. For the sake of argument, we might think of three different types of states that fall into this category: those that have no alliance with each other because they are hostile to each other, those that have no alliance because they are irrelevant to each other's security, and those that have no alliance because of an implicit alignment that renders a formal treaty unnecessary.

---

[8] Technically, the changes in the value of τ_b in this example arise through the effects of ties on equation (3). In (e) and (h), there are two types of ties: those where $x_{kl} = 0$ and $y_{kl} = 0$ and those where either $x_{kl} = 0$ or $y_{kl} = 0$ but not both (see equations (1) and (2)). The former contribute nothing to either the numerator or the denominator of equation (3). However, while the latter contribute nothing to the numerator, they do contribute to the denominator. Because of this, the numerator of equation (3) is small compared to the denominator, resulting in a τ_b score near zero. In (f) and (g), there is only one type of tie: where $x_{kl} = 0$ *and* $y_{kl} = 0$. Since these contribute nothing to the numerator and denominator of equation (3) and since all other pairwise comparisons are discordant, τ_b is –1.

[9] Previous debates over Bueno de Mesquita's use of τ_b and alliance portfolios have not explicitly addressed whether τ_b measures the similarity of portfolios per se. See Majeski and Sylvan, 1984, Wagner, 1984, Nicholson, 1987, Khong, 1984, Levy, 1989; but see also the replies by Bueno de Mesquita (1984a, 1984b, 1987) and by Bueno de Mesquita and Lalman (1992:286–91).

[10] Wallace (1973:579–80) was one of the first to raise concerns about the ordinality of the standard coding of alliance data. Other authors have questioned the assumption of ordinality among the categories, often focusing on the "neutrality or nonaggression pact" category. See Sabrosky, 1980:196; Liska, 1962:32; Levy, 1981; and Fearon, 1997:86, fn. 37.

In empirical applications, researchers usually calculate the similarity of states' alliance commitments with all of the states in Europe or with all of the states in the world. It is therefore common for states' alliance portfolios to consist of fifteen to thirty states, and in some cases more than 120 states. Most states do not have alliances with each other, however, and the inclusion of strategically irrelevant states in alliance portfolios renders the portfolios poor indicators of common policy interests. Say, for example, we were examining the similarity of the two states *i*'s and *j*'s alliance policies. Let us assume *i* and *j* are neighbors, but their policy similarities are measured using their alliance portfolios over all the states in the international system in a given year. It might be the case that states outside of *i*'s and *j*'s region are simply irrelevant to *i*'s and *j*'s foreign policy decisions. Unfortunately, since *i* and *j* have no alliance with most of the states in the international system, their alliance portfolios would include mostly zeros—and the contingency table representing these portfolios will have most of its elements in the (0,0) cell.[11] As a result, because they have the same type of alliance commitments (i.e., "no alliance") with most of the other states in the world, their alliance portfolios may look very similar even if they have diametrically opposed alliance commitments with the states in their region. The inclusion of irrelevant states in a pair of alliance portfolios will therefore tend to produce a more positive τ_b than would otherwise be the case, although it is difficult to characterize the amount of bias precisely because it depends upon the locations of the other elements in the contingency table.

The obvious solution to this problem would be to ensure that the domain of alliance portfolios is appropriately specified for each empirical application. When used as indicators of foreign policy similarity, states' alliance portfolios should include only those states that can reasonably be considered "policy relevant." For example, Bueno de Mesquita and Lalman (1992) limit their consideration to alliances between members of the European region (as defined by Small and Singer, 1982:47–48) or to alliances between states that are sufficiently active in European affairs to influence political decision-making (e.g., Turkey and the United States). Since the alliances data themselves do not allow us to distinguish irrelevant observations of "no alliance" from strategically meaningful observations of "no alliance," narrowing the domain of relevant states by geography or by diplomatic activity is probably the best we can do for now.

This raises a second, related, issue. Even if we assume *i*'s and *j*'s portfolios to include only strategically relevant states, are alliances with each of those states equally important to *i*'s and *j*'s foreign policies? There may be theoretical or empirical reasons to assume that not all system members are of equal importance in terms of the resources they can bring to bear in an alliance. For example, most theories under the realist, neorealist, and neoliberal rubrics assume that states ally to increase their security. Therefore, in operationalizing such theories, it might be appropriate to weight the states proportionally to their military power in order to avoid exaggerating the importance of small states. For example, imagine *i* and *j* are major powers that are directly allied in a mutual defense pact in order to pursue their regional interests, but *i* also has a defense pact with some small, weak, distant state *k* in order to secure harbor rights or for reasons of diplomatic protocol. If *j* and *k* are not allied, *i*'s and *j*'s divergent commitments to irrelevant state *k* probably should not greatly offset their shared commitments to each other in determining the similarity of their foreign policy positions. Unfortunately, there is no meaningful way to "weight" ordinal concordance, so τ_b cannot incorporate the differential values of the same alliance with different partners.

[11] Note that we are not saying that *all* (0,0) elements represent irrelevant states. Rather, since we should expect to see "no alliance" with irrelevant states, there should be more (0,0) entries if irrelevant states are included in alliance portfolios than would exist if only relevant states were included.

Finally, states that fall into the "no alliance" category because they are implicitly aligned pose a somewhat different problem from "irrelevant" states in that they violate the assumption that the data are ordinal. While states that are hostile or indifferent to each other may be grouped in the "no alliance" category without destroying the ordinality of the data, states that are implicitly aligned with each other clearly share a higher level of commitment—perhaps as high as a defense pact. A number of examples easily come to mind. In 1914 it was common knowledge among the major powers that Russia styled itself the protector of the Balkan Slavs. However, the special relationships between Russia and the states of Rumania, Bulgaria, and Serbia were never codified in formal treaties of alliance. Similarly, the relationship between the U.S. and Britain prior to World War II and the relationship between the U.S. and Israel prior to Camp David are two more examples of implicit alliances that were widely recognized by policymakers. The "implicit alliance" problem is perhaps harder to fix than the problem of including irrelevant states. The ideal solution would be to collect more detailed alliance data that identified alignments between states. In the absence of such data we must find some other way to mitigate the effect of this violation of the orderings.

Given that the data on states' alliance commitments may not at times provide enough information to permit accurate inferences about the similarity of states' foreign policies, it seems natural that a wider variety of data should be brought to bear. Bueno de Mesquita (1975, 1981a) discusses a number of different data sources—for example, UN votes, diplomatic missions, IGO involvement, and trade—that could be used to measure the similarity of states' policy positions, but he settles on alliances data, arguing that these data are the most relevant to security issues and are available for the longest period of time for the greatest number of countries. Given the evident problems with relying solely on alliances data, we recommend using whatever data sources are available for as long as they are available (assuming, of course, the data are theoretically relevant). For example, even though UN voting only started in 1945, we can probably get better estimates of states' interests from 1945 to the present by supplementing the alliances data with UN data.[12] Supplementing the alliances data with other sources of data should help to distinguish states that are implicitly aligned with each other from decided enemies and should, more generally, provide a richer array of information with which to determine the policy positions of states.

3. Policy Portfolios and a Spatial Measure of Foreign Policy Similarity

Because τ_b can seriously misrepresent the degree to which two states' alliance portfolios are similar, and because of the difficulties in relying on the alliances data for information about the similarity of states' interests, we should consider alternative approaches to measuring foreign policy similarity. We cannot surmount these problems by replacing τ_b with some other correlation coefficient like Spearman's ρ, because in this context "association" is simply not the same concept as "similarity." Measures of "agreement," which measure the extent to which two vectors "agree" on the actual values of their respective elements, would seem to be a step in the right direction.[13] Existing measures of agreement are not perfectly suited to our needs, however, because they are designed to distinguish agreement from randomness

[12] For an attempt to get at similarity of interests using UN voting data, see Gartzke and Simon, 1996. Oneal and Russett (1997) use both alliance data and data on economic interdependence.

[13] See, for example, Bishop et al., 1975; Cohen's (1960) κ; Cohen's (1968) weighted κ; Davies and Fleiss's (1982) generalized κ; and other κ-like measures by Schouten (1982), O'Connell and Dobson (1984), and Berry and Mielke (1988, 1990).

rather than from disagreement. Moreover, some measures of agreement use only the nominal information of the categories, while those that claim to be measures for ordinal data actually impose interval assumptions on the rankings. To our knowledge, no measure of agreement has been developed for ordinal data that uses the rank information *and* respects the ordinality limitations. This is an information problem inherent to the data: granting partial credit for "close" agreement requires specification of the extent of that credit, which cannot be done without imposing a metric on the data space. Because this metric has important implications for the measure, we would prefer not to be bound by the restrictive assumptions developed in existing "agreement" measures, which may not be well-suited to modeling a foreign policy space.

Because no existing measure is appropriate, we develop a new *spatial* measure of foreign policy similarity—one that is consistent with spatial models of international politics.[14] We assume that a state makes choices over a number of policy dimensions; and that the vector of its multiple policy choices—i.e., its revealed *policy portfolio*—represents a point in (foreign) policy space. Based on this, our conception of "similarity" is very specific: the closer two states are in the policy space—i.e., the closer their revealed policy positions—the more "similar" their revealed policy positions. The further apart two states are in the policy space, the more dissimilar their revealed policy positions.[15]

More formally, we assume there are N dimensions to each policy portfolio. Thus far we have considered portfolios consisting entirely of alliance commitments with the N states in a system. However, we now allow policy portfolios to consist of any number of issues over which states make choices, whether they involve types of alliance commitments, amounts of trade, amounts of foreign aid, or levels of support for UN resolutions. We let state i's policy portfolio $P^i = [p_1^i \; p_2^i \cdots p_N^i]'$ represent a point in a compact, N-dimensional policy space, and similarly for state j's portfolio $P^j = [p_1^j \; p_2^j \cdots p_N^j]'$. If our data on policy positions were the true points in that space (versus mappings of intervals into ordered categories), we would then only need to make an assumption about a metric $d(P^i, P^j)$ on that policy space and denote how close policy P^i is to policy P^j by the distance between them, $d(P^i, P^j)$.[16] Letting $d^{\max}$ be the maximum possible distance between any two points in the policy space, we could transform this into a measure of similarity $S^*(P^i, P^j) = 1 - 2d(P^i, P^j)/d^{\max}$ with values on the interval [–1, 1], where –1 denotes two policies that are as far apart as possible and 1 denotes identical policy positions.

Because the available data are mappings of intervals into ordered categories, we must make a few additional assumptions. Let $L = [l_1 \, l_2 \cdots l_N]'$ be a vector of order-preserving scoring rules $l_k : p_k^i \rightarrow \mathbf{p_k}$, which map a data value p_k^i for state i's policy along dimension k to a value on the closed interval $\mathbf{p_k} \equiv [l_k^{\min}, l_k^{\max}] \subset \mathbb{R}^1$. Given the scoring rules for the dimensions, define $\Delta_k^{\max} = l_k^{\max} - l_k^{\min}$ as the maximum difference along dimension k. Finally, let $W = [w_1 \, w_2 \cdots w_N]'$ be a vector of weights over the N dimensions.

We define the similarity S of states i's and j's policy portfolios P^i and P^j, respectively, as

[14] This approach is also similar to those based on "measures of similarity and dissimilarity," which have been widely employed to assess the extent to which two vectors of (nominal, ordinal, interval, or continuous) data differ from each other (e.g., Kotz and Johnson, 1988:397–405).

[15] As with alliance portfolios, we will refer to "policy portfolios" and "revealed (foreign) policy positions" interchangeably.

[16] The absolute distance metric $d(x, y) = \sum_{i=1}^{N} |x_i - y_i|$ and the Euclidean distance metric $d(x, y) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2}$ are examples of two common metrics. See, for example, Protter and Morrey, 1991:133. These two distance functions have also been used to model indifference contours in spatial models (Ordeshook, 1993:22–23).

$$S(P^i, P^j, W, L) = 1 - 2\frac{d(P^i, P^j, W, L)}{d^{\max}(W, L)} \tag{4}$$

where

$$d(P^i, P^j, W, L) = \sum_{k=1}^{N} \frac{w_k}{\Delta_k^{\max}} \left| l_k(p_k^i) - l_k(p_k^i) \right| \tag{5}$$

and

$$\begin{aligned} d^{\max}(W, L) &= \max_{X^i, X^j} d(X^i, X^j, W, L) \\ &= \sum_{k=l}^{N} \frac{w_k}{\Delta_k^{\max}} (l_k^{\max} - l_k^{\min}) \\ &= \sum_{k=1}^{N} w_k \,. \end{aligned} \tag{6}$$

The term $d(P^i, P^j, W, L)$ represents the distance between the points P^i and P^j, given the scoring rules L, dimension weights W, and using an absolute distance metric.[17] $d(P^i, P^j, W, L) = 0$ when P^i and P^j are identical. $d(P^i, P^j, W, L) = d^{\max}(W, L)$ when P^i and P^j are as far apart as possible in the policy space. The $\Delta_k^{\max}$ term is used to normalize dimensions when portfolios contain dimensions with different scoring rules. This allows the maximum difference along each dimension to have the same effect and allows the weights W to determine the relative size (or "importance") of the dimensions. The distance $d(P^i, P^j, W, L)$ between policies P^i and P^j is transformed so that the measure of similarity falls on the interval $-1 \leq S \leq 1$, such that $S = 1$ represents complete similarity of policy portfolios and $S = -1$ represents complete dissimilarity. This standardization allows us to easily substitute S in place of τ_b in any of the various applications for which the τ_b measure of portfolio similarity has been used (e.g., in calculating system poles, state utilities, or risk propensities).

Recall that the inclusion of irrelevant states in alliance portfolios tends to lead to more positive τ_b scores and hence to inflated judgments of the similarity of states' policy positions. S is also affected by the inclusion of irrelevant states: when most of the elements in the contingency table fall into the (0, 0) category, S will correctly indicate that two states' alliance portfolios are quite similar in that they tend to have the same type of alliance commitments with each of their partners. While it remains important that researchers specify the domain of relevant alliance partners carefully, S, unlike τ_b, can also incorporate a vector of weights W directly into the similarity measure in order to reflect the fact that not all alliances are equally important indicators of common interests between states.

In some of the hypothetical examples we discuss in this article we will assume that the states are equally relevant in foreign policy terms. We therefore use a uniform weighting scheme such that $w_k = 1/N\ \forall k$, and for the sake of brevity and clarity we indicate this weighting scheme with the subscript u, so that $S_u = S(P^i, P^j, w_k = 1/N\ \forall k, L)$. In section 2.3 we argued that one reasonable weighting scheme for empirical use might involve weighting each alliance according to the partner's military capabilities to ensure that commitments to "important" states have a larger impact

[17] Again, any distance function could be used in equation (5).

on our measure of policy similarity than commitments to weak states. We use such a weighting scheme later in this article, and refer to the capability-weighted similarity score as S_c to distinguish it from the uniform weighting scheme S_u. Let c_k be nation k's military capabilities and $C = \Sigma_{k=1}^{N} c_k$ be the total military capabilities in the relevant system in that year.[18] Then $S_c = S(P^i, P^j, w_k = c_k/C\ \forall k, L)$.[19] We shall see shortly that the choice of a weighting scheme to use with S is not trivial. Although somewhat crude, capability weighting seems to be a good "first cut" that is theoretically consistent with much of the literature.

As we have suggested, S can be used to calculate the similarity of policies where portfolios include only one type of policy data (e.g., alliances data *or* UN voting data *or* trade data, etc.) or where the portfolios consist of multiple types of portfolio data (e.g., alliances data *and* UN voting data *and* trade data). To calculate policy similarity from several different types of data, one would first code the alliance commitments, UN votes, or trade amounts into their own policy issue portfolios. The information provided by these policy issues may then be combined by creating a stacked vector of the issue portfolios, providing intra- and inter-issue weights, and using S to calculate the similarity of the multi-issue stacked portfolios. Alternatively, one could take the weighted average of the individual issues' similarity scores.

In practice, calculating S is quite simple, as we demonstrate with a brief example. Imagine we wish to measure the similarity of the policy positions of two states A and B based upon their alliance commitments and their UN voting records. Table 4(a) displays the alliance commitments of A and B with the four nations in our hypothetical system (A, B, C, D) along with the capability weights attributed to these nations. Table 4(b) displays A's and B's votes (1="against," 2= "abstain," 3= "for") over four UN resolutions, with the resolutions weighted equally with respect to each other.

In this case, A's and B's alliance policies are somewhat similar, but not overwhelmingly so: they have neutrality pacts with each other, and diametrically opposing alliance commitments with nations C and D. The weighting implies that alliances with A and B are considered more important than alliances with C and D. Clearly, A's and B's UN voting policies are much more similar than their alliance policies: they agree on every resolution but one—and even in that one case, their disagreement is not the maximum possible.

Assuming we use the scoring rules $l_k(a_k^A) = a_k^A$ and $l_k(a_k^B) = a_k^B\ \forall k$ for the alliance portfolios—i.e., set the scores to the alliance data values—and that we do the same for the UN vote portfolios, then the separate similarity scores calculated for the alliance portfolios and UN voting portfolios in Tables 4(a) and (b) are $S = .2$ and $S = .75$, respectively, indicating that while A's and B's alliance policy positions are only slightly similar, their UN voting policy positions are quite close.

The combined policy portfolios for nations A and B can be represented as the stacked vectors $P^A = \left[A^{A'}\ V^{A'}\right]'$ and $P^A = \left[A^{B'}\ V^{B'}\right]'$, respectively. Similarly, let W be the stacked vector of the weights. Note that stacking the two given vectors of weights has intra- and inter-issue weighting implications. Each weight vector in Tables 4(a) and (b) specifies how much each issue element is weighted with respect to the other

[18] We refer here to capabilities as measured by the COW National Material Capabilities data set. In our examples, the "relevant system" will be the European region as defined by Bueno de Mesquita and Lalman (1992).

[19] This approach should not be confused with the "weighted τ_b" scores used in Bueno de Mesquita and Lalman, 1992:295–97. There, Bueno de Mesquita and Lalman use differences in τ_b scores to weight the amount of support (i.e., capabilities) a nation is expected to contribute to one nation versus another. Capabilities are therefore *external* to τ_b. In contrast, capability weights are employed *internally* in S_c to define the relative sizes of each policy dimension. Note that by definition, τ_b cannot incorporate weights internally, since it is based on the concordance or discordance of pairs of rankings.

TABLE 4. Hypothetical Example

Nation	*Weights*	*Alliance Portfolios* A^A	A^B
A	.5	3	2
B	.4	2	3
C	.05	3	0
D	.05	0	3

(a)

UN Case	*Weights*	*UN Vote Portfolios* V^A	V^B
1	.25	3	3
2	.25	3	3
3	.25	1	1
4	.25	2	3

(b)

(a) shows *A*'s and *B*'s alliance commitments with four nations in the system (*A*, *B*, *C*, *D*), as well as the weights assigned to those nations. As can be seen, *A*'s and *B*'s alliance policies are somewhat similar, but not overwhelmingly so. This is reflected in a similarity score of $S = .2$. (b) shows *A*'s and *B*'s votes (1="no", 2="abstain", 3="yes") over four UN resolutions, with weights assigned uniformly to the resolutions. *A*'s and *B*'s UN voting policies are much more similar than their alliance policies, yielding $S = .75$. The similarity of their combined policy portfolios is $S = .475$.

elements within that policy issue. Since each weight vector sums to one, that implies that the alliance issue is weighted equally with the UN voting issue. The maximum difference along each dimension, $\Delta_k^{\max}$, is 3 for the alliance dimensions and 2 for the UN voting dimensions. Then the weighted distance between the two portfolios is

$$\begin{aligned} d(P^A, P^B, W, L) &= \frac{.5}{3}|3-2| + \frac{.4}{3}|2-3| + \frac{.05}{3}|3-0| + \frac{.05}{3}|0-3| + \\ &\quad \frac{.25}{2}|3-3| + \frac{.25}{2}|3-3| + \frac{.25}{2}|1-1| + \frac{.25}{2}|2-3| \\ &= .525 \end{aligned}$$

and the maximum possible distance is

$$\begin{aligned} d^{\max}(W, L) &= .5 + .4 + .05 + .05 + .25 + .25 + .25 + .25 \\ &= 2 \end{aligned}$$

and the similarity score S is

$$\begin{aligned} S &= 1 - 2(.525)/2 \\ &= .475. \end{aligned}$$

This could also have been calculated as the weighted average of the individual similarity scores: $S = [(1).2 + (1).75]/2 = .475$.

Before turning to a comparison of S and τ_b, it should be noted that S does impose two assumptions on the data. First, using a spatial measure requires that we make some assumptions concerning what the ordinal data represent in the policy space. Here, those assumptions take the form of the scoring rule, weights, and distance metric. In effect, these define the space and allow us to measure the distance between two points in it. Since these are modeling assumptions, S will necessarily be an approximation to the "true" similarity of foreign policy positions S^*.[20] The specification of the policy space, of course, has implications for the value of S. One strength of our approach is that it allows for considerable flexibility in modeling the policy space. The scoring rule may be as simple as setting the intervals to the rank values of alliance types: $l\left(p_k^i\right) = p_k^i \ \forall k$. With no other information available, this may be an acceptable scoring rule.[21] However, different scoring rules may be devised. For example, if one does not believe that the difference in commitment between defense pacts and neutrality pacts is the same as the difference between ententes and "no alliances," then one might specify any number of different scoring rules for the ranks, provided they can be convincingly justified on theoretical or empirical grounds. Moreover, one may specify different scoring rules for each of the N dimensions.

The second assumption placed on the data is that nations are assumed to view the alliance categories in the same way. In other words, every nation has a similar conception of the commitment embodied in an entente, in a neutrality pact, and in a defense pact. We do not feel this is an heroic assumption to make. In fact, the COW alliance data set coding rules for the alliance types imply a common understanding among the partners of their alliance responsibilities. A more heroic assumption along these lines concerns the weights W, which are assumed shared by all countries. This might be the case in the hypothetical situation noted above—where nations weight a possible alliance partner by the capabilities it would bring to the table. However, one can easily think of two nations having different weights for alliance partners based on language, culture, or other nonmaterial factors. Still, even this would not be difficult to incorporate into S. In equation (5), one could distribute w_k inside the absolute value term and then define each side's weights as heterogeneous (e.g., as w_k^i and w_k^j). Alternately, one could account for the heterogeneous valuation of the alliances through the scoring rules. We refrain from exploring such complications here.

4. Comparing S and τ_b

It should be clear from a direct comparison of their formulas that S measures portfolio similarity much more directly than τ_b. It is true that both measures take on values on the interval [–1, 1] and that these endpoints are identified as representing complete dissimilarity of policies and complete similarity of policies, respectively. This is, however, the only resemblance between the measures. τ_b measures the extent to which two sets of rankings are concordant: $\tau_b = 1$ indicates complete concordance between two rankings, $\tau_b = -1$ represents complete discordance in the rankings, and $\tau_b = 0$ represents independence in the rankings. As we have seen, however, concordant rank orderings of alliance partners do not necessarily imply

[20] One might well ask if using S simply trades the error in using τ_b for the approximation error in S. We believe this misses the point: similarity and association are not the same concept. Given the specification of a policy space, S by construction measures the similarity of policy positions. τ_b, on the other hand, cannot measure similarity even if all sources of error were eliminated. Simply put, the difference is between approximately measuring the correct concept versus correctly measuring the wrong concept.

[21] In fact, this is what we use in the hypothetical and empirical examples that follow.

similar types of alliance commitments to each of the partners. In contrast, S measures the distance between two policy positions in a policy space, given assumptions about the scoring rules and importance of policy dimensions. $S = 1$ means that the two policy positions are identical — i.e., they are the same point in the policy space. $S = -1$ indicates that the two points are as far apart as possible in the policy space, and $S = 0$ means that the distance between the two points is half the maximum it could be. The spatial positions of the policy portfolios diverge as the values of each of the corresponding elements in the portfolio vectors diverge, mirroring our intuitive definition of portfolio "similarity" from section 2. Moreover, S has three additional advantages over τ_b: it is able to incorporate data from several different sources in order to overcome the limitations of the alliance data as indicators of foreign policy interests; it allows for the weighting of observations in order to minimize the effects of "irrelevant" elements; and as a spatial measure it is more theoretically compatible with the standard applications involving state "utilities" and risk propensities than τ_b is.

In this section, we attempt to provide a sense of how S and τ_b differ in practice. We begin by reexamining the hypothetical cases from section 2.2 in order to clarify how the measures differ in the situations where τ_b does not perform well. We then present two empirical examples, the first illustrating the importance of weighting the different policy dimensions, and the second showing the importance of using multiple sources of policy data. Finally, moving beyond individual and anecdotal examples, we examine the alliance portfolios of two pairs of European states over long periods of time, and we examine the similarity of the alliance portfolios of every dyad of European major powers since the Napoleonic Wars in order to demonstrate that the similarity scores generated by S differ substantially from those generated by τ_b.

While we realize that many of our readers will be interested in how S alters previous research findings based on τ_b, we have chosen not to conduct replications of prior work here. Such replications are beyond the scope of this article and simply are not required to prove our central argument. In fact, replications would introduce a host of other issues that might distract from our main point. To the extent that we are able to show that S is a better measure of alliance portfolio similarity and that S and τ_b often take on very different values empirically, conducting replications with S can only lead to more accurate estimates of the substantive parameters and their standard errors, regardless of whether previous results are strengthened or weakened. For the present, we restrict ourselves to showing both logically and empirically that S more accurately measures the similarity of alliance portfolios than does τ_b—and that the difference is substantial enough that international relations scholars should take notice.

4.1. S Applied to the Hypothetical Examples

We now reexamine the hypothetical examples of Table 3, comparing S to τ_b in those cases where τ_b did not perform well as a measure of alliance policy similarity. Again, we are not arguing that these cases appear in the empirical data more often than others or that these are the only cases where τ_b does not perform well; we present them to highlight certain differences between S and τ_b. Rather than assign arbitrary capability weights to the imaginary states in these examples, we simply assume the states are all equally strategically relevant and we compare τ_b to S_u for each of the examples in Table 5.

Tables 5(a)–(c) were used in section 2.2 to demonstrate that, contrary to the conventional interpretation of τ_b, perfectly negative association does not necessarily imply complete dissimilarity of alliance policies. Although τ_b generates a value of

–1 for each of these cases, S_u changes based on the distance between the policy positions. In Table 5(a), S_u = –.33, reflecting the fact that *i* and *j* do not have completely dissimilar alliance policies because they share alliances (of different types) with the states in the (1,2) and the (2,1) cells. Completely dissimilar policy positions would occur only if all the elements fell in the (0,3) and (3,0) cells, and *S* (regardless of the weighting scheme used) would take on a value of –1 in that case.

In Table 5(b), *i* and *j* have moved closer in their alliance commitments to each other and actually agree on the exact level of the alliance commitments with the two other states. While τ_b = –1, reflecting the discordance of the rankings, S_u = .33, indicating that the alliance policies in Table 5(b) are more similar than those depicted in Table 5(a). In Table 5(c), *i* and *j* have relatively closer alliance commitments with each other, but they disagree by one category on the level of alliance commitment with the other two states. S_u = .33, reflecting moderately similar alliance policies—which are, again, more similar than in Table 5(a). τ_b would once again incorrectly imply complete disagreement.

Because *S* is a spatial measure of similarity, it is defined when there is no variation in the elements of a policy portfolio. In Table 5(d), *i* and *j* have identical alliance commitments with each other and with the other two states. Any measure of foreign policy similarity should identify these two portfolios as identical. As we saw in section 2.2, τ_b is undefined because there is no variation in the rankings. However, *S* = 1 (regardless of the weighting), correctly indicating that *i* and *j* have identical alliance portfolios. Note that for τ_b to be undefined, it only takes one portfolio without variation in the rankings. The assumption that states have mutual defense pacts with themselves limits the potential for cases like Table 5(d) to arise in empirical alliance data, but once we move away from strict reliance on the alliance data we run the risk that other types of policy portfolios (e.g., UN voting portfolios) might be more prone to exhibit patterns for which τ_b would be undefined. In any of these cases, *S* would be defined—and the results would be interpretable with respect to a foreign policy space.

The final set of hypothetical examples is shown in Tables 5(e)–(h). Recall that in all four cases, *i* and *j* have no alliance with each other, but they have identical alliance commitments to the other ten states in the system. In effect, although the level of commitment to the ten other states increases steadily from (e) to (h), the similarity of the states' alliance commitments remains constant across all of the examples. As we noted in section 2.2, the τ_b score changes from (e) to (f) and from (g) to (h) due to the differences in the ties in the rankings of pairs of alliances. In contrast, because the distance between *i*'s and *j*'s policy positions is the same in each example, S_u indicates that the four cases have the same (fairly high) level of policy similarity.[22] In sum, then, in each of the examples in Table 5, using τ_b as a measure of policy similarity would yield misleading results, whereas *S* produces substantively meaningful measures of alliance policy similarity.

4.2. Irrelevant States and the Weight of Different Foreign Policy Dimensions: Germany and Russia, 1914

In empirical applications, the alliance portfolios employed are often much larger than those presented in the hypothetical examples or even the empirical examples in Table 1. While one can easily think of applications where one might restrict

[22] This example also raises an important issue concerning the difference between similarity of policy positions and the overall (or joint) level of commitment. *i*'s and *j*'s alliance portfolios are perfectly similar if they have the same types of alliances with all the states in the system, regardless of the level of alliance commitments they agree on. Thus, S_u should *not* increase steadily from Table 5(e) to (h). The level of overall commitment may be an interesting variable, but it is not pertinent to measuring the similarity of policy positions.

TABLE 5. Comparing S_u and τ_b for Hypothetical Cases

	0	1	2	3
0	0	0	0	1
1	0	0	1	0
2	0	1	0	0
3	1	0	0	0

$\tau_b = -1$
$S_u = -.33$
(a)

	0	1	2	3
0	0	0	0	0
1	0	0	0	1
2	0	0	2	0
3	0	1	0	0

$\tau_b = -1$
$S_u = .33$
(b)

	0	1	2	3
0	0	0	0	0
1	0	0	0	0
2	0	0	0	2
3	0	0	2	0

$\tau_b = -1$
$S_u = .33$
(c)

	0	1	2	3
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	4

τ_b = undefined
$S_u = 1$
(d)

	0	1	2	3
0	10	0	0	1
1	0	0	0	0
2	0	0	0	0
3	1	0	0	0

$\tau_b = -.09$
$S_u = .67$
(e)

	0	1	2	3
0	0	0	0	1
1	0	10	0	0
2	0	0	0	0
3	1	0	0	0

$\tau_b = -1$
$S_u = .67$
(f)

	0	1	2	3
0	0	0	0	1
1	0	0	0	0
2	0	0	10	0
3	1	0	0	0

$\tau_b = -1$
$S_u = .67$
(g)

	0	1	2	3
0	0	0	0	1
1	0	0	0	0
2	0	0	0	0
3	1	0	0	10

$\tau_b = -.09$
$S_u = .67$
(h)

In (a)–(c), S_u identifies that the alliance policies in (b) and (c) are more similar than those in (a). In (d), S_u identifies that i and j have identical portfolios. And in (e)–(h), S_u identifies that the similarity of the alliance policies remains constant across each case. (The alliance categories along the top and left of the tables are 0=no alliance, 1=entente, 2=neutrality pact, and 3=defense pact.)

alliance portfolios to covering only major powers, it is more typical to see portfolios over all the nations in a particular region or over every nation in the entire international system for a given year. As we have noted, the inclusion of strategically irrelevant partners in states' alliance portfolios can seriously distort the assumed relationship between the similarity of states' alliance commitments and the similarity of their policy interests. Even when portfolios are limited to strategically relevant partners, it may be inappropriate to count all alliances as equally informative indicators of states' policy positions. We have suggested that S_c, by incorporating capability-weighting into the measure of similarity, can cope with these problems more effectively than τ_b.

The differences between τ_b, S_u , and S_c are nicely illustrated by a comparison of Germany's and Russia's alliance commitments in 1914, as displayed in Table 6(a). The leftmost column lists the states identified by the Correlates of War as members of the European region in 1914, plus Turkey.[23] The second column indicates each state's material capabilities measured as a proportion of the sum of all states' capabilities (i.e., c_k/C). The remaining two columns display Germany's and Russia's alliance portfolios across the European region. Table 6(b) shows the contingency table generated by cross-classifying Germany's alliance portfolio with Russia's, along with the τ_b, S_u, and S_c measures of the similarity of those alliance portfolios.

The value of τ_b is easy to understand in light of our earlier discussions. Most of the elements in Table 6(b) are concentrated in the (0,0) cell. Moreover, most of the non-(0,0) elements are in the (3,0) and (0,3) cells. The situation is similar to that shown in Table 5(e), and τ_b is close to zero because of the effect of ties (see the discussion in footnote 8). It is hard to see how the value of τ_b makes any sense either as a measure of the similarity of Germany's and Russia's alliance portfolios or as a measure of the similarity of their foreign policy positions.

[23] This follows Bueno de Mesquita and Lalman's practice of coding the U.S. as a member of the European system only after 1916 (Bueno de Mesquita and Lalman, 1992:281).

TABLE 6. Comparison of Alliance Similarity for Germany and Russia, 1914

(a)

Nation	*System Cap*	*Portfolios GMY*	*RUS*
GMY	.25	3	0
RUS	.21	0	3
UK	.20	0	1
FRN	.08	0	3
AUH	.08	3	0
ITA	.05	3	1
BEL	.03	0	0
SPN	.03	0	0
TUR	.02	0	0
NTH	.01	0	0
SWD	.01	0	0
RUM	.01	3	0
POR	.01	0	0
SWZ	.01	0	0
GRC	.00	0	0
DEN	.00	0	0
YUG	.00	0	0
BUL	.00	0	0
NOR	.00	0	0
ALB	.00	0	0

(b)

		RUS 0	1	2	3
GMY	0	13	1	0	2
	1	0	0	0	0
	2	0	0	0	0
	3	3	1	0	0

$\tau_b = .03$
$S_u = .4$
$S_c = -.45$

(a) shows Germany's and Russia's alliance portfolios with all the European nations in 1914, along with the nations' proportion of system capabilities. (b) displays the contingency table based on those portfolios and the values of τ_b, S_u , and S_c.

In contrast, even though they take on opposite signs, the scores for S_u and S_c are very intuitive, provided we remember how the different weighting schemes affect the substantive interpretations of the numbers. If we ignore the states' capabilities and treat each possible alliance partner as being equal, the two portfolios look fairly similar in that they are both mostly composed of zeros. Germany and Russia agree precisely on the type of alliance commitment they share with thirteen of their twenty potential partners, and they very nearly agree on their relationship with a fourteenth.[24] This moderate level of similarity is reflected in $S_u = .4$.

If, on the other hand, we assume that a state's relevance as an ally is proportional to the capabilities it can bring to the battlefield, then Germany's and Russia's alliance policies appear to be quite different. Although Germany and Russia are coded as having exactly the same level of alliance commitment with thirteen of the twenty

[24] Once again, it may seem counterintuitive to treat "no-alliance" relationships as points of agreement, but if the states included in the portfolio are relevant to the referent nations' decision making, then (0,0) reflects the same degree of agreement along a particular policy dimension as does (3,3). A joint ranking of (3,3) may indicate greater degree of joint military commitment to a partner than (0,0), but not a "more similar" joint policy with that partner.

possible allies, those states only account for approximately 12 percent of the region's military capabilities. Germany and Russia have substantially different alliance policies with states that account for about 67 percent of the region's capabilities. That Germany and Russia have such different alliance policies with respect to the most powerful nations is reflected in $S_c = -.45$.

4.3 Using Multiple Sources of Data on Revealed Foreign Policy: The United States and the European Major Powers, 1947

As we noted in section 2.3, even capability-weighting is not always sufficient to permit us to accurately infer similarity of states' foreign policy interests from their revealed alliance commitments. Sometimes, states with such common foreign policy interests that they are commonly seen as de facto allies do not actually share formal alliance ties. While Bueno de Mesquita (1975:194) is probably correct that such tacit alignments "are often recognized by signing a formal agreement to ally," it is not difficult to think of exceptions to that rule. Therefore, we suggested supplementing data on alliance commitments with additional types of foreign policy data in order to reveal common and conflicting interests that are imperfectly reflected in formal pledges.

We now turn to a fairly striking example of how relying only on alliance data can lead to faulty inferences about the similarity of states' foreign policy interests and how supplementing alliance data with UN voting data can help us get closer to the true similarity of interests. Table 7(a) displays the surprising pattern of formal alliance commitments among the European major powers in 1947. Britain and the Soviet Union signed a wartime alliance in July 1941 and extended their commitment for twenty years in the Anglo-Soviet Alliance of May 1942 (Ulam, 1974:318, 335; Werth, 1964:355). In 1944, DeGaulle secured a Franco-Soviet alliance in Moscow, hoping it would provide him with greater political independence from Britain and the U.S. (Werth, 1964:835). Thus, in the earliest days of the Cold War, as the relationships between Russia and the Western allies deteriorated due to disputes over Soviet policies in Eastern Europe, Britain and France held stronger formal alliance ties to the Soviet Union than they did to the United States. Simultaneously, however, Marshall Plan aid had begun to arrive in Western Europe and the Western allies managed to cooperate relatively effectively in occupied Germany. By the following year, the Brussels Treaty would pave the way for the NATO alliance. To the casual observer, unaware of the pattern of formal alliance commitments, France and Britain surely would have appeared closer to the U.S. than to the USSR in 1947.

Table 8(a) shows the capability-weighted similarity scores S_c for the four nations based on the alliance data alone. The similarity scores indicate that Britain and France have almost identical alliance policies ($S_c = .99$) and that the similarity of their alliance policies with the Soviet Union's is extremely high ($S_c = .87$ and $S_c = .88$, respectively). However, Great Britain's, France's, and Russia's alliance policies are quite divergent from those of the United States ($S_c = -.59$, $S_c = -.58$, and $S_c = -.7$, respectively). If we were to follow the standard practice of interpreting the similarity of states' alliance policies as an indication of the similarity of their interests, we would naturally conclude that in 1947, Britain, France, and the Soviet Union had very similar interests that stood quite opposed to those of the United States.

The alliance data, however, are not our only source of information on the similarity of Anglo-Franco-Soviet-American interests. By 1947, the young United Nations was already the scene of contentious debates among the powers over issues ranging from the redrawing of European borders to the management of atomic weapons. From its earliest days, the United Nations reflected the conflicts between the Soviets and the Western allies (Ulam, 1974:412–18). Table 7(b) shows the votes (1=no, 2=abstain, 3=yes) of the U.S., Britain, France, and the Soviet Union over

TABLE 7. Alliance and UN Vote Portfolios of European Major Powers and U.S., 1947

		Alliance Portfolios			
Nation	*System Cap*	*USA*	*UK*	*FRN*	*RUS*
USA	.42	3	0	0	0
RUS	.20	0	3	3	3
UK	.12	0	3	3	3
FRN	.05	0	3	3	3
POL	.04	0	0	0	3
ITA	.04	0	0	0	0
SPN	.02	0	0	0	0
CZE	.02	0	0	0	3
BEL	.01	0	0	0	0
NTH	.01	0	0	0	0
TUR	.01	0	0	0	0
SWD	.01	0	0	0	0
YUG	.01	0	0	0	3
RUM	.01	0	0	0	0
HUN	.01	0	0	0	0
POR	.00	0	3	0	0
SWZ	.00	0	0	0	0
GRC	.00	0	0	0	0
DEN	.00	0	0	0	0
NOR	.00	0	0	0	0
BUL	.00	0	0	0	0
LUX	.00	0	0	0	0
FIN	.00	0	0	0	0
IRE	.00	0	0	0	0
ALB	.00	0	0	0	0
ICE	.00	0	0	0	0

(a) Alliance portfolios

resolutions brought before the United Nations in 1947.[25] Even a cursory review of the UN voting portfolios reveals that Britain and France voted relatively similarly to the United States and in apparent opposition to the Soviet Union.

While the alliance data show that Britain and France held formal alliances with the Soviet Union in 1947, the UN voting data show that their announced positions on a variety of specific political issues were quite different from the Soviet positions. Given our knowledge of the events at that time, the UN voting data appear to be a "better" indicator of the immediate similarity of interests among the European major powers in 1947. Nevertheless, it may be of substantive importance that states still retain high levels of alliance commitments with each other rather than repudiating those commitments, even though they have significant disagreements on particular policy dimensions. Unless there is a specific theoretical requirement for it, rather than choosing one type of data or the other, we suggest enlarging the foreign policy space by combining the information from the two data sets.

When we supplement the alliance policy dimensions with the UN voting dimensions, the results reflect the influence of the more similar UN voting policies of the Western allies. Table 8(b) displays the similarity scores for the four nations, calculated by stacking the alliance portfolios and UN voting portfolios, weighting within

[25] The table does not show the actual resolution numbers. Rather, the resolutions are simply listed sequentially as they appear in the data.

TABLE 7. (continued)

	Vote Portfolios			
UN Case	*USA*	*UK*	*FRN*	*RUS*
1	3	3	3	1
2	3	3	3	1
3	3	3	3	1
4	1	1	1	3
5	1	1	1	3
6	1	1	1	3
7	3	3	3	1
8	3	2	3	3
9	3	1	3	3
10	1	1	1	3
11	3	3	3	1
12	3	3	3	1
13	1	1	1	3
14	1	1	1	3
15	2	3	1	1
16	3	1	3	3
17	3	3	3	1
18	3	3	3	1
19	3	3	3	1
20	3	3	3	1
21	3	3	3	1
22	3	3	3	1
23	3	3	3	3
24	1	3	3	3
25	3	3	3	3
26	1	1	3	3
27	3	3	1	1
28	3	3	3	3
29	3	1	3	1
30	3	1	3	1
31	3	2	3	2
32	3	2	3	3

(b) UN vote portfolios

(a) shows the alliance portfolios of the European major powers and U.S. in 1947. Britain, France, and the Soviet Union all have defense pacts with each other and no alliances with the U.S. (b) displays UN voting portfolios for the same four nations in 1947. In this case, Britain and France both appear to vote much more similarly to the U.S. than to the Soviet Union. (Alliance coding: 0=no alliance, 1=entente, 2=neutrality pact, 3=defense pact. UN vote coding: 1=no, 2=abstain, 3=yes.)

alliances by capabilities and within UN cases uniformly, and then weighting alliances versus UN votes uniformly. While Britain and France still appear to have very similar policies overall (S_c = .73), their policies are not nearly as similar to the Soviet Union's as before (now S_c = .19 and .3, respectively). And, while the combined policies are by no means overwhelmingly similar to those of the U.S., they have moved substantially in that direction (now S_c = –.01 and .1, respectively). These results are much more satisfying than the results reported in Table 8(a), both because they come closer to matching our intuitive understanding of the similarity of interests among the major powers at that time and because the method used to generate them makes better use of the different types of information that are available to us.

TABLE 8. S_c Similarity Scores Based Only on Alliance Data Versus Those Based on Both Alliance and UN Voting Data

	USA	*UK*	*FRN*
UK	–.59		
FRN	–.58	.99	
RUS	–.70	.87	.88

(a) Alliance data only

	USA	*UK*	*FRN*
UK	–.01		
FRN	.10	.73	
RUS	–.60	.19	.30

(b) Alliance and UN voting data

Alliance and UN vote portfolios are those displayed in Table 7. (a) shows the S_c scores calculated using only the alliance data. Based on these scores we infer that Britain's and France's interests were almost identical to the Soviet Union's and quite divergent from the U.S.'s. However, the UN voting patterns of Britain and France are much closer to the U.S.'s than to the Soviet Union's, which is reflected in the similarity scores shown in (b), where the alliance data has been supplemented with the UN voting data.

4.4. Evidence of Extensive Empirical Differences Between Sc and τ_b : The European Major Powers, 1816–1965

We feel that our analytical comparison of τ_b and S provides sufficient grounds to conclude that future studies of international politics should use S (or, properly, S_c) rather than τ_b to estimate the similarity of states' revealed policy positions: S_c measures revealed policy similarity more accurately than τ_b and it is easier to calculate than τ_b. In addition, the hypothetical and empirical examples provide clear if anecdotal evidence that the advantages of S over τ_b are of practical, rather than just theoretical, interest. We have not yet demonstrated, however, that τ_b and S_c, when applied to empirical alliance data, generate sufficiently different policy similarity values for many pairs of states over long periods of time so that we should be concerned about whether or not established research findings based upon τ_b should be reexamined. In this final set of comparisons, we seek to do exactly that.

One of the simplest methods for determining the extent of the differences in the τ_b and S_c measures in empirical applications is to plot the scores for particular dyads over a large period of time and to calculate the correlation between the similarity scores generated by the two different measures. Two such examples appear in Figures 1 and 2. Figure 1 plots the similarity of Britain's and Germany's alliance portfolios for every year from 1816 to 1945, while Figure 2 displays the similarity of France's and Germany's alliance portfolios.

As the two graphs show, S_c and τ_b draw very different pictures of the similarity of these states' alliance policies. The differences in the values of τ_b and S_c are not localized in specific time periods, and the magnitude of those differences is often quite substantial. Additionally, they are only moderately correlated (in these two examples, U.K.-Germany $\rho = .63$, France-Germany $\rho = .53$), indicating that the values of τ_b stray from those of S_c in tendency as well as in scale. For scholars testing systemic theories of international politics, it is particularly noteworthy that S_c is often a different sign than τ_b, as this is likely to affect the composition and characteristics of alliance "clusters."

Still, although this example shows the divergence between τ_b and S_c over nearly two hundred pairs of alliance portfolios, this represents a rather small subset of the alliance data. To get a better idea of the extent to which τ_b and S_c differ empirically, we have calculated τ_b and S_c for every pair of European major powers (plus the U.S.) for every year from 1816 to 1965. This region and time period involves seven major powers over 150 years, although not every state qualified as a major power in every year of that time period. Table 9(a) displays the number of years from 1816 to 1965 that each of the relevant states qualified as a major power according to Correlates of War standards. While Britain and France qualify as a major power dyad for 146 of those years, the U.S. and Austria-Hungary overlapped as major powers for only three years. Our analysis of the dyadic similarity of the alliance portfolios of European major powers includes a total of 1791 dyad-years. We follow the practice established by Bueno de Mesquita and followed by most other researchers of calculating policy similarity by comparing the major powers' alliance portfolios over all the states in the European region, including the U.S., for a given year.[26]

One indicator of the extent to which τ_b and S_c might lead to different substantive conclusions is the extent to which they are correlated over time for the major-power dyads. Table 9(b) displays these correlations.[27] As the table shows, the correlation between S_c and τ_b varies quite a bit, depending on the major-power dyad. Dyads that include the U.S. and either Britain, France, Austria-Hungary, or Russia all display a very high degree of positive correlation between the two similarity measures. On the other hand, dyads with Italy and either Britain, France, or Russia all show approximately no correlation. The remaining dyads fall somewhere in between, leading to an average correlation of S_c and τ_b of 0.63. While this is moderately positive, it is weak enough to warrant concern that empirical studies using τ_b may suffer from enough measurement error to lead to incorrect parameter estimates, especially since such studies generally employ these indicators not as explanatory variables but in complex operationalizations of explanatory variables (see, e.g., Bueno de Mesquita's (1985) measure of risk propensity).[28]

Finally, Table 9(c) displays the percentage of the time that τ_b and S_c are of the same sign, and the results here are also troubling. For the major powers we examine, τ_b and S_c take on different signs surprisingly often. Overall, τ_b and S_c have the same sign just over half the time. For scholars testing systemic theories of international politics, these sign differences are certain to affect the composition and characteristics of alliance "clusters," and they may very well also have significant implications for the operationalizations of choice variables.

5. Concluding Remarks

We have made two general claims. First, contrary to over twenty years' practice, Kendall's τ_b should not be interpreted as an indicator of the similarity of states' alliance policies. τ_b measures the association of two alliance portfolios interpreted as rankings—in fact, measures it quite well. However, this is not at all the same as measuring the similarity of the alliance portfolios. We have demonstrated analytically that τ_b is inappropriate for this task and have provided hypothetical and empirical examples that illustrate some of the problems. Rather than simply identifying the

[26] Since there are only 120 UN-relevant dyad-years in this sample and previous studies using τ_b did not make use of the UN data, we omit the UN data in the interest of keeping the comparison as straightforward as possible.

[27] The dyad-year correlations of τ_b and S_u are roughly similar to that shown in Table 9(b). These results are available from the authors.

[28] In a multiple regression context, if a single explanatory variable suffers from measurement error, its coefficient is biased toward zero. However, the coefficients of the *other* variables are biased in unknown directions. The problem is worse if multiple variables are measured with error. See Greene, 1993:283–84.

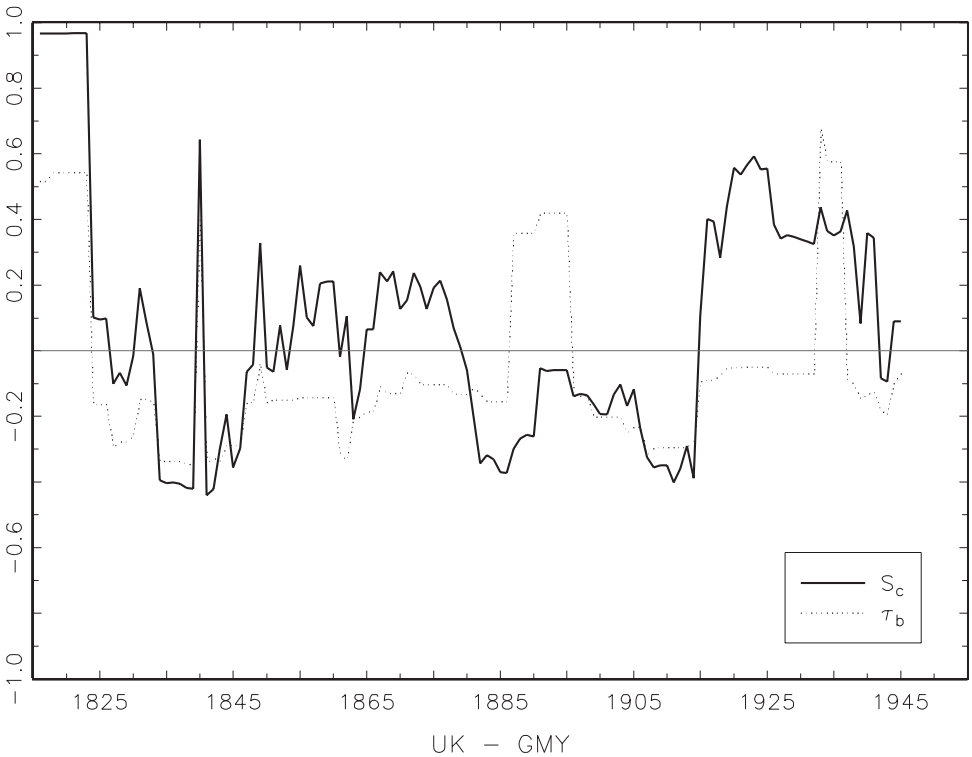

FIG. 1. Comparison of S_c and τ_b for Britain and Germany, 1816–1945. *The similarity of Britain and Germany's alliance portfolios is plotted from 1816 to 1945 for both S_c and τ_b. As the graph shows, there is often empirically substantial difference between the two measures of similarity. The correlation between S_c and τ_b for the UK–GMY case is $\rho = .63$.*

shortcomings of τ_b in this context, we have also developed a spatial measure of policy position similarity, S, which we believe more directly and accurately captures this concept. In addition, using data on European alliances, we have shown that S produces substantively different results than τ_b for measures of alliance portfolio similarity—not just for a few select cases, but over a broad period of time.

Second, we have claimed that inferring states' interests from their alliance commitments is problematic, even given a perfect measure of portfolio similarity. However, we believe our method in combination with other sources of data (e.g., on UN votes, diplomatic missions, trade, and disputes) can provide leverage by enlarging the foreign policy space over which comparisons are made. The example of the 1947 European major powers and the United States amply shows both the problem of inferring similarity of foreign policies from alliance data alone and the benefit of supplementing alliance data with UN voting data to obtain a more accurate measure of the similarity of the states' foreign policy positions.

We suspect, however, that S_c may not be the final word in this matter. In the process of reopening the issue of how to measure the similarity of states' revealed foreign policy positions, we have identified at least two promising avenues for future research. First, most students of international politics would agree that alliance commitments result from complex forms of strategic interaction between states. This strategic interaction may lead states to adopt alliance commitments that reflect not their most preferred outcomes, but the best possible outcomes "in equilibrium,"

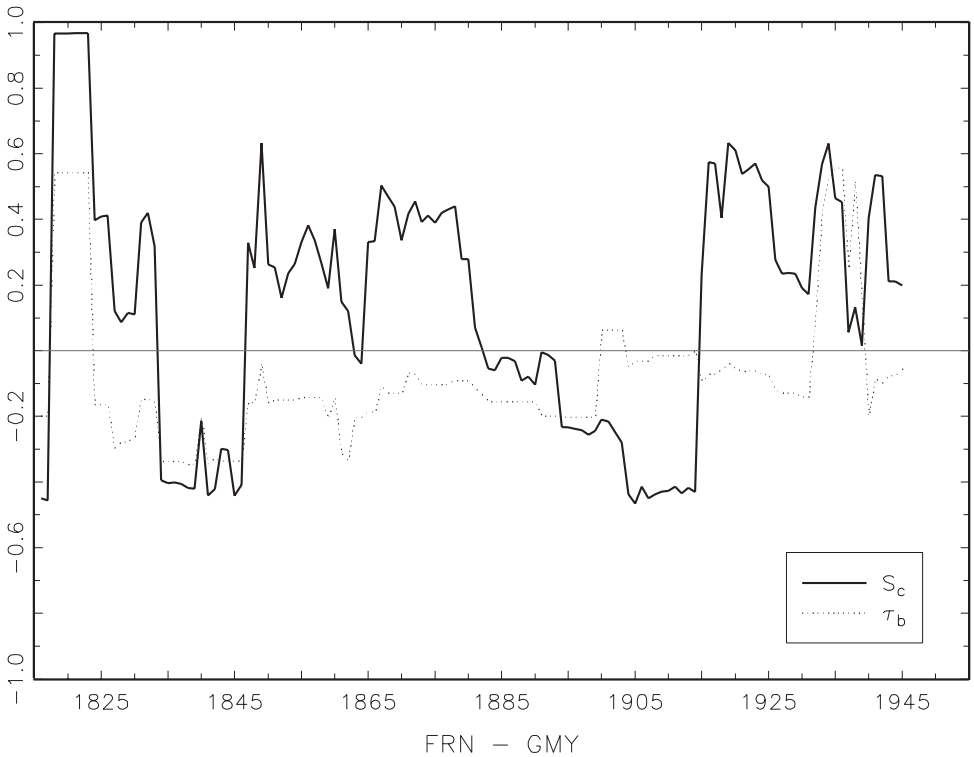

FIG. 2. Comparison of S_c and τ_b for France and Germany, 1816–1945. *The similarity of France and Germany's alliance portfolios is plotted from 1816 to 1945 for both* S_c and τ_b. *The graph shows there is empirically even greater difference between the two measures of similarity in this case. The correlation between* S_c and τ_b *for the FRN–GMY case is* $\rho = .53$.

given the structure of the strategic interaction.[29] There are numerous ways strategic interaction and "budget" constraints can affect the formation of alliances. If alliances are politically costly, states may "underconsume" alliances with demanding partners relative to their ideal level of consumption. States may have incentives to send misleading signals by adopting alliances with states that do not share their interests. In some cases, states with very similar interests may not need to sign formal alliance agreements. Alliances may be used to restrain potential adversaries or to reduce the set of possible alliance partners available to an adversary. And, to the extent that alliance decisions reflect security interests, they can do so only imperfectly by virtue of the fact that the set of potential alliance partners is not continuous. Ideally, we would like a coherent theory of alliance formation that is able to relate a state's observed alliance commitments back to its underlying preferences. Clearly, the development of such a theory is an important avenue of future research. Without such a theory, political scientists will have to continue to assume that revealed policy positions represent some approximation of states' actual interests, despite having reasons to think that this may not be the case.

Second, the available data on states' alliance commitments were not designed for use as indicators of states' common interests; they have been adopted for this purpose due to the absence of better alternatives. We feel the available alliance data

[29] We are, of course, not the first to note these issues. See, for example, Bueno de Mesquita, 1981a:112–14, Bueno de Mesquita and Lalman, 1992, Niou and Ordeshook, 1994, Smith, 1995, and Gartzke and Simon, 1996.

TABLE 9. Comparison of S_c and τ_b for the U.S. and the European Major Powers from 1816 to 1965

	USA	*UK*	*FRN*	*GMY*	*AUH*	*ITA*
UK	50					
FRN	46	146				
GMY	24	124	120			
AUH	3	103	103	103		
ITA	28	84	81	78	59	
RUS	46	146	142	123	102	80

(a) Major power dyad-years. Total = 1791

	USA	*UK*	*FRN*	*GMY*	*AUH*	*ITA*
UK	.92					
FRN	.93	.84				
GMY	.63	.63	.53			
AUH	1	.72	.47	.71		
ITA	.77	.06	.07	.80	.75	
RUS	.91	.84	.86	.64	.60	–.04

(b) Correlation. Weighted average = .63

	USA	*UK*	*FRN*	*GMY*	*AUH*	*ITA*
UK	50					
FRN	61	54				
GMY	42	53	49			
AUH	0	70	40	89		
ITA	39	48	35	56	56	
RUS	85	60	52	59	52	39

(c) Percent same sign. Weighted average = 55%

(a) displays the number of years for which τ_b and S_c could be calculated for the row and column dyad. (b) and (c) show that the correlation and similarity in sign of S_c and τ_b vary considerably over the different dyads.

might be used as the basis for a revised and expanded data set that more accurately reflects states' announced security policy positions. For example, the coding rules for the alliances data set explicitly excluded unilateral security guarantees like the 1960 U.S.-Japan Security Treaty and certain multilateral asymmetric treaties like the Locarno Agreement of 1925 (Singer and Small, 1966:135–36). While it may be appropriate to exclude some types of security treaties from studies of international alliances, it seems perfectly reasonable to take them into account when measuring states' policy positions. In fact, just as our similarity measures vary from –1 to 1, policy data might vary from observable signs of active animosity to signs of joint security interests. By making use of data on sanctions, embargoes, severed diplomatic relations, and the like, it may be possible to reduce or eliminate the analytical problems currently posed by the "no alliance" classification.

Nevertheless, even given these limitations, *S* will lead to more accurate measurement of the similarity of foreign policy positions, and we believe that the ability to incorporate data from multiple sources makes *S* an important step toward better measures of the similarity of states' interests. We hope that our efforts here will contribute to the improvement of several very valuable and interesting research programs.

References

ALTFELD, M. F. (1984) The Decision to Ally: A Theory and Test. *Western Political Quarterly* **37**(4):523–544.

ALTFELD, M. F., AND B. BUENO DE MESQUITA (1979) Choosing Sides in Wars. *International Studies Quarterly* **23**(2):87–112.

ALTFELD, M. T., AND W. K. PAIK (1986) Realignment in International Treaty Organizations: A Closer Look. *International Studies Quarterly* **20**:107–114.

BERKOWITZ, B. D. (1983) Realignment in International Treaty Organizations. *International Studies Quarterly* **27**(1):77–96.

BERRY, K. J., AND P. W. MIELKE, JR. (1988) A Generalization of Cohen's Kappa Agreement Measure to Interval Measurement and Multiple Raters. *Educational and Psychological Measurement* **48**(4):921–933.

BERRY, K. J., AND P. W. MIELKE, JR. (1990) A Generalized Agreement Measure. *Educational and Psychological Measurement* **50**(1):123–125.

BISHOP, Y. M. M., S. E. FIENBERG, AND P. W. HOLLAND (1975) *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

BUENO DE MESQUITA, B. (1975) Measuring Systemic Polarity. *Journal of Conflict Resolution* **19**(2):187–216.

BUENO DE MESQUITA, B. (1978) Systemic Polarization and the Occurrence and Duration of War. *Journal of Conflict Resolution* **22**(2):241–267.

BUENO DE MESQUITA, B. (1980) An Expected Utility Theory of International Conflict. *American Political Science Review* **74**(4):917–931.

BUENO DE MESQUITA, B. (1981a) *The War Trap*. New Haven, CT: Yale University Press.

BUENO DE MESQUITA, B. (1981b) Risk, Power Distributions, and the Likelihood of War. *International Studies Quarterly* **25**(4):541–568.

BUENO DE MESQUITA, B. (1984a) Theory and the Advancement of Knowledge About War: A Reply. *Review of International Studies* **10**(1):65–75.

BUENO DE MESQUITA, B. (1984b) A Critique of "A Critique of the War Trap." *Journal of Conflict Resolution* **28**(2):341–360.

BUENO DE MESQUITA, B. (1985) The War Trap Revisited: A Revised Expected Utility Model. *American Political Science Review* **79**(1):156–176.

BUENO DE MESQUITA, B. (1987) Conceptualizing War: A Reply. *Journal of Conflict Resolution* **31**(2):370–382.

BUENO DE MESQUITA, B., AND D. LALMAN (1986) Reason and War. *American Political Science Review* **80**:113–131.

BUENO DE MESQUITA, B., AND D. LALMAN (1988) Empirical Support for Systemic and Dyadic Explanations of International Conflict. *World Politics* **41**(1):1–20.

BUENO DE MESQUITA, B., AND D. LALMAN (1992) *War and Reason: Domestic and International Imperatives*. New Haven, CT: Yale University Press.

COHEN, J. (1960) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* **20**:37–46.

COHEN, J. (1968) Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit. *Psychological Bulletin* **70**(4):213–220.

CONYBEARE, J. A. C. (1992) A Portfolio Diversification Model of Alliances: The Triple Alliance and Triple Entente, 1879–1914. *Journal of Conflict Resolution* **36**:53–85.

DAVIES, M., AND J. L. FLEISS (1982) Measuring Agreement for Multinomial Data. *Biometrics* **38**:1047–1051.

FEARON, J. D. (1997) Signaling Foreign Policy Interests. *Journal of Conflict Resolution* **41**(1):68–90.

GARTZKE, E., AND M. W. SIMON (1996) Contracts Between Friends? Counterintuitive Determinants of Alliance Formation. Paper presented at the Annual Meeting of the International Studies Association (Midwest).

GREENE, W. H. (1993) *Econometric Analysis*, 2nd ed. New York: Macmillan.

HUTH, P., D. S. BENNETT, AND C. GELPI (1993) System Uncertainty, Risk Propensity, and International Conflict Among the Great Powers. *Journal of Conflict Resolution* **36**(3):478–517.

IUSI-SCARBOROUGH, G. (1988) Polarity, Power and Risk in International Disputes. *Journal of Conflict Resolution* **32**(3):511–533.

KENDALL, M. G., AND J. D. GIBBONS (1990) *Rank Correlation Methods*, 5th ed. London: Edward Arnold.

KENDALL, M. G., AND A. STUART (1961) The Advanced Theory of Statistics, vol. 2. London: Charles Griffin.

KHONG, Y. F. (1984) War and International Theory: A Commentary on the State of the Art. *Review of International Studies* **10**(1):41–63.

KIM, C. H. (1991) Third-Party Participation in Wars. *Journal of Conflict Resolution* **35**(4):659–677.

KIM, W. (1989) Power, Alliance, and Major Wars, 1816–1975. *Journal of Conflict Resolution* **33**(2):255–273.

KIM, W. (1991) Alliance Transitions and Great Power War. *American Journal of Political Science* **35**(4):833–850.

KIM, W., AND J. D. MORROW (1992) When Do Power Shifts Lead to War? *American Journal of Political Science* **36**(4):896–922.

KOTZ, S., AND N. L. JOHNSON (1988) *Encyclopedia of Statistical Sciences*. New York: John Wiley.

LALMAN, D. (1988) Conflict Resolution and Peace. *American Journal of Political Science* **32**:590–613.

LALMAN, D., AND D. NEWMAN (1991) Alliance Formation and National Security. *International Interactions* **16**(4):239–253.

LEVY, J. S. (1981) Alliance Formation and War Behavior. *Journal of Conflict Resolution* **25**:581–614.

LEVY, J. S. (1989) "The Causes of War: A Review of Theories and Evidence." In *Behavior, Society, and Nuclear War*, edited by P. E. Tetlock, J. L. Husbands, R. Jervis, P. C. Stern, and C. Tilly. New York: Oxford University Press.

LIEBETRAU, A. M. (1983) Measures of Association. Newbury Park, CA: Sage.

LISKA, G. (1962) *Nations in Alliance: The Limits of Interdependence*. Baltimore, MD: Johns Hopkins University Press.

MAJESKI, S. J., AND D. J. SYLVAN (1984) Simple Choices and Complex Calculations: A Critique of the War Trap. *Journal of Conflict Resolution* **28**(2):316–340.

NICHOLSON, M. (1987) The Conceptual Bases of the War Trap. *Journal of Conflict Resolution* **31**(2):346–369.

NIOU, E. M. S., AND P. C. ORDESHOOK (1994) Alliances in Anarchic International Systems. *International Studies Quarterly* **38**(2):167–192.

O'CONNELL, D. L., AND A. J. DOBSON (1984) General Observer-Agreement Measures on Individual Subjects and Groups of Subjects. *Biometrics* **40**:973–983.

ONEAL, J. R., AND B. RUSSETT (1997) Escaping the War Trap: Evaluating the Liberal Peace Controlling for the Expected Utility of Conflict. Mimeo.

ORDESHOOK, P. C. (1993) *Game Theory and Political Theory*. Cambridge: Cambridge University Press.

ORGANSKI, A. F. K., AND J. KUGLER (1980) *The War Ledger*. Chicago: University of Chicago Press.

OSTROM, C. W., JR., AND J. H. ALDRICH (1978) The Relationship Between Size and Stability in the Major Power International System. *American Journal of Political Science* **22**(4):743–771.

PROTTER, M. H., AND C. B. MORREY, JR. (1991) *A First Course in Real Analysis*, 2nd ed. New York: Springer-Verlag.

SABROSKY, A. N. (1980) "Interstate Alliances: Their Reliability and the Expansion of War." In *The Correlates of War II: Testing Some Realpolitik Models*, edited by J. D. Singer. New York: Free Press.

SCHOUTEN, H. J. A. (1982) Measuring Pairwise Interobserver Agreement When All Subjects Are Judged by the Same Observers. *Statistica Neerlandica* **36**(2):45–61.

SINGER, J. D., AND M. SMALL (1966) Formal Alliances, 1815–1939. *Journal of Peace Research* **1**:1–32.

SMALL, M., AND J. D. SINGER (1982) *Resort to Arms: International and Civil Wars, 1816–1980*. Newbury Park, CA: Sage.

SMITH, A. (1995) Alliance Formation and War. *International Studies Quarterly* **39**:405–425.

STOLL, R. J. (1984) Bloc Concentration and the Balance of Power. *Journal of Conflict Resolution* **28**(1):25–50.

STOLL, R. J., AND M. CHAMPION (1985) "Capability Concentration, Alliance Bonding, and Conflict Among the Major Powers." In *Polarity and War: The Changing Structure of International Conflict*, edited by A. N. Sabrosky. Boulder, CO: Westview Press.

ULAM, A. B. (1974) *Expansion and Coexistence: Soviet Foreign Policy 1917–1973*. Fort Worth, TX: Holt, Rinehart and Winston.

WAGNER, R. H. (1984) War and Expected Utility Theory. *World Politics* **36**(3):407–423.

WALLACE, M. (1973) Polarization, Cross-Cutting, and International War, 1815–1964. *Journal of Conflict Resolution* **17**(4):575–604.

WALLACE, M. (1985) "Polarization: Towards a Scientific Conception." In *Polarity and War: The Changing Structure of International Conflict*, edited by A. N. Sabrosky. Boulder, CO: Westview Press.

WERTH, A. (1964) *Russia at War: 1941–1945*. New York: Avon.