DISPUTES, DEMOCRACIES, AND DEPENDENCIES: A RE-EXAMINATION OF THE KANTIAN PEACE

MICHAEL D. WARD, RANDOLPH M. SIVERSON, AND XUN CAO

Abstract. Militarized interstate disputes are widely thought to be less likely among democratic countries that have high levels of trade and extensive participation in international organizations. We re-examine this broad finding of the Kantian peace literature in the context of a model that incorporates the high degree of dependency among countries. Based on in-sample statistical tests, as well as out-of-sample, predictive cross-validation, we find that results frequently cited in the literature are plagued by over-fitting and can not be characterized as identifying the underlying structure through which international conflict is influenced by democracy, trade, and international governmental organizations. We conclude that much of the statistical association typically reported in this literature apparently stems from three components: a) geographical proximity, b) dependence among militarized interstate disputes with the same initiator or target, and c) the higher-order dependencies in these dyadic data. Once these are incorporated, covariates associated with the Kantian peace tripod lose most of their statistical power. We do find that higher levels of joint democracy are associated with lower probabilities of militarized interstate dispute involvement. We find that despite high statistical significance and putative substantive importance, none of the variables representing the Kantian tripod is associated with any substantial degree of predictive power.

Kantian Peace

Are democracies less likely to have conflict with one another than other pairs of states? Yes, according to a growing body of research known as the "Democratic Peace." This literature has evolved from exploring whether pairs of democratic countries have a reduced probability of war with one another (Maoz & Abdolali 1989) into a continued and broader examination of the covariates of militarized interstate disputes.[1] Most empirical scholarly literature on this topic uses a regression perspective, identifying those variables and factors thought to be important in explaining the occurrence of militarized interstate disputes (MIDs), assessing by statistical significance the power of various hypotheses concerning the democratic peace.[2]

Authors' Note: Version of February 14, 2007. Ward's research was supported by grants from the Methods, Measurement, and Statistics Program at the National Science Foundation, numbers: SES-0417559 and SES-0631531. This research greatly benefitted from many helpful suggestions from Kristin M. Bakke, Nathaniel Beck, Bear F. Braumoeller, Bruce Bueno de Mesquita, John R. Freeman, Jeff Gill, Kristian Skrede Gleditsch, Peter D. Hoff, Gary King, Aseem Prakash, Adrian Raftery, Dennis L. Ward, Anton Westveld, and Erik Wibbels. David Callaway and John E. Daniels of the Social Science Data Service at the Institute of Governmental Affairs at the University of California, Davis were especially helpful to us at a critical point in our quest for distributed computing resources, which we found at the San Diego Supercomputer Center at the University of California, San Diego. We thank Chris Adolph for sharing his R code for "ropeladder" plots. Finally, we thank reviewers for helpful and invigorating comments.

[1]Summaries of the literature on the democratic peace include Chan (1997), Ray (1998), and Morrow (2002). An incomplete list of articles on this topic would include Rummel (1983, 1985), Doyle (1983*a*, 1983*b*, 1986), Lake (1992), Maoz & Russett (1993), Farber & Gowa (1995), Ray (1995), Maoz (1996), Rousseau, Gelpi, Reiter, & Huth (1996), Farber & Gowa (1997), Gleditsch & Hegre (1997), Mansfield & Snyder (1997), Mousseau (1997), Oneal & Russett (1997), Raknerud & Hegre (1997), Thompson & Tucker (1997), Gartzke (1998), Maoz (1998), Russett, Oneal & Davis (1998), Cederman (2001), Russett & Oneal (2001), Schultz (2001), Huth & Allee (2002), Dixon & Senese (2002), Russett, Oneal & Berbaum (2003), Goertz, Jones & Diehl (2005), among many others.

[2]For example, see recent work by Leeds (2003), Gartzke (2003), and Lemke & Reed (2001). Cederman (2001) and Cederman & Rao (2001) are examples of different approaches. More recently, several scholars have undertaken a "logical" examination of the Kantian peace (Rosato 2003, Rosato 2005, Slantchev, Alexandrova, & Gartzke 2005, Doyle 2005).

Within this larger research program, one particularly prominent line of inquiry has explored the hypothesis that pairs of democracies that trade with each other and belong to and participate in the same international organizations will have reduced probabilities of being involved in militarized disputes with one another. "The Kantian Peace" (Oneal & Russett 1999) exemplifies this line of research on the democratic peace. Like many similar studies, it uses a general linear model to estimate the occurrence of militarized interstate disputes over the period 1886 to 1992. Many other studies have employed these or similar data along with various perturbations of the specified regression equation. The basic insight of this line of research has been that democracy, international commerce, and participation in international organizations are each shown to be independently pacifying forces in world politics. Oneal and Russett (1999, page 33) summarize their work by noting that "... analyses for the years 1885 to 1992 indicate that Kant was substantially correct: democracy, economic interdependence, and involvement in international organizations reduce the incidence of militarized interstate disputes." Similar declarations are by now legion in scholarly writing and are also frequently found in the policy pronouncements of the governments of industrialized democracies.[3]

In this article, we report results that cast doubt on whether the literature has accurately captured the mechanism that is thought to embody the Kantian peace. We do this not by a "battle of the stars", comparing the significance of our variables to those reported in various other studies, but by examining the ability of standard kinds of approaches to stand up to in- and out-of-sample validation. Thus, we do not question the statistical significance of any published work on the Kantian peace, nor do we question the logic or the theoretical bases guiding these studies. Rather, we use an example to illustrate that such statistical models may not fare well in predicting conflicts, either in or out-of-sample.

Why is this important? It might be that the conventional models are correctly specified, but that there is simply not much systematic variation to be captured. Alternatively, it might be that such models are missing a great deal of systematic variation because they ignore important features of the data generating process in international relations.

Beck, King & Zeng (2000) illustrate that large-N studies, such as those found in the Kantian peace literature and more broadly the empirical literature on international conflict, can be improved via the use of an out-of-sample forecasting heuristic. They note (p. 21) that "statistical analysts must be concerned about whether they are taking advantage of some idiosyncratic features of the data to improve fit at the expense of detecting structure" and assert that out-of-sample forecast are considered the "gold standard" in terms of assessing model fit. In a subsequent paper, de Marchi, Gelpi & Grynaviski (2004) challenge some claims in Beck, King & Zeng (2000), but support the use of out-of-sample forecasting as a important heuristic, a point also made in Beck, King & Zeng (2004). However, this is not yet among the standard practices in the international relations literature, and few scholars have made predictive out-of-sample assessments of their models.

The conjecture of Beck, King & Zeng (2000) implicitly raises the question of how well the results of the Kantian Peace capture the existence of militarized interstate disputes. Using the data in Oneal & Russett (1999) as an example from the literature, we replicated (exactly) their main empirical results (i.e., Table 1 therein, which analyzes data from 1886 to 1992). Using two plausible cut-points, Table 1 illustrates the *in-sample* predictive power of one version of the Kantian peace (Oneal & Russett 1999). As is clear, this estimated model is not very accurate in identifying disputes, either predicting none at all, or having a high rate of false positives.[4] This buttresses another finding of the study by Beck, King & Zeng (2000): namely, standard approaches will tend to over-predict conflicts among dyads with low probabilities of conflict, but is likely to under-predict conflict among dyads with high prior probabilities of conflicts.

The second part of the table uses a cutpoint of 0.009 and is open to two drastically different interpretations. On the one hand, for every twenty-three dyads that the model predicts to be in a dispute, only one turns out to be an actual dispute, while the other twenty-two (approximately) are not in a dispute. On the other hand, cases that are predicted to have disputes are about 23 times more likely to have disputes than are cases predicted to be absent

[3]See also important, recent work on this topic in Oneal & Russett (2005).

[4]Indeed, the residuals of the estimated model produces perfect results, for almost any cut point. The residuals actually contain all of the correct information about the existence of dyadic, militarized interstate disputes. The converse is that the originally estimated canonical model of the Kantian peace contains virtually no information about the dependent variable. Hence, better in-sample predictions are produced by the residuals than are produced by the model.

TABLE 1. *Using the basic results from the Oneal and Russett (1999) model (Table 1, column 1) it is impossible to predict accurately the occurrence and non-occurrence of dyadic, militarized interstate disputes over the period from 1886 to 1992. In both choices of cut-points, the model has difficulty in distinguishing between disputes and the absence of a dispute.*

Cut-point: 0.50	Non-Disputes	Disputes
Predicted Non-Disputes	147,989	1,415
Predicted Disputes	none	none
Cut-point: 0.009		
Predicted Non-Disputes	125,502	274
Predicted Disputes	22,487	1,141

disputes.[5] This Rashomon-like effect is partially dictated by the very large number of dyads in the study not involved in disputes, thus yielding very low ratios whenever the non-conflicts are included in the calculations.

What is the nature of the disputes that are accurately predicted by this model of the Kantian peace? Major powers are frequent members in the set of countries that are correctly predicted to be involved in a militarized dispute. The United States is most frequently correctly identified as being in a dispute (some 151 times), followed by the Soviet Union (139), the U.K. (125), China (76), and France (70). The most frequently correctly predicted dyads are also unsurprising: the Soviet Union-Japan (35 instances), the U.S.-Soviet Union (26), India-Pakistan (27), Greece-Turkey (24), Soviet Union-China (23), and the U.K.-Soviet Union (21). The years immediately preceding World War I and II account for over 20% of the correct predictions. Recent predictions are also better, with an annual average of 19 disputes being correctly predicted in the post-1938 period, compared to 8 in the pre-1939 era. In short, this model produces a large number of false positive predictions and illustrates that major and regional powers are more likely to be involved in disputes. The estimates support an overconfidence in the results based on statistical significance that is not borne out in the accuracy of predictions.

Our approach builds on the standard covariates employed in the Kantian peace, but examines them in the context of reciprocity, transitivity, balance, and clusterability in the network of interactions that comprise international politics. Rather than dealing with each action between countries as being independent, our modeling strategy explicitly embraces the idea that some countries will initiate lots of conflict, and other are likely to be recipients of conflict. As a result, two conflicts involving a country that is often involved in militarized disputes may not be completely independent. The same is true for even a single conflict. Consider for example, the 1991 Gulf War that began with the Iraqi invasion of Kuwait in the context of alleged Kuwaiti transgressions in the oil fields along their border. By the time of the counter-invasion in January 1991, the United States and a coalition of more than two dozen other independent countries joined together in this conflict.[6] This raises the question, for example, of how likely it was that the U.S. and Britain would both be members of this coalition, given their patterns of interactions and prior cooperation in the Middle East? Many dyadic studies consider each of these countries' interactions with Iraq to be an independent event, i.e., the instance of a militarized interstate dispute. Thus, a MID would exist in the U.S.-Iraq dyad, as well as the U.K.-Iraq dyad, and between each of the other members of the coalition and Iraq. When combined in a large-N study, these dyads would be considered to be independent events in most regression based approaches. Yet, these events were not independent but were predicated on having a single target, and having been organized by the U.S. and others in the several months leading up to the actual counter-invasion of Iraq. The history of international disputes, and consequently the extant data on militarized interstate disputes, is replete with these kinds of dependencies. We formally incorporate and estimate the extent of these kinds of dependencies in our model of the Kantian peace in order to more precisely determine the effects of the Kantian tripod on international conflict.

Our findings will demonstrate that the dependencies are powerful, overwhelming the impact of covariates typically associated with reduced probabilities of interstate conflict: joint democracy, high levels of trade, and

[5]The odds of having a dispute in the set of dyads predicted to not have a dispute is $\frac{274}{125502} = 0.00218$, while the odds of a dyad that is predicted to have a dispute actually having one is $\frac{1141}{22487} = 0.0507$. The ratio of these two odds is approximately 23.

[6]Afghanistan, Argentina, Australia, Bahrain, Bangladesh, Canada, Czechoslovakia, Denmark, Egypt, France, Germany, Greece, Hungary, Honduras, Italy, Kuwait, Morocco, The Netherlands, Niger, Norway, Oman, Pakistan, Poland, Portugal, Qatar, Saudi Arabia, Senegal, South Korea, Spain, Syria, Turkey, The United Arab Emirates, the United Kingdom and the United States.

participation in international organizations. Using out-of-sample predictive tests, our framework for modeling these dependencies dominates the standard approach offered in the bulk of the empirical literature in international relations. In the next section we offer a model that focuses on the frequently ignored dependencies that characterize international politics in order to offer an improved model of the Kantian peace.

A Bayesian, Hierarchical, Bilinear, Mixed Effects Model

Dyadic data in international relations are rife with dependencies. For example, because states have relatively stable policies it is reasonable to expect that data emanating from a single country are likely to be correlated, as are data directed to a single country by others. Similarly, dyads are likely to have data correlated over time. The social relations model introduced a method to decompose variance in such data into sender and receiver effects as well as permit within-dyad correlations via the analysis of variance (ANOVA) protocol (Warner, Kenny & Stoto 1979, Wong 1982).[7]

The social relations model (SRM) grows out of interest by psychologists to separate the independent and interactive effects of groups versus individuals. The basic insight provides a way of conceptualizing and acquiring *unbiased* estimates of actor, partner, and relationship effects and their interrelationships. The social relations model emerged in the study of family relations as systems of interacting, adaptive relationships. Strong, affectionate relations among the various dyads in a family, for example, have been shown to have powerful effects on the developmental health of children in that family. There was a dissatisfaction in the field of psychology with studying these dynamics using conceptual frameworks and methods that assumed all the objects of study were independent of one another and that the appropriate locus of inquiry was *either* the individual *or* the dyad or triad. Kenny (1985) introduced the social relations model in part to focus attention on the analysis of dyads as dependent phenomena. The development of the social relations model is presented first by David Kenny (1981), but the canonical citation is from the mid-1980s (Malloy & Kenny 1986). A widely cited volume was penned on this topic in the mid-1990s (Kenny 1994), and recently a book devoted to data analytic approaches to the SRM was published (Kenny, Kashy & Cook 2006). Literally hundreds of articles in the field of psychology use this widely appreciated approach to confronting the dependencies in dyadic data. In short, by separating actor, partner, relationship and reciprocity effects from one another, this model focuses on, rather than ignores, the interdependencies among actors in a family, or other systems of highly interconnected and dependent actors such as international politics.

The idea of further decomposing the variance in the context of dyadic data was developed (Gill & Swartz 2001, Li & Loken 2002) to permit the statistical analysis of normally distributed dyadic data using additive effects. These ideas have been extended to a generalized linear model that incorporates third-order dependence via a bi-linear effect (Hoff 2005) similar in spirit to that first introduced by Gabriel (1998).

There is strong interdependence among the observations in many international relations data sets. A first-order dependency is the inclination for states to behave toward others in a consistent manner, or, alternatively, for a state to be the object of consistent policies from others. Often this is captured in a time-series context, though below we offer a different conceptualization.[8] Consider, for example, the behavior of the United States toward Cuba in the years since Fidel Castro announced he was a Marxist-Leninist. The range of variation is narrow. By the same token, for an extended period Libya was the object of a consistent set of policies by a relatively large number of other states. The existence of first-order dependencies is widely recognized, and many forms of time series analysis take them into account, including, of course, pooled cross-sectional methods. The problem, as alluded to above, is the the dependencies do not stop with the first-order, but become more complex, and these have not thus far been accommodated, or even generally recognized, in most analyses of international relationships. Controlling for time dependence does not eliminate other higher-order dependencies.[9]

When analysis moves away from the policies of individual states over time to embrace other complicated sets of relationships, higher-order dependencies are likely to become rife in the data, as we have argued above. But

[7]In the context of transactions, the language of sender and receiver is clear. However, in the context of conflict, it is more complicated, since many disputes are protracted and while initiators of episodes can often be specified, it is over-simplistic to assign prime mover status to that country. In what follows, we use sender and receiver to represent different sides of a dyadic relationship. Every country is a potential sender and a potential receiver. This sociomatrix is square having all countries arrayed along the columns and rows. A country is the sender in a particular militarized dispute if it is among that set of countries identified to have first taken militarized action, according to the MID data.

[8]There is a long tradition of looking at these kind of dependencies in pairs of countries, and even among triads. Much of this literature began with the insights of Richardson (1960) and continued with disaggregated events data (Goldstein & Freeman 1990).

[9]Raknerud & Hegre (1997) use time dependence on common trends to induce mean independence among dyads.

what does "higher-order dependencies" mean? Second-order dependence refers to what is often described as reciprocity in the context of directed relationships. The study of arms races is one area in which this has been frequently investigated in a single pair of countries. In the years before World War I, did the French military budget depend in some significant part on what the French believed the Germans would spend? Similarly, did the Germans gear their spending on what they anticipated the French would do? Given the obvious answers to these questions, is it possible to consider the arms acquisition policies of a state in isolation from what other salient states are doing? Interactions in international conflicts also often evince similar reciprocity at the dyadic level. For instance, India's aggressive behavior towards Pakistan is likely to induce similar behavior from Pakistan. Many pairs of countries show evidence of this type of reciprocity, both in the conflictual as well as cooperative realms, even though reciprocity can still be asymmetric such that one actor in the dyad responds more strongly, or weakly, than the other. The prevalence of reciprocity—a second-order dependence among observations—in directional network data in the study of international relations challenges the basic assumption of observational independence. It can neither be ignored nor assumed away, and therefore needs to be explicitly modeled.

To discuss fully third-order dependence, some formalism is helpful. We define y_{ij} as some measure of the relationship between actors i and j. This may be a measure of the flow between actors, as in the case of data on international commerce or even a measure of whether a "linkage" exists between the actors. If the value of y_{ij} is identical to y_{ji} for all ij pairs, the data are said to be undirected (and symmetric). Third-order dependence includes (a) transitivity, (b) balance, and (c) clusterability (Wasserman & Faust 1994). *Transitivity* follows the familiar logic of "a friend of a friend is a friend." In particular, for directed binary data, triad ijk is transitive if whenever $y_{ij} = 1$ and $y_{jk} = 1$, we also observe that $y_{ik} = 1$. A triad ijk is said to be *balanced* if all pairs of actors relate to one another in an identical fashion, specifically: $y_{ij} \times y_{jk} \times y_{ki} > 0$. The idea is that if the relationship between i and j is "positive" then both will relate to another unit k identically. If y_{ij} is positive, then to observe balance, y_{jk} and y_{ki} are either both positive or both negative. *Clusterability* is a relaxation of the concept of balance. A triad is clusterable if it is balanced or the relations are all negative. The idea is that a clusterable triad can be divided into groups where the measurements are positive within groups and negative between groups.

The notion that "a friend of a friend is a friend" captures the concept of *transitivity* in directed binary network data setting and is illustrated in Figure 1. Consider a triad of countries $\{ijk\}$. This triad is composed of three dyads $\{ij\}$, $\{jk\}$, and $\{ki\}$. In a network context, there are six possible links if the data are directional and three if data are non-directional. If we know that country i considers j as an ally and country j is allied with k, then the probability that k will also be allied with i is likely to be higher than for another country outside of this network, since these countries are at least indirectly connected in the alliance network by virtue of their separate linkages to country j. In other words, knowing something about the relationship between countries in the first two dyads in a triad often tells us something about the relations in the third dyad. Treating dyads $\{ij\}$, $\{jk\}$, and $\{ki\}$ as independent from each other, as the way we routinely do in empirical analyses of international relations data, usually ignores important patterns in these data.

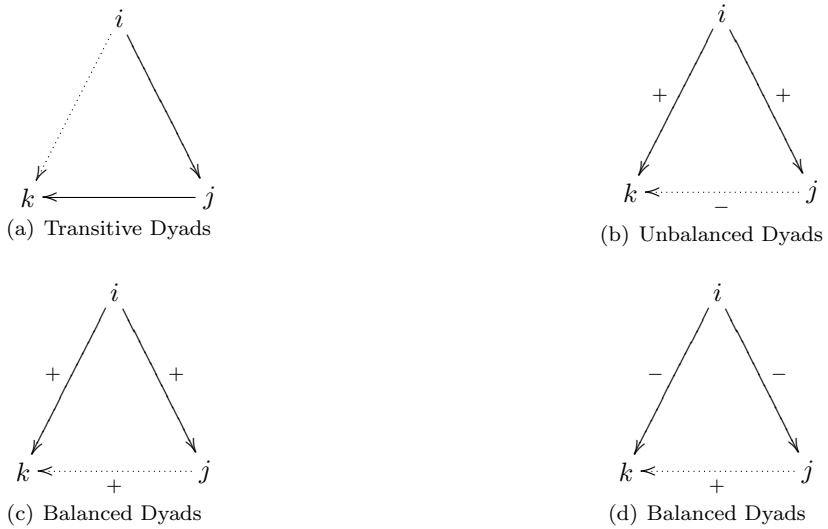
The concept of *balance* implicates the putative stability in a triad; a large set of balanced triads is therefore called a balanced graph or balanced network. *Clusterability* is a relaxed version of balance: a triad is clusterable if it is balanced or the relations are all negative.[10] During the cold war, for example, members of the NATO pact all had mostly cooperative relations with one another, as did members of the Warsaw Treaty Organization (the first WTO), but members in each "cluster" had relatively hostile relations across these two blocs. Thus, while the entire network was not necessarily clustered into one grouping, individual countries could be clustered in terms of their probability of having positive relations with those clustered nearby and negative relations with those located in different clusters, further away. This kind of third-order dependence is generally prevalent in network data, and transitivity, balance, and clusterability are commonly found in dyadic IR data sets that capture international transactions among countries.

As we think about third-order dependencies, knowing the relation between i and j as well as between j and k may reveal something about the relationship between i and k, even when we do not directly observe it. Hoff, Raftery, and Handcock (2002) note:

> In some social network data, the probability of a relational tie between two individuals may increase as the characteristics of the individuals become more similar. A subset of individuals in the population with a large number of social ties between them may be indicative of group of individuals who have nearby positions in this space of characteristics, or "social space." Note if

[10]Networks composed entirely of negative links are quite rare outside of academic departments.

FIGURE 1. *Transitivity and Balance in Triads*



(a) Transitive Dyads (b) Unbalanced Dyads

(c) Balanced Dyads (d) Balanced Dyads

some of the characteristics are unobserved, then a probability measure over these unobserved characteristics induces a model in which the presence of two individuals is dependent on the presence of other ties.

In other words, the "social space" summarizing these unobserved characteristics is another "image" of the third-order dependence in these dyadic data. We describe how to estimate and map these latent positions such that positions in a latent space represents these dependencies. Stated differently, once the higher-order dependencies are taken into account, the dyadic data can be analyzed by techniques such as regression that assume the data are independent of one another.

Conditional on the inclusion of a latent dimension that captures the dependence of the observation on one another, the dyadic data can be treated as independent, and the coefficients can be estimated by well known techniques. If the data are truly independent, estimations under these two scenarios will be equivalent, indeed identical. However, if there is clustering or dependence of the observations on one another in the manner specified above, they will diverge substantially. Our hypothesis is that evidence for the Kantian peace argument is based on a faulty assumption of the independence of international events and that once these dependencies are incorporated into the model, a different picture will emerge. The latent variables capture this dependence and allow us to more precisely estimate the effects of the independent variables upon the probability of conflict.

Using the social relations re-formulation to capture sender (initiator), receiver (target), and higher-order dependencies in the MID data, we explore a model analogous to the original specification in Oneal and Russett. This model is analogous, but has two distinct advantages. First, it permits the direct estimation of important dependencies in the dyadic data. These dependencies have implications not only in the means but also in the covariance structures of the data. Second, because the model has a hierarchical format it is unnecessary to force non-dyadic variables to have a dyadic character (or vice versa), an unfortunate practice that is widespread in the empirical literature. Each partner to any particular interaction has associated covariates at both the country and dyad level. The hierarchical specification allows relational as well as country specific linkages to be coherently included in a single model. Often researchers have tried to deal with dyadic data by creating dyadic versions of all the variables. Joint democracy is an example of such a variable. But we will have much richer analyses when we have models that allow for covariates that are dyadic as well as for covariates that may selectively affect the initiators and targets of dyadic actions. For example, Oneal & Russett (1999) includes only covariates that are entirely dyadic in operationalization, a procedure that is typical in the literature. Many extant studies have multilevel covariates, but often these levels are homogenized to one level.[11] As a result, there is a distinction in the literature between dyadic and monadic studies of the Kantian Peace. Our conceptualization emphasizes that both "monadic" and "dyadic" covariates co-exist, at least plausibly, and this framework provides a standard way of

[11]An exception to the contrary is Hegre's early study using Cox regression (2004).

incorporating them that also has the benefit of taking advantage of the social relations conceptualization. There are individual and dyadic effects that are neither transparent nor identical in groups that have highly dependent relations. A multilevel approach helps to tease out these specificities. As a result of these advantages we have been able to adopt a simple approach to measuring the three putative foundations of the Kantian peace: democracy, trade, and international organizations.

The Kantian strand of the democratic peace theory asserts that dyadic levels of democracy will tend to suppress the likelihood of international conflict and will absolutely prevent the occurrence of war. High levels of economic interaction also have an independent effect in suppressing the probability of conflict, as does joint participation in international organizations. Each of these causal covariates is thought to operate independently. A variety of control variables are often employed, typically including a measure of the capability of the state, such as the level of economic development, measured as gross domestic product (GDP), the size of the country as measured by population, and the amount of democracy in a particular country. At the same time it is often argued that countries that are proximate have a greater opportunity and therefore likelihood of interstate conflict. Thus, some measure of distance between countries is almost always included in contemporary models. We include all of these in our specification, as well. What is different is that the standard models assume that the observations are independent.[12] Thus, they assume that the estimated coefficients apply equally to every one of a set of independent observations.

The specific variables we employ are patterned after the major results in the literature. Dyadic variables (x_{ij}) capture the pillars of the Kantian peace: (1) joint governance structures measured as the product of the polity score of sender and receiver, (2) the imports of the sender from the receiver, and (3) the number of international organizations in which both are members. Like most empirical studies, we also added a measure of the distance between pairs of countries; we use the distance in kilometers between the capital cities. In the dyadic context, i and j index the sending and receiving countries. For both sender and receiver, three country-level covariates (x_i) capture the population, the economic size measured in terms of gross domestic product, as well as the regime type of each country. Detailed information on each of the covariates and their sources may be found in Appendix A.[13]

The model is:

$$\theta_{ij} = \beta_d^{'} x_{ij} + \beta_s^{'} x_i + \beta_r^{'} x_j + a_i + b_j + u_i^{'} v_j + \epsilon_{ij}, \tag{1}$$

where

$$\begin{aligned}
\beta_d x_{ij} &= d \in \text{dyadic effects: joint democracy, imports, joint IO membership, distance} \\
\beta_s x_i &= s \in \text{sender effects: population, GDP, democracy} \\
\beta_r x_j &= r \in \text{receiver effects: population, GDP, democracy} \\
a_i &= \text{random effect of sender} \\
b_j &= \text{random effect of receiver} \\
u_i^{'} v_j &= \text{separate latent positions for sender and receiver} \\
\epsilon_{ij} &= \text{error.}
\end{aligned}$$

We modeled the probability of a militarized dispute y_{ij} with a logistic link function:

$$P(y_{ij} = 1 \mid \theta_{ij}) = \frac{e^{\theta_{ij}}}{1 + e^{\theta_{ij}}}.$$

This setup not only adds to the standard logistic model covariates that are specific to senders and receivers, but also includes both sender and receiver random effects, along with an estimate of the unmeasured latent positions of each country in the militarized interstate dispute network. These latent positions (u_i and v_j) index the propensities for country pairs to have similar interaction patterns toward other countries as a result of *unmeasured*

[12]The often made claim that correcting standard errors by clustering on dyads controls for the dependence of dyadic observations is incomplete and frequently misleading. The adjustment of standard errors within panels does not address the dependence of the observations on one another in the same time period, but is a useful post-estimation strategy to obtain better estimates of standard errors in certain situations.

[13]Open source software written for the $\mathcal{R}$ computer language is available to estimate general bilinear mixed effects models at: `http://www.stat.washington.edu/hoff/Code/GBME/`.

similarities.[14] The latent positions of two countries will be similar if countries are responsive to one another or if they have similar response patterns involving other countries. Latent similarity captures transitivity and balance, and reveals clustering. One strong benefit of this specification is that in addition to the latent positions, four additional quantities of interest are developed. Each reflects some aspect of the higher-order dependencies. Since we model the random effects as being multivariate normal, we can estimate their covariance structure: σ_a^2 is the variance of the sender random effects and σ_b^2 the variance of the receiver random effects. Additionally, the covariance between these two components is given by σ_{ab}. These terms may be thought of as *first-order* components. In a similar way, the covariance of the errors across dyads, i.e., the covariance of the errors between ϵ_{ij} and ϵ_{ji}, can also be parameterized as $\rho\sigma_\epsilon^2$, allowing a specific measure of reciprocity to be estimated (ρ). This is a *second-order* effect. Finally, the latent vectors for senders and receivers capture the *third-order* dependencies. In this simple way, higher-order dependencies are posited in new quantities of interest that have been previously assumed to be nonexistent or unimportant. For the first time, it is possible to estimate these quantities. Hoff (2005) provides details and full conditionals for this model. The following moments are defined for the ϵ_{ij}:

$$\begin{aligned} E(\epsilon_{ij}^2) &= \sigma_s^2 + \sigma_r^2\sigma_\gamma^2 \\ E(\epsilon_{ij}\epsilon_{ji}) &= \rho\sigma_\gamma^2 + 2\sigma_{rs} \\ E(\epsilon_{ij}\epsilon_{ik}) &= \rho\sigma_r^2 \\ E(\epsilon_{ij}\epsilon_{kj}) &= \sigma_s^2 \\ E(\epsilon_{ij}\epsilon_{ki}) &= \sigma_{rs} \end{aligned}$$

where r represents the receiver and s the sender. The quantity σ_r^2 is the dependence among dyadic observations with a common target; σ_s^2 the dependence among dyadic observations with a common source; and ρ captures the reciprocity, or correlation within a dyad. Standard approaches assume that all of these quantities of interest, as well as the latent positions, are zero. But this assumption is neither plausible theoretically, nor borne out in empirical analyses we report below.

Statistical Results

Since it has been demonstrated that the empirical results for the Kantian peace are stronger in the period following the end of the World War II (Gowa 1999, Box-Steffensmeier, Reiter & Zorn 2003), we have focused on militarized interstate disputes over the last fifty years, estimating the model in eleven specific years: $\{1950, 1955, \ldots, 1995, 2000\}$. Table 2 provides one "big table of numbers" for the Bayesian estimation of the hierarchical, bilinear random effects model for the year 2000, as an illustration of our basic results.[15] It is important to note that the 95% empirical credible limits for all the dyadic variables representing the Kantian peace argument include zero. Stated differently, the three traditional explanatory variables that undergird the notion of a Kantian peace do not have strong, unambiguous empirical effects in the most recent year available for analysis. However, as we show below, in most other years the patterns are more supportive of the notion that higher levels of democracy tend to reduce conflict probabilities; in most years dyadic democracy levels exert a modest, negative impact on the probability of a militarized interstate dispute.[16]

The results in Table 2 also demonstrate that the higher-order dependencies posited are prominent in these data. These dependencies are represented, in part, by the common sender variance, the sender-receiver covariance, and the common receiver variance. Each is large and important. Taken together, these three components exert a vastly stronger influence on the probability of a militarized dispute among pairs of countries than the standard covariates. Together they are substantially larger than the error variance. It is also worth noting that the dyadic reciprocity of the democratic peace in this particular year is large, a result similar to the large autoregressive coefficients widely reported in empirical literature on the Kantian peace. The size of these components serves to underscore the importance of dependencies in these dyadic data on militarized interstate disputes. These results hold in all years we examined.[17]

[14] We estimate these positions in three dimensions each, but employ multidimensional scaling to display them in two dimensions below. See Cao (2007) for another application of the bilinear mixed effects model in the study of domestic economic convergence.

[15] Full results for the other ten years estimated are available on the replication archive for this project.

[16] Similar results can be found in Bennett (2006).

[17] All the results from the generalized bilinear model in each of the eleven years analyzed may be found on the replication Web site for this project: `http://faculty.washington.edu/mdw/data.html`.

TABLE 2. *Bayesian estimates for 2000, using equation (1), are the posterior means for the estimated quantities. Quantile-based, empirical credible intervals of 95% are presented. The mean posterior log likelihood is -30.139.*

Posterior Distributions of Estimated Coefficients for Year 2000.

	2.5%	*Mean*	*97.5%*
Constant	−82.78	−55.59	−34.32
Dyadic Effects: $Polity_i \times Polity_j$	−0.08	−0.02	0.05
$Import_{ij}$	−0.40	−0.18	0.01
IGO_{ij}	−0.22	0.02	0.26
$Distance_{ij}$	−8.18	−5.55	−3.94
Sender Effects: $Population_i$	−0.01	0.01	0.03
GDP_i	0.00	0.01	0.01
$Polity_i$	−0.54	0.39	1.37
Receiver Effects: $Population_j$	−0.03	−0.00	0.02
GDP_j	0.00	0.01	0.01
$Polity_j$	−0.65	0.35	1.30
Dependencies: Common Sender σ_a^2		170.4	
Sender-Receiver $\sigma_{a,b}$		125.4	
Common Receiver σ_b^2		148.0	
Reciprocity ρ	0.99	0.99	1.00
Error Variance σ_ϵ^2		270.9	

We present a graphical display of the estimated coefficient for each of the four dyadic variables for each of the eleven years we analyzed in Figure 2. Note that the variances of the estimates are reduced in more recent yearly time slices, and that the effect of each variable moves around considerably from year-to-year, in contrast to the assumptions made in standard, pooled analyses. In particular, the impact of the number of joint memberships in international organizations moves from exerting a modest suppression on militarized interstate disputes in the 1950 through 1970 era, to exhibiting a slight, but always positive impact thereafter. This suggests that international organization membership does not, in recent years, reduce the probability of militarized interstate conflict, but may actually be associated with a greater number of dyadic disputes. In the same way, with the exception of 1990, both the impact of joint democracy and international trade within dyads appears to exert a much smaller impact on the probability of militarized disputes in recent years. This goes against any claim that the forces of the Kantian peace are getting stronger as more countries become democratic. On the other hand, the second row of estimates illustrates that the level of democracy in each country in a dyad generally has an association with increased probability of conflict in militarized interstate disputes. Thus, democracies appear to be more likely to be the first-mover, or initiator, in a militarized dispute *and* they are more likely to be targets of militarized disputes, though this effect seems to be diminishing. It is the confluence of democracy on both sides of the dyad that is associated with fewer instances of conflict, a result that is entirely consistent with the bulk of research on the broader topic of the democratic peace.

Analyzing Decades. In addition, we analyzed each decade since 1950, using the a bilinear model as specified above. One distinct difference in these analyses is that we counted the number of disputes for each dyad pair during the decade, rather than a simple binary indicator of dispute. As a result, we employed a Poisson link function in the bilinear model. The results are presented visually in Figure 3. These results show that imports tend to reduce the probability of disputes, though in the 1970s this effect was positive, and the 1975 analysis had credible intervals spanning zero. Similar to results for individual years, the impact of IGO membership is frequently positive, rather than negative, as the Kantian peace argument would suggest. Distance has a strong negative impact on the probability of MIDs, and its decade averaged impacts appear to be increasing in magnitude.

Compared to the annual results, however, the joint democracy variable has strong and consistently negative impacts on the probability of interstate conflict, exactly as predicted by the Kantian peace argument. To the extent that both partners are more democratic, their joint probability of getting into a dispute with each other

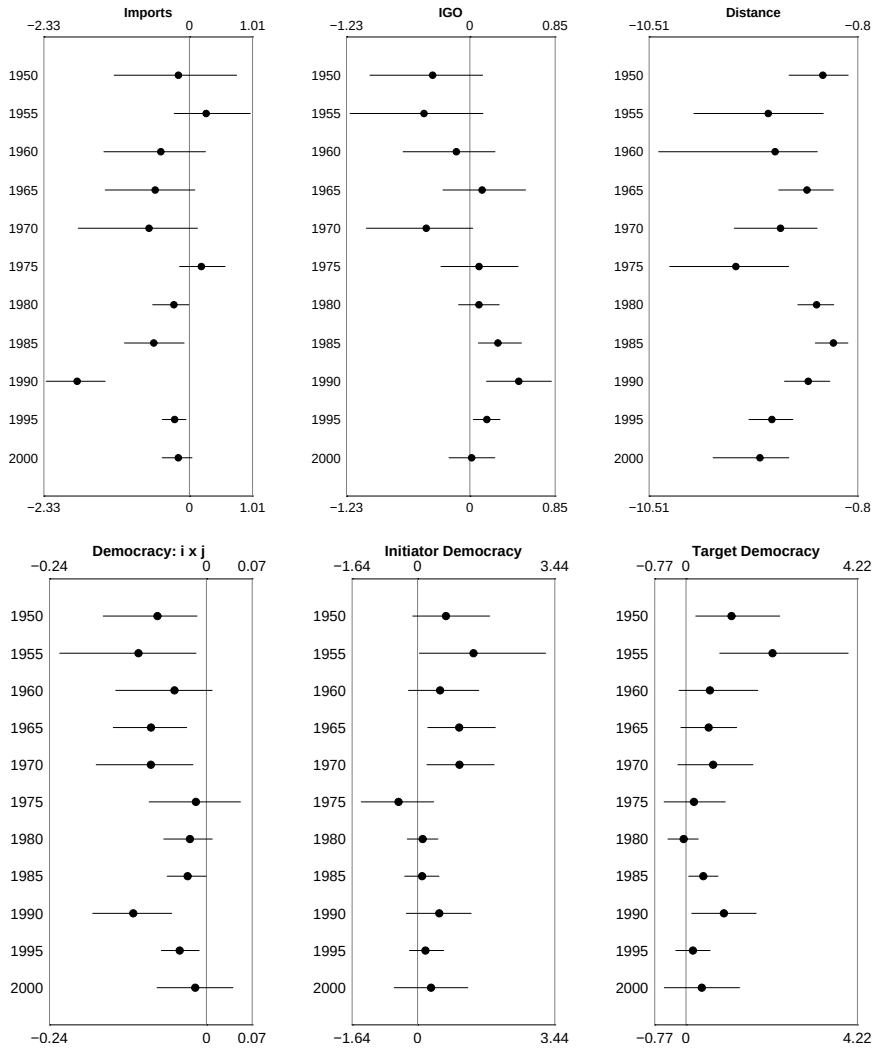


FIGURE 2. *Changing effects of covariates, 1950 to 2000. The dot in each panel represents the location of the posterior mean–i.e., the estimate–and the lines present the 95% credible intervals. The top row of displays portrays the estimates for dyadic covariates. The first panel in the second row illustrates the estimates for the measure of joint democracy, followed by the separate coefficients for democracy for initiators and targets of militarized interstate disputes. In general, more recent periods have effects that are closer to zero, with a smaller variance around these estimates.*

is reduced. The separate impacts of the level of democracy on both sides of a dyad is shown to be generally negative, but often indistinguishable from zero. Generally, the level of democracy in the initiator is more a powerful retardant for dispute involvement than is the level of democracy in the target. In short, we find more support for the democratic pillar of the Kantian peace in the decade-based analyses.

Predicting Militarized Interstate Disputes. How well does the general bi-linear model predict militarized interstate disputes? First, we return to the yearly analyses. Incorporating the higher-order dependencies for the annual data from the year 2000, the model correctly classifies 63 out of 114 militarized interstate disputes (52%, for a cut point of 0.0048), accompanied by only 12 false positives. We also provide predictions mirroring our earlier experiment reported in Table 3, which reflects data for a single year. Thirty-seven disputes are correctly identified by our model using a cut-point of 0.50, with only six false positives. There is a substantial number of false negatives (77). Using the *ex post ante* mean to determine the prediction threshold, 63 correct dyadic MID involvements are predicted, with 12 false positives and 51 false negatives. It appears that the model is making

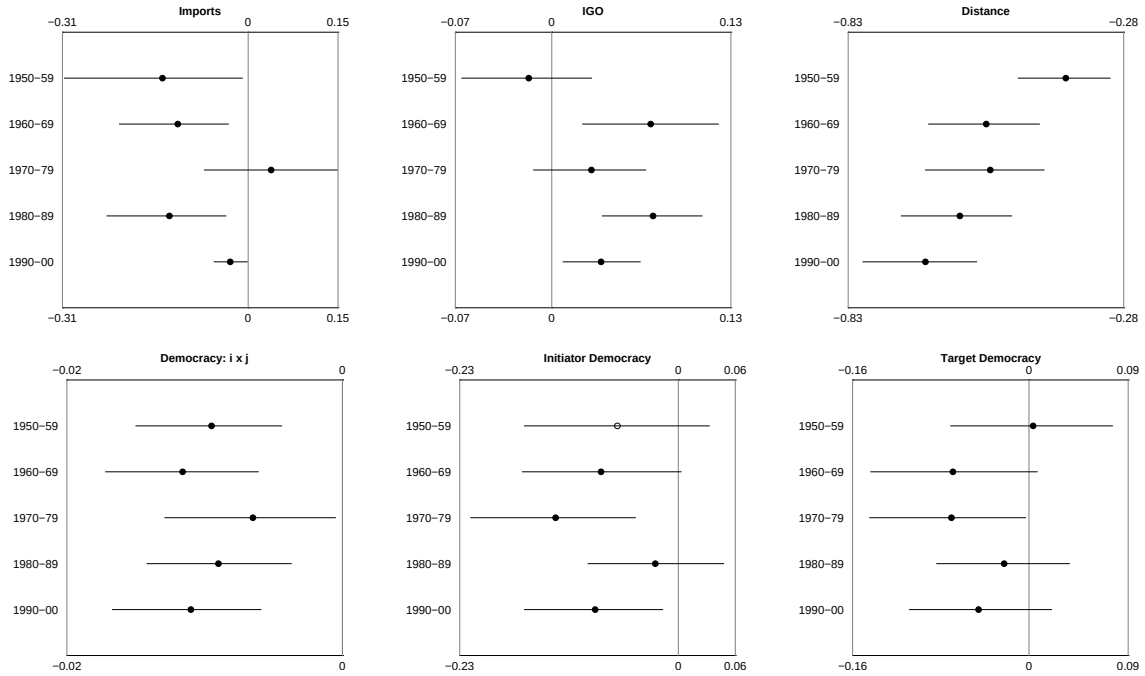


FIGURE 3. *Changing effects of covariates, for five decades between 1950 to 2000. The dot in each panel represents the location of the posterior mean–i.e., the estimate–and the lines present the 95% credible intervals. The top row of displays portrays the estimates for dyadic covariates. The first panel in the second row illustrates the estimates for the measure of joint democracy, followed by the separate coefficients for democracy for initiators and targets of militarized interstate disputes.*

reasonable predictions, correctly identifying about one-half of the observed disputes, accompanied by a relatively low rate of false predictions.[18] Importantly, of the 75 predicted disputes about 85% are actual disputes. Not every year has results comparable to those for the most recent year. In some single years, there are very few actual MIDs. Other years have much more modest predictive success. For example, in 1950 (shown in Table 4) only a few militarized dispute involvements are correctly predicted; however, seven disputes are correctly predicted out of a possible 36–almost 20%, accompanied by only three false positives. As shown in Table 5 in 1960, 1975, and 1980 the model fails to predict a single conflict.

TABLE 3. *The predictions from the bilinear, mixed effects model for the year 2000.. Thirty-seven dispute involvements are correctly identified with a* 0.50 *cut-point, sixty-three when the posterior mean is used. These basic patterns do reflect a model that is accurate, however, in comparison to most published studies. The model on which these results depend is that presented in Table 2 above.*

Cut-point: 0.50	Non-Disputes	Disputes
Predicted Non-Disputes	23442	77
Predicted Disputes	6	37
Cut-point: 0.0048		
Predicted Non-Disputes	23436	51
Predicted Disputes	12	63

[18]One metric of "fit" is the Proportional Reduction in Error, variously known as the PRE or λ_P statistic, which is an index of the association between two nominal scales. Using 0.50 as the cut-point yields a PRE in 2000 of 0.27 and 0.45 when the cut-point is set to the posterior mean of 0.0048. We are indebted to Bruce Bueno de Mesquita for elaborating this point for us.

We present in Figure 4 the latent positions of countries in 1950. These are shown in two plots, one for the MID initiator latent space, on the left, the other the target latent space, shown on the right. Colors correspond to geographical positions, so that similar countries are proximate to one another. For example, the countries of North and South Asia are shown in an aquamarine/turquioise tone. China and North Korea besides being the largest vectors in the latent space, are located close to one another, virtually overlapping. This illustrates that they have similar patterns (in 1950) of initiation of militarized disputes with other countries. Other groupings in this latent mapping illustrate the clustering of European countries (shown in maroon and also in purple hues). The corresponding illustration of receiver latent space for 1950 also shows considerable geographic clustering. The basic idea of these latent clusters is that they capture the extant dependencies in the data, incorporating them in to the model, so that conditional on these latent clusters, the dyadic data may be treated as independent and identically distributed.

Even in years that are sparse in the number of MID involvements, the model still performs better than previous models. Table 5 illustrates the predictions of the model using the posterior mean for each year as the cut-point. These results, of course, were generated in-sample. It is a much more demanding task to generate good out-of-sample results, where the predictions are not a direct consequence of the data that generated them. We now turn to such an approach, introduced in political science by Beck, King & Zeng (2000).

TABLE 4. *The predictions from the bilinear, mixed effects model for the year 1950. The cut point of* 0.0063 *is the mean frequency of militarized interstate disputes in 1950.*

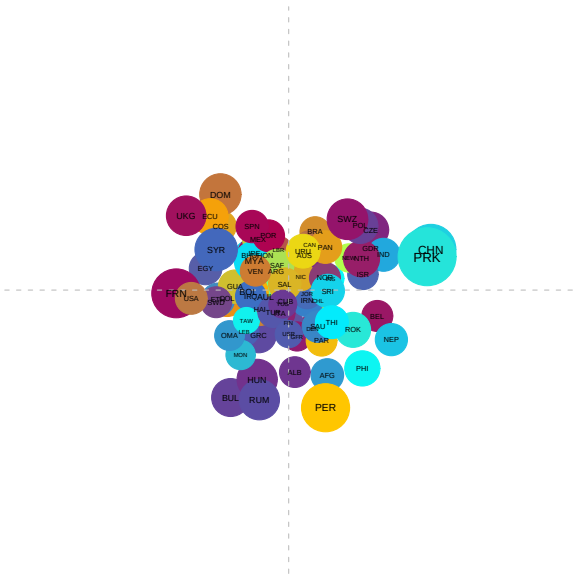
Cut-point: 0.50	Non-Disputes	Disputes
Predicted Non-Disputes	5663	33
Predicted Disputes	3	3
Cut-point: 0.0063		
Predicted Non-Disputes	5661	29
Predicted Disputes	3	7

TABLE 5. *The predictions of the basic bilinear model for each year show a modest improvement in most years over a model that always predicts zero. Predictive success is best in the most recent year.*

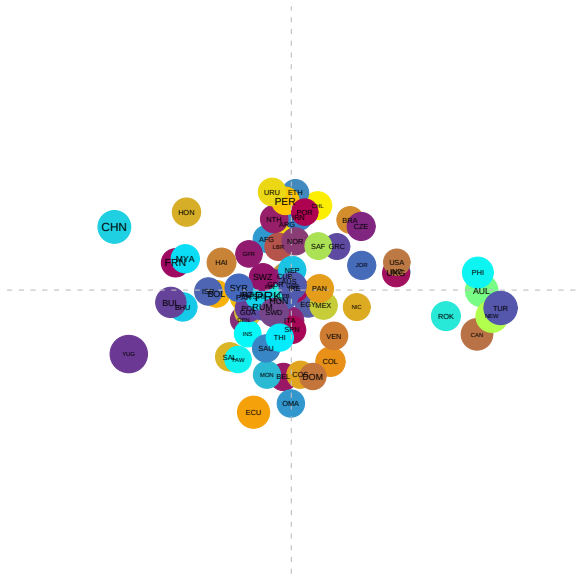
Year	Cut-point	Correctly Predicted MID Involvements	False Negatives	False Positives	Correctly Predicted Zeros	MIDs	Dyads	Countries
1950	0.0063	7	29	3	5661	36	5700	76
1955	0.0045	1	29	0	6612	30	6642	82
1960	0.0031	0	34	0	10886	34	10920	105
1965	0.0039	6	50	4	14220	56	14280	120
1970	0.0029	12	36	2	16462	48	16512	129
1975	0.0022	0	42	0	19140	42	19182	139
1980	0.0021	0	40	4	18588	40	18632	137
1985	0.0030	2	54	2	18574	56	18632	137
1990	0.0030	2	53	1	18576	55	18632	137
1995	0.0023	6	49	6	23809	55	23870	155
2000	0.0048	63	51	12	23436	114	23562	154

A Cross-validation Experiment. We focus our discussion on the most recent ten year period from 1990-2000.[19] There are 130 countries for which data are available in this period, resulting in a dyadic matrix of 130×130, with the diagonal set by definition to 0. We randomly divide these dyads into two equally sized sets, a training (fit) set and a test set. The model is estimated using data from the training set. Estimated coefficients are then used in

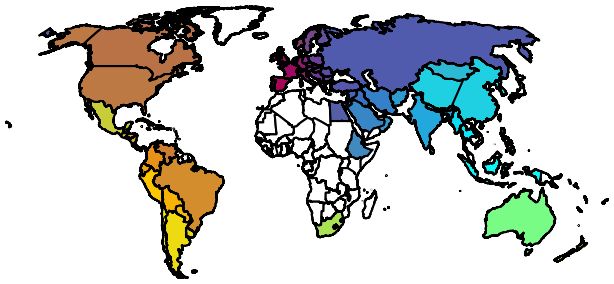
[19]Each decade was examined, and produced similar results. Because of space limits, we only present the most recent decade.



(a) Latent Space of MID Initiators, 1950



(b) Latent Space of MID Targets, 1950



(c) Legend

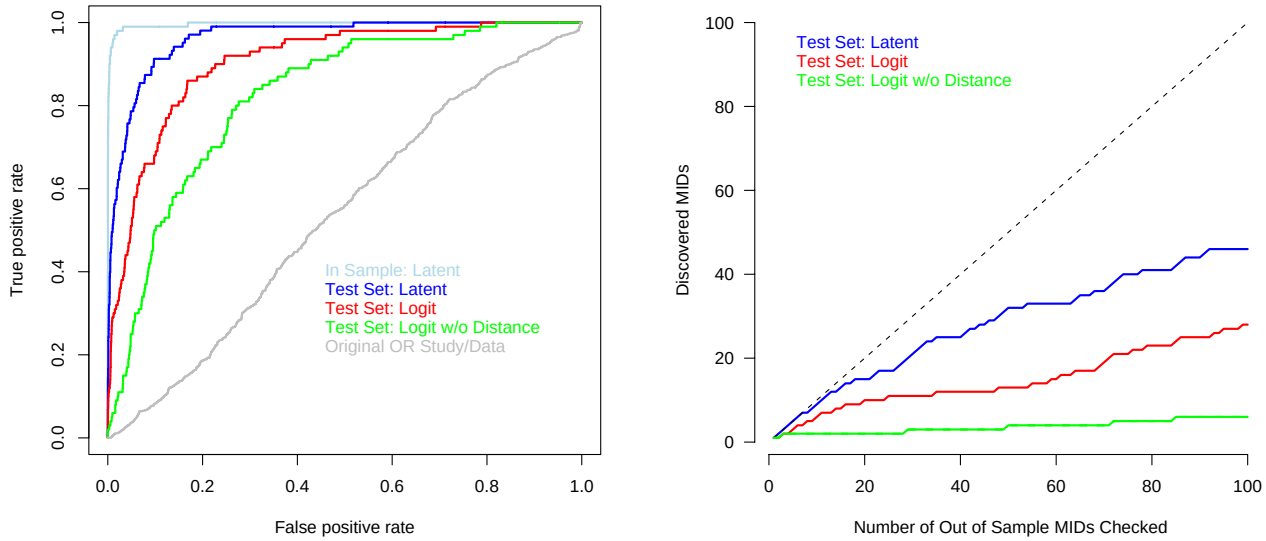
FIGURE 4. *The estimated latent space for initiators and targets of militarized interstate disputes in 1950 illustrate the 3rd order dependencies in the dyadic militarized interstate dispute data. Conditional on these latent positions, the data may be treated as independent in order to estimate regression models of the effects of covariates on the probability of militarized interstate dispute. The colors of countries are chosen so that proximate countries have proximate colors. The geographical map displays these colors by way of providing a visual legend.*

combination with the covariate data from the test set to generate predicted probabilities in the test set. This is an out-of-sample, cross-validation of the model. We treat the dependent variable as binary for any conflict during the decade, and average covariates over the period. Several models are estimated. First we estimate the general bilinear latent space model, with three latent dimensions as specified in Equation 1. This model was estimated for the full data, then it was estimated for the randomly generated 50% sample which comprise the training set. The estimates from the fit set estimation were used to predict MIDs in the test set of data; this is labeled "Test Set: Latent," in figures to follow. Next, using the fit set, we estimated a model that deleted the higher-order dependencies from Equation 1; this model is essentially a multilevel logistic regression that includes the dyadic as well as sender and receiver covariates. This is labeled "Test Set: Logit," below. Another model was estimated in the same fashion, except that the dyadic covariate that measures the distances among actors was deleted; this model is labeled "Test Set: Logit w/o Distance." Finally, for comparability we conducted a separate 50% cross-validation on the Oneal and Russett model based on Column 2 of Table 1 in their study; this model uses the Oneal and Russett data, their covariates, and their specification, including the four splines to represent the peace years correction. This is labeled "Original OR Study/Data." We include the peace years correction in the Oneal and Russett model in this cross-validation experiment in order to gauge the relative performance of that model which specifies a time dependency within dyads, but ignores higher-order dependencies. This compares to the model developed herein which ignores the time dependencies, but focuses on the higher-order dependencies.

Figure 5(a) presents the Receiver Operator Characteristics curves for these estimations. The ROC curve presents the true positive prediction rate as a function of the false positive prediction rate, for all possible cut-points dividing the estimated probabilities into predicted events and non-events. The light blue line illustrates that the in-sample accuracy of the latent space model is quite high. The out-of-sample ROC curve for this model is shown in blue, and is below the in-sample fit, illustrating that the model performs worse in out-of-sample tests, but that it still is fairly accurate. The red line presents the ROC curve for the standard logistic implementation of the model (which does not specify any higher-order dependencies). This line is interior to both blue lines, illustrating that it has worse fit compared to the other two estimations. We present a green line which is a standard logistic model that is estimated without the inclusion of the distance variable in the specification. This model captures the ability of the Kantian tripod variables to accurately predict the absence and presence of militarized disputes in out-of-sample tests. Finally, the gray line close to the diagonal presents cross-validation results for the original Oneal and Russett model (including peace years corrections), using their specification and their data for the same decade. This model is the least successful in the cross-validation experiment. In general, including the higher order dependencies allows one to have improved model performance, and it turns out that including distance as a covariate is more important than whether any of the covariates frequently used in this literature is *included*.

Figure 5(b) illustrates the out-of-sample predictive ability of a standard logistic regression compared to the specification which includes the latent representations of the higher-order dependencies in the data. In this figure we examine the 100 largest predicted probabilities of each of the estimated models, sorted from largest to smallest. We then examine each corresponding dyad to see whether a militarized dispute existed in that dyad. We keep a running total of the number of correctly predicted MIDs and plot it against the running number of dyads checked (from one to 100). For any given model, if the largest 100 probabilities corresponded to 100 dyads that were in militarized interstate disputes, then the plot of the cumulative success would fall along the line $y = x$; if no MIDs were found in the dyads with the 100 largest predicted probabilities, the line would be indicated by the line $y = 0$. Better models will have a cumulative predictive success that are closer to the diagonal. This kind of comparison complements the ROC curve and provides a more complete picture of the out-of-sample predictive performance than do single number summaries.

The standard approach substantially under-performs models that include the latent representation of the higher-order dependencies in these data. The latent space approach is able to capture approximately twice the number of actual disputes that the logistic specification identifies. In order to gauge further which covariates were important, we selectively eliminated joint democracy, joint memberships in intergovernmental organizations, and measures of economic interdependence, as well as the country level covariates. None of these made any difference in the outcome. Instead, we found that deleting the geographic proximity variable drastically undercut the out-of-sample predictive ability of the standard model. This is portrayed in Figure 5(b) by the green line. In short, these results suggest that (a) despite a strong in-sample descriptive performance, a typical approach to examining the Kantian peace is out-performed substantially by a model that includes the higher-order dependencies that characterize dyadic data and (b) most of the power of the standard Kantian peace model in out-of-sample tests is derived from the inclusion of geographic proximity of countries.



(a) ROC Curves

(b) Correct Predictions of MID among largest probabilities.

FIGURE 5. *Out-of-sample cross-validations of a logistic and latent space model, estimated on training data from 1990 to 2000 are examined against test data from the same period. Results are also presented for a standard logistic model in which the distance covariate is deleted. This illustrates that the out-of-sample predictive power of the standard Kantian peace model rests in large part on the inclusion of the geographic proximity of dyad pairs, rather than on the explanatory power of the covariates capturing the substance of the Kantian peace argument; Receiver Operator Characteristic Curves for the Latent Space Model (In and out-of-sample), along with the out-of-sample logistic with and without the distance variable. The ROC curve (gray line) is also given for an 50% out-of-sample cross validation with the original Oneal and Russett data and model.*

Figure 6 presents a map of countries' positions in three latent dimensions estimated for the period from 1990 to 2000. The latent dimensions, represented here in two dimensions, are scaled to have unit length, but the size of the latent vector is proportional to the size of the acronym representing the country. Further, the initiators are scaled to be located interior to the latent representation of the targets. Three clusters of initiation are evident: one in Africa (indicated by the acronyms for Uganda and Rwanda), another in the Middle East (flagged by Iraq and Jordan), and a third, led by the United States. In terms of the dependencies at the other end of the Militarized Interstate Dispute, Namibia and the Democratic Republic of Congo lead an African cluster, Iraq is revealed as a likely target of U.S.-initiated disputes, while Jordan, Iraq, and North Korea are shown to be hostile toward the United States and many of its traditional allies. These dependencies are shown by the closeness of country positions in this representation of latent space. Actual militarized interstate disputes are shown in this figure by dashed lines connecting the countries involved.

By way of summary, we observe that a bilinear, latent space approach allows coherent heuristics about models that are based on dyadic data. In particular, we learned that it is possible to make new inferences about the Kantian peace. Prime among these is that the Kantian peace tripod is shakier in some years than in others, and we need to understand better this temporal fragility. It seems most solid in years in which scant conflict occurs, but worse in years in which there is more conflict. Ideally a predictive model of conflict should be as strong when there are more conflicts as when there are fewer. Without such a property it is a less reliable guide to policy than would otherwise be the case and, moreover, it is less satisfying theoretically.

Perhaps more importantly, we also learned that interaction dependencies abound in existing data used to study the Kantian peace. Under the standard approach to the Kantian peace–and, we suspect–many other problems in international relations research, second- and third-order dependencies among the data overwhelm the

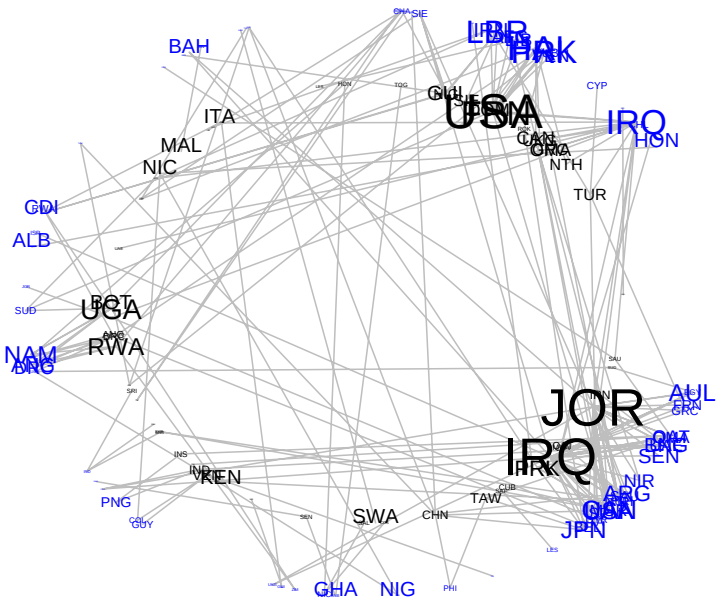


FIGURE 6. *A representation of the estimated latent space for both initiators (inner ring) and targets (outer ring, in blue) of Militarized Interstate Disputes (MID) over the period from 1990 to 2000. Actual MIDs are shown by a line connecting the disputants.*

impacts of the standard covariates thought to explain why democracies do not fight one another. This does not mean that empirical theories are incorrect, necessarily. It does mean that they are incomplete if they fail to take into account these contextual dependencies. In particular, we must recognize that countries have a character to their foreign policies that is broadly consistent, imposing a correlation in their actions toward others. Often this is the result of explicit coordination. For example, at the turn of the century, then Chancellor of Germany Gerhard Schröder and President of France Jacques Chirac held *informal* bilateral dinners every six weeks, reportedly in part to coordinate their foreign policies in opposition to the United States.[20] Similarly, some countries elicit similar responses from many others, imposing a correlation in their dyadic behavior patterns. These features of the social relations model are missing from most models in international relations, even those that recognize consistent patterns within dyads over time. These results also show that context is neither impossible to take into account systematically, nor solely the domain of qualitative analysis.

Conclusion

The broad research program of the democratic peace has been an attractive research program to many in the field of international relations. This attraction is probably due to the coherence of the program, as well as to the fact that it offers the basis for theoretically informed, empirically supported policy advice, that, if adopted, ought to motivate the enhancement of democratic political institutions. The results of the Kantian Peace research program contain similar implications for the promotion of democracy and increasing through increasing levels of trade and membership in international governmental organizations. However, in our re-examination none of these aspects

[20]These were the so-called Blaesheim meetings, named after the French village in which the first one occurred–in January 2001.

receives strong empirical support in the analysis of data from single years. Instead the dependencies in the dyadic data among common senders and common receivers overwhelm the standard covariates thought to be supportive of the Kantian peace. However, when we pool the analysis for decades, we see that these components regain their statistical prominence. We might conclude that the Kantian peace is a longer term phenomenon that might not be visible in separate, yearly slices of data. However, there is an inescapable fact that undercuts this conclusion. When we conduct out-of-sample tests of the basic model, we find that despite high statistical significance, none of the variables representing the Kantian tripod is associated with any substantial degree of predictive success. Instead, it is only the distance variable that undergirds the predictions of the standard representation of the Kantian peace. This suggests that the extant results frequently cited in the quantitative literature on the Kantian peace are probably plagued by over-fitting. However, the goal of these analyses is an identification of the process whereby international conflict is influenced by democracy, trade, and international organizations.

In comparison, the model that captures the dependencies among countries performs remarkably well in out-of-sample tests. While it is true that more complicated models will always do better in in-sample statistical tests, this is not true of out-of-sample examinations. Consequently, we have greater confidence in the contribution that our specifications of the dependencies can bring to the study of the Kantian peace and other international phenomena.

Our results should raise some skepticism about the robustness of the Kantian peace argument. That is not to suggest jettisoning the Kantian ideas. But it does suggest that our certitude about it needs tempering. To be sure, shared democratic political institutions in a dyad does reduce the probability of conflict, but the overall effect is modest. Modest effects can, of course, be important. The idea that shared memberships in intergovernmental institutions reduces conflict shows great variation over time, and in recent years actually appears to have a modest positive effect–it is associated with more conflict, not less. Finally, support is virtually nonexistent for the idea that high levels of trade dampen conflict. The Kantian tripod rests on a weaker empirical basis than was previously thought.

In addressing these substantive problems we also attempted to cast light on three more general problems in some empirical work in international relations. The first problem is that dyadic data contain dependencies that are omitted by most popular procedures for analyzing international relations.[21] Because regression-based approaches assume that the data are exchangeable and the errors independently and identically distributed, they are unable to capture the extent to which dependent data may actually reflect the ebb and flow of international politics. In reviewing the difficulties of analyzing dyadic international relations data, King (2001) noted that the thorniest problem is to unravel the dependencies in dyadic data that result in biased estimates of coefficients and covariance structures, and further suggested that "[a] logical methodological starting point for addressing the problems at hand would be based on Bayesian hierarchical, random effects, or split population models" (page 506).[22]

A second important problem concerns the basic inferential model that dominates popular practice in the field of international relations and other disciplines. Since much research in international relations is based on observational studies that have a large number of cases, there is an appearance of statistical significance in almost all the findings in the literature, even considering the well-known publication bias for positive and statistically significant results. Actually, the prevailing popular statistical approach reflects the ability of the tests to detect small differences–i.e., the power of the tests–as well as the number of observations more than it reveals any underlying statistical significance of hypothesized causal linkages (Savage 1957, Gill 1999). This research tradition somehow expects more from tests of significance, an expectation that probably creates more problems than it solves. While observational data rarely comprise a random sample, they nevertheless are frequently treated as an enormous one. Large *samples* have immense statistical power that large *collations* of observational data do not. As a result, statistical significance tests can be badly misleading in terms of producing information about the underlying data generating processes by allowing scholars to attribute statistical significance to summary characteristics known *a priori* to be different.

Finally, many studies have focused solely on statistical significance as a measure of "fit," while at the same time avoiding evaluations of the out-of-sample fit of estimated model predictions with actual data. Regression diagnostics are only one heuristic, not the terminus of scientific investigation. Indeed, they can often be misleading by suggesting confidence in results that are highly uncertain. We show that a predictive (out-of-sample) heuristic

[21]One early effort examined strategic dependencies in the context of endogenous choice calculus (Smith 1999). Another important initiative (Signorino 1999) extends the Quantile Response Equilibrium in the context of a solution to certain game-theoretic models of strategic choice.

[22]Clark & Regan (2003) have examined split population models, examining heterogeneity, but not dependencies, in different samples.

helps to protect scholars from making statistical inferences that may be problematic. In this regard, we underscore Beck, King, and Zeng's insights about using predictive cross-validation for model assessment in international relations. The basic idea is quite simple: a good model will not only have estimated quantities of interest that are informative, but it also should accurately map the covariate information into the dependent variable in a new set of similar data. In terms of predicting the pattern of militarized interstate disputes, this does not appear to be the case for Kantian peace theory.

One weakness our of work on this topic to date is the absence of any substantial consideration of time dependencies, despite our demonstration that other dependencies are important. Given that many research reports use the time series corrections of Beck and Katz (1995) (who have now somewhat revised their earlier recommendations, (Beck & Katz 2004)) but ignore the second and third order dependencies, it seemed reasonable to focus on the latter as a way of opening up a new line of thinking about dependencies. In the long run it will be important to include temporal as well as higher-order dependencies in our models of interstate interaction. However, no one has yet solved this problem.[23]

[23]Research on this topic is presently underway on several fronts. Ward and Peter Hoff are supported by a National Science Foundation grant (SES-0631531) to study "Longitudinal Network Modeling of International Relations Data." Anton Westveld's recent (2007) doctoral dissertation worked out the covariance structures for simple time dependencies in social networks. Other ideas being explored include the incorporation of both partial adjustment mechanisms and hidden Markov models.

References

Beck, Nathaniel, Gary King and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94(1):21–35.

Beck, Nathaniel, Gary King and Langche Zeng. 2004. "Theory and Evidence in International Conflict: A Response to de Marchi, Gelpi, and Grynaviski." *American Political Science Review* 98(2):379–389.

Beck, Nathaniel and Jonathan Katz. 2004. "Time-Series-Cross-Section Issues: Dynamics, 2004." Web published at `http://www.nyu.edu/gsas/dept/politics/faculty/beck/beckkatz.pdf`.

Beck, Nathaniel and Jonathan N. Katz. 1995. "What to Do (and Not to Do) With Pooled Time-Series Cross-Section Data." *American Political Science Review* 89(3):634–647.

Bennett, D. Scott. 2006. "Towards a Continuous Specification of the Democracy-Autocracy Connection." *International Studies Quarterly* 50(2):513–537.

Box-Steffensmeier, Janet M., Dan Reiter and Christopher J.W. Zorn. 2003. Temporal Dynamics and Heterogeneity in the Quantitative Study of International Conflict. In *Economic Interdependence and International Conflict: New Perspectives on an Enduring Debate*, ed. Edward D. Mansfield and Brian M. Pollins. Ann Arbor, MI: University of Michigan Press pp. 273–289.

Cao, Xun. 2007. Convergence, Divergence, and Networks in the Age of Globalization: A Social Network Analysis Approach to IPE. Working dissertation chapters, University of Washington, Seattle, WA, USA, `http://students.washington.edu/xuncao/dissert/draft_jan2007.pdf`.

Cederman, Lars-Erik. 2001. "Modeling the Democratic Peace as a Kantian Selection Process." *Journal of Conflict Resolution* 45(4):470–502.

Cederman, Lars-Erik and Mohan Penubarti Rao. 2001. "Exploring the Dynamics of the Democratic Peace." *Journal of Conflict Resolution* 45(6):818–833.

Chan, Steve. 1997. "In Search of Democratic Peace: Problems and Promise." *Mershon International Studies Review* 41(2):59–91.

Clark, David H. and Patrick M. Regan. 2003. "Opportunities to Fight: A Statistical Technique For Modeling Unobservable Phenomena." *The Journal of Conflict Resolution* 47(1):94–115.

de Marchi, Scott, Christopher Gelpi and Jeffrey D. Grynaviski. 2004. "Untangling Neural Nets." *American Political Science Review* 98(2):371–378.

Dixon, William J. and Paul Senese. 2002. "Democracy, Disputes, and Negotiated Settlements." *Journal of Conflict Resolution* 46(4):547–571.

Doyle, Michael W. 1983*a*. "Kant, Liberal Legacies, and Foreign Affairs, Part I." *Philosophy and Public Affairs* 12(3):205–235.

Doyle, Michael W. 1983*b*. "Kant, Liberal Legacies, and Foreign Affairs, Part II." *Philosophy and Public Affairs* 12(4):323–353.

Doyle, Michael W. 1986. "Liberalism and World Politics." *American Political Science Review* 80(4):1151–1169.

Doyle, Michael W. 2005. "Three Pillars of the Democratic Peace." *American Political Science Review* 99(3):463–466.

Farber, Henry & Joanne Gowa. 1995. "Politics and Peace." *International Security* 20(2):123–146.

Farber, Henry and Joanne Gowa. 1997. "Building Bridges Abroad." *Journal of Conflict Resolution* 41(3):45–56.

Gabriel, Kuno Ruben. 1998. "Generalised Bilinear Regression." *Biometrika* 85(3):689–700.

Gartzke, Erik. 1998. "Kant We All Just Get Along? Opportunity, Willingness and the Origins of the Democratic Peace." *American Journal of Political Science* 42(1):1–27.

Gartzke, Erik. 2003. "War, Peace, and the Invisible Hand: Positive Political Externalities of Economic Globalization." *International Studies Quarterly* 47(4):561–586.

Gill, Jeff. 1999. "The Insignificance of Null Hypothesis Significance Testing." *Political Research Quarterly* 52(3):647–674.

Gill, Paramjit S. and Tim B. Swartz. 2001. "Statistical Analyses for Round Robin Interaction Data." *The Canadian Journal of Statistics. La Revue Canadienne de Statistique* 29(2):321–331.

Gleditsch, Kristian Skrede. 2002. "Expanded Trade and GDP Data." *Journal of Conflict Resolution* 46(5):712–724.

Gleditsch, Nils Petter and Håvard Hegre. 1997. "Peace and Democracy: Three Levels of Analysis." *Journal of Conflict Resolution* 41(2):283–310.

Goertz, Gary, Bradford Jones and Paul F. Diehl. 2005. "Maintenance Processes in International Rivalries." *Journal of Conflict Resolution* 49(5):742–769.

Goldstein, Joshua S. and John R. Freeman. 1990. *Three Way Street: Strategic Reciprocity in World Politics*. Chicago, IL: University of Chicago Press.

Gowa, Joanne. 1999. *Ballots and Bullets: The Search for the Elusive Democratic Peace*. Princeton, N.J.: Princeton University Press.

Hegre, Håvard. 2004. "Size Asymmetry, Trade, and Militarized Conflict." *Journal of Conflict Resolution* 48(3):402–429.

Hoff, Peter D. 2005. "Bilinear Mixed Effects Models for Dyadic Data." *Journal of the American Statistical Association* 100:286–295.

Hoff, Peter D., Adrian E. Raftery and Mark S. Handcock. 2002. "Latent Space Approaches to Social Network Analysis." *Journal of the American Statistical Association* 97(469):1090–1098.

Huth, Paul K. and Todd L. Allee. 2002. *The Democratic Peace and Territorial Conflict in the Twentieth Century*. New York: Cambridge University Press.

Kenny, David A. 1981. Interpersonal Perception: A Multivariate Round Robin Analysis. In *Scientific Inquiry and the Social Sciences: A Volume in Honor of Donald T. Campbell*, ed. Marilynn B. Brewer and Barry E. Collins. San Francisco, CA: Jossey-Bass pp. 288–309.

Kenny, David A. 1994. *Interpersonal Perception: A Social Relations Analysis*. New York: Guilford Press.

Kenny, David A., Deborah A. Kashy and William L. Cook. 2006. *Dyadic Data Analysis*. New York: Guilford Press.

King, Gary. 2001. "Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Data." *International Organization* 55:497–507.

Lake, David. 1992. "Powerful Pacifists: Democratic States and War." *American Political Science Review* 86(1):24–37.

Leeds, Brett Ashley. 2003. "Do Alliances Deter Aggression? The Influence of Military Alliances on the Initiation of Militarized Interstate Disputes." *American Journal of Political Science* 47(3):427–439.

Lemke, Douglas and William Reed. 2001. "The Relevance of Politically Relevant Dyads." *Journal of Conflict Resolution* 45(1):126–144.

Li, Heng & Eric Loken. 2002. "A Unified Theory of Statistical Analysis and Inference for Variance Component Models For Dyadic Data." *Statistica Sinica* 12(2):519–535.

Malloy, Thomas E. and David A. Kenny. 1986. "The Social Relations Model: An Integrative Method for Personality Research." *Journal of Personality* 54(1):199–225.

Mansfield, Edward and Jack Snyder. 1997. "A Reply to Thompson and Tucker." *Journal of Conflict Resolution* 41:457–461.

Maoz, Zeev. 1996. *Domestic Sources of Global Change*. Ann Arbor, Michigan: University of Michigan Press.

Maoz, Zeev. 1998. "Realist and Cultural Critiques of the Democratic Peace: A Theoretical and Empirical Re-assessment." *International Interactions* 24(1):1–89.

Maoz, Zeev and Bruce M. Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946-1986." *American Political Science Review* 87(3):624–38.

Maoz, Zeev and Nasrin Abdolali. 1989. "Regime Types and International Conflict, 1816-1976." *Journal of Conflict Resolution* 33(1):3–35.

Morrow, James D. 2002. International Conflict: Assessing the Democratic Peace and Offense-Defense Theory. In *Political Science: The State of the Discipline*, ed. Ira Katznelson and Helen V. Milner. New York: W. W. Norton and Company pp. 172–196.

Mousseau, Michael. 1997. "Democracy and Militarized Interstate Conflicts, 1816-1992." *Journal of Peace Research* 34(1):73–87.

Oneal, John and Bruce M. Russett. 1999. "The Kantian Peace: The Pacific Benefits of Democracy, Interdependence, and International Organization." *World Politics* 52(1):1–37.

Oneal, John and Bruce M. Russett. 2005. "Rule of Three, Let It Be: When More Is Better." *Conflict Management and Peace Science* 22(4):293–310.

Oneal, John R. & Bruce M. Russett. 1997. "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950-1985." *International Organization* 41(2):267–295.

Raknerud, Arvid and Håvard Hegre. 1997. "The Hazard of War: Reassessing the Evidence for the Democratic Peace." *Journal of Peace Research* 34(4):385–404.

Ray, James. 1998. "Does Democracy Cause Peace." *Annual Review of Political Science* 1:27–46.

Ray, James Lee. 1995. *Democracy and International Politics: An Evaluation of the Democratic Peace Proposition*. Columbia, S.C., USA: University of South Carolina Press.

Richardson, Lewis F. 1960. *Arms and Insecurity*. Chicago and Pittsburgh, PA: Quadrangle/Boxwood.

Rosato, Sebastian. 2003. "The Flawed Logic of Democratic Peace Theory." *American Political Science Review* 97(4):585–602.

Rosato, Sebastian. 2005. "Explaining the Democratic Peace." *American Political Science Review* 99(3):467–472.

Rousseau, David, Christopher Gelpi, Dan Reiter, and Paul Huth. 1996. "Assessing the Dyadic Nature of the Democratic Peace, 1918-1988." *American Political Science Review* 90(3):512–33.

Rummel, Rudolph J. 1983. "Libertarianism and International Violence." *Journal of Conflict Resolution* 27:27–71.

Rummel, Rudolph J. 1985. "Libertarian Propositions on Violence within and between Nations." *Journal of Conflict Resolution* 29:419–455.

Russett, Bruce M. and John Oneal. 2001. *Triangulating Peace: Democracy, Interdependence, and International Organizations*. New York: W.W. Norton.

Russett, Bruce M., John Oneal and Michael Berbaum. 2003. "Causes of Peace: Democracy, Interdependence, and International Organizations, 1885-1992." *International Studies Quarterly* 47(3):371–393.

Russett, Bruce M., John R. Oneal and David R. Davis. 1998. "The Third Leg of the Kantian Tripod for Peace: International Organizations and Militarized Disputes, 1950-1985." *International Organization* 52:441–468.

Savage, I. Richard. 1957. "Nonparametric Statistics." *Journal of the American Statistical Association* 52:331–344.

Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy*. New York: Cambridge University Press.

Signorino, Curtis. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 92(2):279–298.

Slantchev, Branislav L., Anna Alexandrova, and Erik Gartzke. 2005. "Probabilistic Causality, Selection Bias, and the Logic of the Democratic Peace." *American Political Science Review* 99(3):459–462.

Smith, Alastair. 1999. "Testing Theories of Strategic Choice." *American Journal of Political Science* 43(4):1254–1283.

Thompson, William R. and Richard M. Tucker. 1997. "A Tale of Two Democratic Peace Critiques." *Journal of Conflict Resolution* 41:428–54.

Warner, R., David A. Kenny and Michael A. Stoto. 1979. "A New Round Robin Analysis of Variance for Social Interaction Data." *Journal of Personality and Social Psychology* 37:1742–1757.

Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Westveld, Anton. 2007. Statistical Methodology for Longitudinal Social Network Data. Ph. D. thesis. University of Washington, Seattle, Washington, USA.

Wong, George Y. 1982. "Round Robin Analysis of Variance via Maximum Likelihood." *Journal of the American Statistical Association* 77(380):714–724.

Appendix A. Data

All the data employed in this study, as well as more detailed results from our empirical estimations are available on the replication archive for this project.

A.1. Country Level Data.

A.1.1. *GDP.* Data on the annual Gross Domestic Product were taken from Kristian Skrede Gleditsch's "Expanded Trade and GDP Data, version 4.1" (2002). These data in 1999 U.S. dollars are available at `http://privatewww.essex.ac.uk/~ksg/exptradegdp.html`.

A.1.2. *Democracy and Autocracy.* We use the annualized polity score (called Polity2), which ranges from -10 for highly authoritarian states to +10 for highly democratic societies, to gauge the domestic institutions in each country. These data are available from `http://www.cidcm.umd.edu/inscr/polity/`, with registration. Some of these data were updated with information on so-called "micro-states" from the separate database maintained at the University of California at San Diego by Kristian Skrede Gleditsch (2003) *Modified Polity P4 and P4D Data, Version 1.0.*, available at `http://privatewww.essex.ac.uk/~ksg/Polity.html`.

A.1.3. *Population Data.* Data on population were taken from Kristian Skrede Gleditsch's population data archive, "Expanded Population Data," University of Essex.

A.2. Dyadic Data.

A.2.1. *Militarized Interstate Disputes.* We employed the Militarized Interstate Disputes database (version 3.02) which is maintained by Faten Ghosn and D. Scott Bennett, and described in *Codebook for the Dyadic Militarized Interstate Incident Data, Version 3.0, 2003.* These data and the documentation are available at `http://cow2.la.psu.edu/`. The dependent variable in this study is the existence of a militarized dispute between any two pairs of countries in a given year, coded as a 1. Senders are those countries designated by the MID data to have first initiated action.

A.2.2. *Dyadic Democracy.* The product of the polity score for any given pair of countries defines the jointly democratic characteristics of regime types in the two countries.

A.2.3. *Trade.* Imports were taken from Kristian Skrede Gleditsch's "Expanded Trade and GDP Data, version 4.1" (2002). These data are available at `http://privatewww.essex.ac.uk/~ksg/exptradegdp.html`. The trade data are used in billions of current year U.S. dollars.

A.2.4. *International Governmental Organizations.* We used the IGO data on international governmental organizations with at least three independent states as members (Version 2.1) maintained by Jon Pevehouse and Timothy Nordstrom and described in Jon Pevehouse, Timothy Nordstrom, and Kevin Warnke, "Intergovernmental Organizations, 1815-2000: A New Correlates of War Data Set, 2003." These data are at `http://cow2.la.psu.edu/`.

A.2.5. *Distance.* Distance was calculated using the Haversine formula with data on latitude and longitude of capital cities taken from the world.cities database maintained as part of the `maps` package in the *R* statistical programming package. These are available from cran.r-project.org. Distance was calculated in 1000s of Kilometers.

A.3. **Countries Analyzed.** We used all the data available to us for each year, which resulted in a superset of 165 countries over the period from 1950-2000: Afghanistan, Albania, Algeria, Angola, Argentina, Armenia, Australia, Austria, Azerbaijan, Bahrain, Belgium, Benin, Burkina Faso, (Upper Volta), Bhutan, Belarus, Bangladesh, Bolivia, Botswana, Brazil, Burundi, Bulgaria, Cambodia, Canada, Cameroon, Cote D'Ivoire, Central African Republic, Chad, Chile, China, Colombia, Comoros, Congo, Costa Rica, Croatia, Cuba, Cyprus, Czechoslovakia, Czech Republic, Denmark, Djibouti, Dominican Republic, Democratic Republic of Congo (Zaire), Democratic Republic of Vietnam, Ecuador, Egypt, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Finland, Fiji, France, Gabon, Gambia, German Democratic Republic, German Federal Republic, Ghana, Germany, Guinea-Bissau, Greece, Georgia, Guatemala, Guinea, Guyana, Haiti, Honduras, Hungary, India, Indonesia, Ireland, Iran, Iraq, Israel, Italy, Jamaica, Jordan, Japan, Kenya, Kuwait, Kyrgyz Republic, Kazakhstan, Laos, Latvia, Liberia, Lebanon, Lesotho, Libya Lithuania, Mauritania, Former Yugoslav Republic of Macedonia, Madagascar, Malaysia, Mauritius, Malawi, Mexico, Moldova, Mali, Mongolia, Morocco, Myanmar, Mozambique, Namibia, Nepal, New Zealand, Nicaragua, Nigeria, Niger, Norway, Netherlands, Oman, Pakistan, Panama, Paraguay, Peru, Philippines, Papua, New Guinea,

Poland, Portugal, People's Republic of Korea, Qatar, Republic of Korea, Romania, Russia, Republic of Vietnam, Rwanda, South Africa, El Salvador, Saudi Arabia, Senegal, Serbia, Sierra Leone, Singapore, Slovakia, Slovenia, Somalia, Spain, Sri Lanka , Sudan, Swaziland, Sweden, Switzerland, Syria, Tajikistan, Taiwan, Tanzania, Thailand, Turkmenistan, Togo, Trinidad and Tobago, Tunisia, Turkey, United Arab Emirates, Uganda, United Kingdom, Ukraine, Uruguay, United States of America, Soviet Union, Uzbekistan, Venezuela, Arab Republic of Yemen, Yemen, People's Republic of Yugoslavia (Serbia), Zambia, and Zimbabwe.

About the authors:

Michael D. Ward: Department of Political Science, University of Washington, Seattle, Washington, USA, 98195-3530
E-mail address: mdw@u.washington.edu

Randolph M. Siverson: Department of Political Science, University of California, Davis, One Shields Avenue, Davis, California, USA, 95616
E-mail address: rmsiverson@ucdavis.edu

Xun Cao: Department of Political Science, University of Washington, Seattle, Washington, USA, 98195-3530
E-mail address: xuncao@u.washington.edu