The Operationalization of Democracy and the Strength of the Democratic Peace: A Test of the Relative Utility of Scalar and Dichotomous Measures

Reşat Bayer and Michael Bernhard

The idea that democracies are less apt to engage in conflict with each other is a central finding in international relations. Yet the operationalization of democracy in this literature has been relatively unreflective. Since the mid-1990s the majority of studies have used Polity. In this paper we raise substantial concerns about its use, notably that there is a mismatch between conceptualization of democracy as a regime type and using an interval scale to measure it. If our contention is correct, we would expect to find that models that use a dichotomous coding should either provide different results from Polity or at minimum fit the data better. We then test this contention by comparing the results of tests of the democratic peace using Polity in its interval scalar form and several common dichotomous codings of democracy. The tests are supportive of the contention that dichotomous coding better captures the notion of “democracy.” At minimum we believe that findings using Polity should be verified for robustness using a dichotomous coding.

Introduction

The democratic peace has been one of the most influential and durable findings produced in international relations. The operation of the main independent variable in this set of findings, democracy, has used the Polity scale pioneered by Eckstein and Gurr. The preponderance of Polity in the literature on democracy and conflict was not always so stark. Influential earlier studies used dichotomous codings of democracy (Chan, 1984; Bremer, 1992; Ray, 1995) and a number of authors have argued for their continued use (Ray, 1995; Chan, 1997; Rioux, 1998). The shift toward Polity began with Maoz and Russett (1993). In this piece they used both scalar and dichotomous codings based on Polity. Since this time, Polity largely has crowded out the dichot-

omous codings used to produce earlier findings.¹ Its predominant position has been challenged even though past studies that have used other measures of democracy have reported different findings. For example, Buhaug (2005) replicates Bremer's (1992) influential "Dangerous Dyads" piece and finds that results for democracy are affected by whether Chan's measure or Polity is used.

We believe that there are several reasons why this consensus should be reexamined. First, there is a considerable literature on concept formation and operationalization which is highly critical of Polity. Second, there is a lively debate in comparative politics on how democracy should be operationalized, continuously or dichotomously. Given comparative's importance in the study of regimes in the discipline, international relations scholars should pay attention to this debate. Third, explicit tests that comparatively evaluate the performance of interval scalar and dichotomous codings of democracy or even parallel results that confirm similar findings using both operationalizations are relatively rare and less than definitive.

To advance this discussion, we frame a mainstream test of the impact of the measurement and operationalization of democracy on conflict. Our test is based on influential work by Russett and Oneal, but we substitute alternative codings of democracy for Polity. We then comparatively evaluate the performance of a base model using Polity to those which employ alternative codings to see if exclusive reliance on Polity is warranted. We find evidence that dichotomous codings perform better, leading us to advocate the parallel testing of findings on regime and conflict with both interval scalar and dichotomous operationalizations to assure the robustness of findings.

¹ Polity is cited with much greater frequency in international relations journals in the Social Sciences Citation Index's than binary codings (e.g. Chan, 1984; Doyle, 1986; Ray, 1995).

In the sections that follow we elaborate on some of the limitations of Polity that led us to question its unchallenged place in the literature on democracy and conflict. We then frame the test and discuss our alternative operationalizations of democracy. Finally, we report our findings, and then draw conclusions from those results.

Questions About Polity

Although Polity is most strongly associated with Tedd Gurr, Harry Eckstein, also had a very important role in its conceptual underpinning. Writing in the early 1970s, at the height of functionalism's influence in the discipline, Eckstein argued that authority relations should be central to political science. Rather than seeing politics as an equilibrium produced by the broad involvement of atomized individuals, Eckstein argued that politics was a group enterprise characterized by asymmetric power relations (1973). Eckstein's ideas were further elaborated in his collaboration with Gurr at the Princeton Workshop on Comparative Politics (Eckstein and Gurr, 1975).

Polity scores are composed of a set of measures derived from Eckstein's scheme for describing patterns of authority. The measure that is used in most conflict studies is actually not the Polity democracy score itself, but the so-called combined score (which subtracts the autocracy score from the democracy score) resulting in a single twenty-one interval scale that runs from -10 to 10. Most researchers consider that countries with a combined score higher than six to be democracies in a conventional sense (Marshall and Jaggers, 2002, 15).

The combined score includes four of the Eckstein/Gurr authority indicators -- competitiveness of political participation (parcomp -- five points), competitiveness of executive recruitment (xrcomp -- four points), openness of executive recruitment (xropen -- two points), and constraints on the chief executive (xconst -- seven points). The autocracy component of the com-

bined score also includes a component on how political competition is regulated (parreg – 2 points) (Gurr, Jagers, and Moore, 1991: 73-81; Marshall and Jagers 2002, 12-15). For some reason the kind of regulation associated with democratic polities (highly institutionalized, without strong factionalization or restrictions on types of action) is omitted.

Polity has been embraced by political scientists because it is readily available, is used extensively by others, and covers many countries over a long time. It is unquestionably one of the most successful long-term comprehensive quantitative variable construction projects in the history of discipline. Not only has it been embraced by the quantitative international relations community, but plays a role in comparative politics where it is utilized as a dependent, independent, or a control variable in many studies that try to establish the correlates of democratization and development. Despite the obvious convenience and utility of Polity as a measure, we believe that there are good reasons to question its universality in the quantitative study of conflict. We now turn to a more detailed discussion of the concerns introduced above that lead us to be skeptical of an exclusive reliance on Polity.

First, despite major criticism of Polity as a variable, it continues to be used exclusively and unreflectively by many scholars. We can only briefly review the most important of these concerns:

[A] Both Munck and Verkuilen (2002, 26) and Goertz (2005) point out that there seems to be no explicit rationale for the weighting of Polity's component parts other than the number of points on the scales developed by Eckstein and Gurr. And Gleditsch and Ward have determined that in the present scheme executive constraint (with a weight of seven points) almost single-handedly determines where countries lie on the Polity scale.

At the same time defining characteristics of democracy such as “the extent and character of popular participation in the selection of leaders, is either totally absent or relatively unimportant in determining the degree of democracy” (1997, 376-378). Goertz’s proposal for reweighting some of the components may make more sense, but it still does not solve the other problems discussed below.

[B] Polity is subject to the general criticism that Munck and Verkuilen make with regard to aggregative scalar measures of democracy: “they have not done much to test whether the lower levels of aggregation do tap into a unidimensional phenomenon, and thus whether aggregation can be carried out without forcing a multidimensional phenomenon into a common metric, a practice that weakens the validity of the resulting scores” (2002, 23). Gleditsch and Ward do much to show that while the upper part of the Polity scale does a reasonably good job of grouping democracies together, the middle range of the scale is anything but coherent. They note that there are ten different combinations of characteristics that will yield a score of six on the democracy scale. They provide concrete examples of five of these, including Spain (1895-1899), Taiwan (1991-94), United Kingdom (1837-79), Fiji in 1991, Napoleonic France and the Fifth Republic (1997, 368). Is there any theory of regimes under which it makes sense to group these states as sharing a single authority pattern? Each has a different score on the different components of the scale. In aggregating them they all happen to add up to the same number.

[C] The omission of the high end of the participation scale from Polity leads to an anachronism – the appearance of countries that are coded as fully democratic before universal

adult suffrage had emerged historically. Polity codes over 270 country-years with scores of 10,0 (as well another 175 with combined scores of eight or nine) between 1800 and 1919. For those not so familiar with the history of nineteenth century rule, one example may serve to point out the problems with this. The United States rates a score of 10,0 for several years in the 1840s despite legal slavery and the disenfranchisement of women. Such anachronistic codings represent between four and five percent of the observations for this time period. Many authors have noted how this can bias research both on democracy and on conflict by ignoring issues of gender and race (Paxton, 2000; Russett and Oneal, 2001; Caprioli and Boyer, 2001; Henderson, 1995; Johnson, 1999). We note that it also ignores issues of class disenfranchisement, which plays an important role in the struggles for democracy in both Latin America and Western Europe.

Second, Polity by no means enjoys the kind of monopoly it has in empirical conflict studies in comparative politics, the subfield of the discipline that has worked most extensively on political regimes. Its dominant conceptualization of regime is in fact based on work other than Eckstein's and there is a substantial debate over how democracy should be operationalized, continuously or dichotomously, among comparativists. Its dominant view of regimes is typological and not scalar (i.e. classificatory). In comparative politics, the study of regimes is characterized by a series of differentiations: traditional vs. modern forms of rule, within the latter – democracy versus dictatorship, and among various types of dictatorship. The early fundamental distinctions were drawn in classic works including Dahl's *Polyarchy* (1971) and Linz's fundamental work on modern dictatorship (1975). Regimes are not seen as something that can be arrayed along a single property scale but as a series of types, which are by no means differentiated by a single defining characteristic. The criteria that comparativists use to differentiate traditional from modern

rule, democracy from dictatorship, and types of dictatorship from each other are multidimensional and differ at each stage of the ordering.

Because regimes are seen as mutually exclusive types by a substantial portion of the community of comparativists, there is no consensus on how to best measure democracy. Bohlen stands out as the most passionate defender of a scalar measure of democracy, and has also contributed his own coding scheme (i.a. 1980; 1991). On the opposite side of this argument one finds Sartori (1970; 1987) and Przeworski et al. (2000) arguing for a dichotomous coding.

Most conflict scholars seem to ascribe to Bohlen's position. They assume democracy is unidimensional and can be scaled. Countries are more or less democratic in this world view. However if democracy is conceptualized as a regime type, Sartori's work (1987) on concepts would lead one to measure it differently. Here his distinction between *contraries* and *contradictories* is quite useful for thinking about concepts that are posed in opposition to each other, like democracy and dictatorship. Contraries stand at opposite ends of a scale, but there are other attributes that lie between them (e.g. large versus small). Contradictories stand in opposition to each other exhaustively and exclusively (e.g. alive versus dead). Sartori uses this distinction to differentiate "object concepts" and "property concepts." When the semantic field of a concept is populated by distinct multidimensional types (literally a set of objects rather than a quality that many observations share to a lesser or greater degree), a dichotomous coding that distinguishes that object from others is preferable to scaling which only makes unidimensional distinctions.

From the perspective of a comparativist who sees democracy as a type, Polity compresses something that is multidimensional into one dimension. From this point of view a regression using Polity in this way essentially is placing the same coefficient in front of all of Polity's com-

ponent parts.² Thus a large number of comparativists who study regime change do not rely on Polity but instead code democracy (and other regime types) dichotomously to understand why some countries experience transition (i.a. Gasiorowski, 1995; Przeworski et al., 2000; Bernhard, Reenock, and Nordstrom 2001; Boix, 2003; Geddes, 2003).

The debate in comparative politics is far from fully resolved. Collier and Adcock (1999), whose contribution summarizes the most important issues in the debate, take a middle position arguing that the choice should be made pragmatically on the basis of theory and concept. If we follow this advice, those who think of democracy as something that exists in degrees will continue to use interval scalar measures, whereas those who think of democracy as a regime type will continue to use dichotomous codings. Munck and Verkuilen suggest that such conceptual and theoretical considerations should also be subject to testing (2002, 25). It is precisely such considerations that motivate us to engage in this exercise.

Third, the kind of explicit tests that comparatively evaluate the performance interval scalar and dichotomous codings of democracy or even parallel results using both kinds of operationalizations are relatively rare and less than definitive. Even at the key junctures when researchers began to rely on Polity, the results it produced were fragile depending on whether one used it as a scale or dichotomized it. In Maoz and Russett (1993) democracy's effect on conflict disappears when they use a scalar coding, and is robust with the dichotomous coding. Similarly in one of the studies that introduced trade dependence as an important control variable, the effect of democracy on peace disappeared when democracy was measured continuously using Polity. But when a dichotomized version was used, the effect of democracy on peace came back strongly (Oneal and Ray, 1997).

² I heard this reservation first expressed by William Clark in a discussion a few years ago.

The one explicit comparative test of different operationalizations of democracy is Elkins (2000). One of his tests is drawn from the democratic peace literature, based on a dyadic study by Rousseau, Gelpi, Reiter, and Huth (1996). Elkins replicates the original findings which show that level of democracy of both sides of a conflict is significant and negatively correlated to conflict initiation using an interaction term composed of Polity scores. When Elkins attempts to replicate this using a dichotomous coding based on Polity, the interaction is not significant.

Elkins argues that this finding demonstrates “superior construct validity” of graded measures. However, there are aspects of this test that might lead one to question the definitiveness of this test. First, the sample employed is small and selective, looking at only those 303 dyads where conflict occurred in 272 international crises (n=606). Very few of these dyads are democratic on both sides and the sample excludes all dyads in which two democracies did not enter into conflict. Second, the model employed by Rousseau et al. is fairly lean with only a few controls. Notably absent from their work is trade interdependence which is well-known to undermine findings on the democratic peace thesis in regressions in which regime is operationalized using Polity in its scalar form (see above). Finally, it is unclear whether the failure to find results in this test is a function of dichotomization or the dichotomization of Polity. Given our reservations about the conclusions that Elkins draws from this particular replication, we will try to frame a more definitive test based on a more mainstream publication from the literature on democracy and conflict, certainly one that relies on a larger sample of countries. We will also comparatively test several operationalizations of democracy.

Framing a More Definitive Test

If the difference between dichotomous and interval scalar measures of democracy are strong, we would expect to find consequential differences in the way the two measures perform in otherwise

identical regressions. Two kinds of findings would be indicative of that. First if the two measures yielded different results in terms of the significance of variables and the magnitude of key coefficients this would indicate that the particular way in which democracy was coded is critical to reliable findings. In such cases we would also expect the results generated by the better measure to fit the data better. Second, even if both measures produced similar results, if one coding provided a noticeably better fit, it would indicate that one of the other measures was over- or underestimating the impact of democracy on conflict.

Reviewing the entire body of literature on democracy and conflict is beyond the scope of this study. For our comparative test we have constructed a series of models that are in line with the contemporary mainstream literature. In the research design and regressions that follow, we will retest the relationship between conflict and democracy with Polity in its interval scalar form as well as different dichotomous measures of democracy and see how they perform differently. Before presenting our results, we present a fuller picture of the elements of our research design.

Research Design

Dependent Variables

We start by replicating a model (Table A5.1) from *Triangulating Peace* from Russett and Oneal (2001). This was important work as it examined the factors associated with the Kantian peace, i.e., democracy, economic interdependence, and intergovernmental organizations, and found that they all decreased involvement in disputes. The dependent variable is annual involvement in militarized disputes for politically relevant dyads from the Militarized Interstate Dispute (MID) data set of the Correlates of War Project (Jones, Bremer, and Singer, 1996). This data set captures not only wars but other forms of violence including threat of force. It is the most

comprehensive data set of its kind temporally and is relied upon extensively in the empirical study of interstate conflict.

Since Russett and Oneal (2001) rely upon politically relevant, nondirected dyads and examine involvement in disputes, we decided that we would perform a second set of analysis that relies upon directed dyads and examines initiation of disputes. Bennett and Stam (2004) point out that analysis that relies on directed dyads has several advantages over nondirected dyads, most notably the ability to model conflict initiation. Analysis using nondirected dyads discards considerable information and the reduction in sample size with the employment of nondirected dyads is not substantial enough to have a profound impact on the standard errors. Thus, we also conduct directed dyad analyses where our dependent variable is initiation of a MID dispute. We rely upon the version of the MID data set updated by Ghosn, Palmer, and Bremer (2004) which extends the analysis up to 2000. Following Bennett and Stam (2004) we drop ongoing disputes and joining instances from our analyses. Finally, since there are issues of independence in directed dyads we discard the possibility of side B initiating a dispute in a year against side A if A has already initiated a dispute with B. We analyze politically relevant dyads (where at least one state is a major power or the pair is geographically contiguous) as well as all directed dyad-years (which increases the number of observations tenfold).

Method

Since our dependent variable is binary, we use logit regression to analyze the data. When replicating Russett and Oneal (2001, 311), we follow their estimation technique and rely upon logistic analyses using the General Estimating Equation (GEE) allowing for an AR1 process. The dependent variable is lagged one year and robust standard errors are used.

In the directed dyad analyses, we use logistic regression because here too the dependent variable is binary. In the directed dyad models we present here, all variables are at time t . We did not find substantive differences when we estimated models with lagged independent variables.

Since we will be relying upon several different measures of democracy it will be essential to gauge how the competing variables perform. We rely upon the Bayesian Information Criteria (BIC) as a goodness-of-fit measure.³

Main Independent Variables – Scalar Versus Dichotomous Codings of Democracy

We operationalize democracy using both Polity (Marshall and Jaggers, 2002), and the dichotomous codings developed by Alvarez, Cheibub, Limongi, and Przeworski (1996, henceforth ACLP), Bernhard, Nordstrom, and Reenock (2001, henceforth BNR), and Doyle (2002). ACLP and BNR are event history data sets used to understand the survival of democratic regimes, whereas Doyle is a more recent conflict data set.⁴ Doyle (1986) does not explicitly operationalize democracy, but “liberal regime,” which is composed of four elements –

³ The formula for BIC is $-2\log(L) + \log(N)K$; where L is the log-likelihood, K is the number of parameters in model, and N is the sample size. We also calculated the Akaike Information Criterion (AIC) for all our models as well. The AIC values were almost identical so we report the BIC values only.

⁴ We use the version of ACLP updated by Cheibub and Gandhi (2004), covering the period 1946-2002. We use a version of BNR that has been updated to cover all democracies from 1919 through the end of 2005. We use Doyle (2002) version 1.0 which covers the period 1816-2000 as downloaded from Eugene version 3.2 (Bennett and Stam, 2000).

private market economy, national sovereignty, judicial rights, and representative “republican” government.⁵

Russett and Oneal (2001) used the lower of the Polity scores in the dyad. Their rationale was that the less democratic side is the weak link in the dyad as it is less constrained in using force and, thus, it sets the tone for violence in the dyad. Russett and Oneal found that as the less democratic side’s score increased, the risk of dispute involvement diminished. We examine what happens when we remove their measure for democracy and add measures for whether both sides were democratic as well as whether only one side was democratic with the dichotomous measures (ACLP, Polity binary, BNR, and Doyle).

It is important to include a measure for joint democracy as it is under this condition that the democratic peace should be at its strongest. The other measure enables us to see whether any of the measures of democracy will be able to capture a significant relationship between a single democracy and conflict involvement. Less work has been done on the properties of dyads in which one side is democratic and the other is a dictatorship. Oneal and Russett (1997) have argued that such dyads will be particularly prone to conflict, the so-called “cat-and-dog effect.” For the models based on the continuous Polity measure, we include instead the less democratic side’s score in place of joint democracy and the higher democracy score in the dyad in place of single democracy following Russett and Oneal (2001). In the directed dyad analyses, we include separate measures of democracy for the initiator and target in the conflict, enabling us to see if

⁵ There is only partial overlap between Doyle’s notion of liberal regime and polyarchy-oriented measures devised by ACLP and BNR. The closer one comes to the present, the closer they are. However, in the nineteenth century there were liberal regimes which do not meet conventional standards of modern democracy.

democracy makes states more aggressive or more likely to be targeted by aggressors. We also include a dichotomous measure of whether both sides were democratic.

We run our models over two different time periods (1919-2000 and 1950-2000). This allows us to examine whether the relationship between conflict and democracy is sensitive to time. Furthermore ACLP only covers the shorter time period. We not uses Polity in its interval scalar form but also operationalized it dichotomously. Countries with scores of six or higher are considered democratic.⁶

In order to facilitate the task of comparing how the data sets perform, we confine ourselves to identical samples for all data sets and our analyses use the same dyad-years. This ensures that measurement rather than sample is responsible for any differences we observe in results. This leads to a diminution of sample of over 10,000 observations for the BNR data set in comparison to Polity in the analysis that follows.⁷

⁶ We used a version of Polity2 where normal range scores replaced the standardized authority scores (-77, -88) (Marshall and Jaggers, 2002). This reduces the missing values in our sample. We examined whether our results were affected by our choice of Polity by running our models without these cases. The results were same as those reported here.

⁷There is fairly substantial variation in the degree to which our measures for joint democracy are correlated in our directed dyad data set. The joint democracy measure correlation varies from .67 (Polity and Doyle) and .82 (ACLP and BNR) for politically relevant dyads between 1950-2000. We examined the cross-tabulation of the measures in relation to MID onset. The measures diverge in the coding of several thousand cases. For example, there are 69 cases where BNR categorizes a democracy as initiating a dispute and Polity does not, and 32 cases where Polity reports a MID initiation but BNR does not. These differences suggest that measures of

Control Variables

In the models in which we utilize Russett and Oneal's (2001) design we include their full set of control variables, including alliances, the balance of dyadic power, contiguity, distance, trade, minor powers, and membership in intergovernmental organizations. They find that the risk of militarized interstate disputes were negatively and significantly correlated with these factors in the period from 1886-1992.

In the directed dyad analyses, we present models with fewer control variables. Our focus is on the conceptualization and measurement of democracy in this study and we want to avoid the possibility of intervening variables (Ray, 2005). Controls in these models include the ratio of initiator state's capabilities to the total in the dyad (using the composite capabilities index developed by Singer, Bremer, and Stuckey, 1972) and the natural logarithm of distance between capitals (and ports in the case of Russia and USA) (Small and Singer, 1982). These were also generated through EUGene (Bennett and Stam, 2000). We also add (but do not report) four lagged-time variables to correct for autocorrelation as suggested by Beck, Katz and Tucker (1998) in these analyses.

Results

The first model in Table 1 replicates Russett and Oneal (2001) for the period from the end of World War I until the end of the Cold War for politically relevant dyads. Like them we find that the lower of the two Polity scores in the dyad is negative and significant indicating that as the less democratic country in the dyad becomes more democratic, the risk of dispute involvement

democracy are not interchangeable and that BNR (or the other measures for that matter) are not just capturing a more stringently coded subset of countries Polity codes as democratic.

decreases. Like them we also find that the higher of the two Polity scores in the dyad is insignificant.

[Table 1]

In the other three models in the first panel we run the same test but with dichotomized democracy variables (Polity, BNR and Doyle). In all three joint democracy decreases the risk of involvement in conflict. In none of the models does the presence of just one democracy produce a significant finding. These models provide consistent and strong support for the joint democracy thesis and little for the ‘cats-and-dogs’ thesis. Thus, there is strong harmony across the different operationalizations of democracy in the first series of models.

When this analysis is run only on the period from 1950-1992 there are a number of changes. In the model that uses Polity in its interval scalar form, once again as the less democratic side becomes more democratic, the chance of dispute involvement decreases. However, in this model the high democracy variable is positive and statistically significant, indicating that more democratic states are more likely to become involved in disputes with less democratic ones. During this time period with this operationalization of democracy, there is now some support for the ‘cats-and-dogs’ thesis.

The dichotomized codings of democracy do not produce consistent results even using identical samples. In the BNR model joint democracy is not statistically significant. This is odd given that it was negative and statistically significant with the previous more extensive sample. This finding underscores the hazards of relying upon on one data set. ACLP, Doyle and the dichotomized version of Polity all support the proposition that joint democracy reduces disputes. In all of the models using a dichotomous operationalization of democracy there is no support for the ‘cats-and-dogs’ argument. In no case does the presence of just one democracy in a dyad lead

to a statistically significant increase in disputes. Russett and Oneal's finding seems to be a product of both the interval scalar operationalization of Polity and the time frame used.

While we had consistent results across operationalizations for the period 1919-92, the results for 1950-1992 varied across operationalizations. The inconsistency of results both across operationalizations and time frames, led us to wonder if the problem lay in our use of politically relevant nondirected dyads. We thus decided to perform further tests using all directed dyads to see if the incorporation of additional information might resolve some of these problems.

[Table 2]

In Table 2, the first five models include all directed dyads for the period 1950-2000. In order to facilitate comparison, the cases were kept the same across models. Again we have very consistent results across models. In all five joint democracy reduces the risk of dispute initiation and democracies are more likely to be targeted in disputes. Democracies are more likely to initiate disputes with all the measures but the Polity interval measure. When we look at the Bayesian Information Criteria to assess how well the data fit these models, Doyle's measure (BIC: -1582.350), does best, followed by BNR, (BIC: -1570.801), ACLP (BIC -1557.471), the dichotomized Polity (BIC: -1535.511), and the standard Polity (BIC-1417.351) measure.⁸

The substantive effect of the different results of each model can be gauged by a marginal effects analysis of the impact of the different operationalizations. Keeping all the other variables

⁸ We also estimated a model where the interaction of the two scalar Polity variables was used. This resulted in a variable ranging from -100 to +100. This is not a common usage of Polity. The results were similar to the Polity interval model in Table 2. The BIC score for this model was -1543.720, and in the marginal effects analysis joint democracy reduced the risk of conflict initiation by 55.1 percent.

in the model at their base or at their mean and varying joint democracy from 0 to 1 results in reductions of the probability of involvement in a dispute from of 36 percent (Polity interval scalar model) to 88.6 percent (Doyle). Between these two lay the Polity binary (75.6 percent), ACLP (87.1 percent), and BNR (87.8 percent). The impact of joint democracy for the three binary operationalizations is strikingly close. The interval scalar version of Polity seems to strongly underestimate the effect of joint democracy, and even the dichotomized form seems to underestimate it slightly.

We also ran directed dyadic models where we limited the sample to politically relevant dyads. These are reported in the second panel of Table 2. In these tests, joint democracy again decreased the risk of dispute initiation across all the models. However, for both dichotomous operationalizations of Polity and ACLP the limitation of the sample to politically relevant dyads led to an important change – democracies were no longer more likely to initiate a dispute. The finding for BNR and Doyle are stable; democracies remained more likely to initiate disputes. Also we consistently found that democracies were more likely to be targeted.⁹

⁹ We varied this model specification in several ways and found substantially the same things with some minor exceptions. For example, for politically relevant dyads for 1919 onwards, the democratic initiator variable was negative and significant only with Polity interval whereas the binary variables were all insignificant. In addition, if we omitted the joint democracy variable, democracies were less likely to initiate disputes except for Doyle's measure when the sample was confined to politically relevant dyads after 1950. In the BIC comparisons for 1919 onwards, the BNR model provides the best fit when all dyads were considered but the Polity binary model fit best when politically relevant cases were used. In both of these tests, the Polity interval model provided the worst fit.

We also calculated the Bayesian Information Criteria to check the fit of the politically relevant, directed dyad models. The best fit was provided by the binary Polity model (BIC: -838.838), followed by BNR (BIC: -830.490), ACLP (BIC: -821.129), Doyle (BIC: -819.675), and Polity interval (BIC -791.435).¹⁰ The calculation of marginal effects showed that all the binary operationalizations reduced the risk of MID initiation by around eighty percent. With the Polity interval scalar variable the reduction was only 63.7 percent. Thus, we again found that the interval scalar operationalization underestimates just how effective democracy is in reducing conflict.

In terms of the way the control variables perform, the findings in Table 1 are identical to those of Russett and Oneal; all of their control variables significantly decrease the risk of dispute involvement. Our results for their controls are also strikingly consistent across all measures of democracy. The control variables in Table 2 are congruent with major studies in the literature that rely on directed dyads (e.g. Bennett and Stam, 2004). Distance decreases the risk of dispute initiation. As the initiator's capabilities exceed those of the target, the risk of a dispute increases.

The results of our tests are congruent with our expectations -- different measures of democracy measures lead to different results in similarly structured tests. The measure of democracy adopted in research designs has ramifications for the results generated. The different operationalizations we tested produced different findings in both directed and in non-

¹⁰ We again examined a model where the interaction of the two scalar Polity variables was used. The results were similar to the Polity interval model discussed in the text with one important exception: the target variable is not negative and significant. The joint democracy variable reduced the risk of initiation only by 52.6 percent. The goodness of fit measure for BIC was as follows -850.372. Thus, with the politically relevant dyads, the fit was much better for Polity.

directed analyses, across different time frames, and on the basis of whether the sample was composed of politically relevant dyads. Thus even before an analysis of goodness of fit or the impact of democracy variables, we saw considerable differences. The BIC tests showed that the binary operationalizations of democracy on balance provided a better fit than the interval scalar operationalization based on Polity. Also, the different operationalizations led to different estimations of the impact of joint democracy. The impact of the dichotomous operationalizations had a greater impact in reducing conflict than the interval scalar version of Polity, leading us to wonder whether a substantial part of the literature has underestimated the impact of democracy on peace.

Conclusions

The purpose behind our exercise in comparative variable testing was to evaluate the near-monopoly that Polity enjoys in the large-n empirical literature on conflict. We showed how the literature on regimes and important debates on the nature of democracy in comparative politics, made a plausible case for the measurement of democracy dichotomously rather than an interval scale. The comparative tests that we framed strongly support our skepticism about the near total reliance on Polity in conflict studies. Across most of our tests the dichotomous operationalizations of democracy that we used fit the data better than Polity, and yielded more stable results across time periods and using different model specifications. Our skepticism was most strongly verified on the use of Polity in its interval scalar form. When it was dichotomized its findings were closer to those obtained using the other data sets. In fact, there was one

instance where it performed marginally better than ACLP, BNR, and Doyle in terms of the goodness of fit measures we used.

Our findings also provide strong evidence for the robustness of the democratic peace. All five operationalizations provided strong support for its basic structure, even if the support provided by Polity, seemed also to underestimate its impact. Here we think the operationalization of joint democracy by the use of the lower Polity score, introduces distortion through measurement error into the tests (e.g. assigning different degrees of democracy to non-democratic regimes). In addition, our results on the propensity of democratic initiators and targets in directed dyad models proved to be more fragile than our simple non-directed dyadic tests. Finally, we found that the ‘cats-and-dogs’ effect hypothesized by Russett and Oneal appeared only with the use of the higher Polity score, whereas with all of the other data sets there was not a statistically significant relationship. This finding may be an artifact of this particular operationalization.

The ultimate lesson that we draw confirms one of the nostrums of contemporary social science practice: one of the surest ways to increase confidence in findings is to use alternative measures and duplicate results. The results we present here again confirm that the democratic peace is no artifact of data. Our results ultimately lead to question the overreliance on Polity, particularly in its interval scalar form, as sole measure of democracy in the study of conflict. Comparative politics has developed a complex literature on regimes that suggests that looking at democracy as a single scale rather than as a complex type is problematic. It is necessary to reevaluate with some regularity whether existing data sets and practices of operationalization really capture the concepts that underlie our hypotheses when we employ them as variables. At

minimum, such questions mean that where possible, results based on Polity should be verified with tests using alternative, especially dichotomous, measures.

References

- Alvarez, Michael, José Antonio Cheibub, Fernando Limongi and Adam Przeworski. 1996. "Classifying political regimes". *Studies in Comparative International Development* 31 (Summer): 1-37.
- Beck, Nathaniel, Jonathan N. Katz and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42:1260-1288.
- Bennett, D. Scott and Allan C. Stam. 2000. "EUGene: A Conceptual Manual." *International Interactions* 26: 179-204. <<http://www.eugenesoftware.org>>.
- . 2004. *The Behavioral Origins of War*. Ann Arbor: University of Michigan Press.
- Michael Bernhard, Timothy Nordstrom, and Christopher Reenock. 2001. "Economic Performance, Institutional Intermediation, and Democratic Survival," *Journal of Politics* 63:775-803.
- Bollen, Kenneth A. 1980. "Issues in the Comparative Measurement of Political Democracy." *American Sociological Review* 44: 572-587.
- . 1991. "Political Democracy: Conceptual and Measurement Traps." In *On Measuring Democracy*, Alex Inkeles, ed. New Brunswick, Transaction. Pp. 3-20.
- Boix, Carles, 2003. *Democracy and Redistribution*. Cambridge, Cambridge University Press.
- Bremer, Stuart A. 1992. "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816-1965." *Journal of Conflict Resolution* 36: 309-341.
- Caprioli, Mary and Mark Boyer. 2001. "Gender, Violence, and International Crisis." *Journal of Conflict Resolution* 45: 503-518.
- Buhaug, H. 2005. "Dangerous Dyads Revisited: Democracies May Not Be That Peaceful After All." *Conflict Management and Peace Science* 22: 95-111.
- Chan, Steve. 1984. "Mirror, Mirror on the Wall...Are the Free Countries More Pacific?" *Journal of Conflict Resolution* 28: 617-648.
- . 1997. "In Search of Democratic Peace: Problems and Promise." *Mershon International Studies Review* 41(2): 59-91.
- Cheibub, José Antonio, and Jennifer Gandhi. 2004. "Classifying political regimes: a six-fold measure of democracies and dictatorships". Prepared for the annual meeting of the American Political Science Association, Chicago, September 2004.

Collier, David and Robert Adcock. 1999. "Democracy and Dichotomies: A Pragmatic Approach to Choices about Concepts," *Annual Review of Political Science* 2:537-565.

Dahl, Robert. 1971. *Polyarchy*. New Haven, Yale University Press.

Doyle, Michael W. 1986. "Liberalism and World Politics," *The American Political Science Review* 80 :1151-1169.

Doyle, Michael W. 2002. "Ideologies and Politics: Liberal Democracy and National Dictatorship in Peace and War." In *War and Peace in the 20th Century and Beyond*, edited by Geir Lundestad and Olav Njolstad. Singapore: World Scientific Publishing Company.

Eckstein, Harry. 1973. "Authority Patterns: A Structural Basis for Political Inquiry," *The American Political Science Review* 67: 1142-1161.

Eckstein, Harry and Ted Robert Gurr. 1975. *Patterns of Authority: A Structural Basis for Political Inquiry*. New York, Wiley-Interscience.

Elkins, Zachary. 2000. "Gradations of Democracy? Empirical Tests of Alternative Conceptualizations," *American Journal of Political Science* 44:293-300.

Gasiorowski, Mark J. 1995. "Economic Crisis and Political Regime Change: An Event History Analysis," *American Political Science Review* 89(4):882-897.

Geddes, Barbara, 2003. *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics*. Ann Arbor, University of Michigan Press.

Ghosn, Faten, Glenn Palmer, and Stuart A. Bremer. 2004. "The MID3 Data Set, 1993-2001: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 21: 133-154 .

Gleditsch, Kristian S., and Michael D. Ward. 1997. "Double Take: A Reexamination of Democracy and Autocracy in Modern Politics," *The Journal of Conflict Resolution* 41:361-83.

Goertz, Gary. 2005. *Social Science Concepts: A Users' Guide*. Princeton, Princeton University Press.

Gurr, Ted Robert, Keith Jagers, and Will Moore. 1991. "The Transformation of the Western State, the Growth of Democracy, Autocracy, and State Power Since 1800." In *On Measuring Democracy*, Alex Inkeles, ed. New Brunswick, Transaction. Pp. 69-104.

Henderson, Errol. 1995. *Afrocentrism and World Politics*. Westport, CT: Praeger.

Johnson, Ollie. A., III. 1999. "Pluralist Authoritarianism in Comparative Perspective: White Supremacy, Male Supremacy, and Regime Classification," Pp. 116-136 in Georgia A. Persons, ed. "Race and Ethnicity in Comparative Perspective," *National Political Science Review* 7:116-136.

Jones, Daniel M., Stuart A. Bremer, and J. David Singer. 1996. "Militarized Interstate Disputes, 1816-1992: Rational, Coding Rules, and Empirical Patterns." *Conflict Management and Peace Science* 15:163-212.

Linz, Juan J. 1975. "Totalitarian and Authoritarian Regimes," In *Handbook of Political Science*, Fred I. Greenstein and Nelson Polsby, (eds). Reading, MA, Addison-Wesley. Volume 3:175-411.

Maoz, Zeev and Bruce Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946-1986." *American Political Science Review* 87: 624-638.

Marshall, Monty and Keith Jaggers. 2002. *Polity IV Data set Users' Manual*. www.cidcm.umd.edu/inscr/polity. Accessed: 12/23/05.

Munck, Gerardo L. and Jay Verkuilan. 2002. "Conceptualizing and Measuring Democracy," *Comparative Political Studies* 35: 5-34.

Oneal, John R., and Bruce Russett. 1997. "The Classical Liberals were Right: Democracy, Interdependence, and Conflict, 1950-1985." *International Studies Quarterly* 41: 267-293.

Oneal, John R. and James Lee Ray. 1997. "New Tests of the Democratic Peace: Controlling for Interdependence, 1950-85," *Political Research Quarterly* 50:751-775.

Paxton, Pamela. 2000. "Women's Suffrage in the Measurement of Democracy: Problems of Operationalization," *Studies in Comparative International Development* 35: 92-110.

Przeworski, Adam, Michael E. Alvarez, Jose Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development, Political Institutions and the Well-Being of the World, 1950-1990*. Cambridge, Cambridge University Press.

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

Ray, James Lee. 2005. "Constructing Multivariate Analyses (of Dangerous Dyads)." *Conflict Management and Peace Science* 22: 277-292.

Rioux, Jean-Sebastien. 1998. "A Crisis-Based Evaluation of the Democratic Peace Proposition." *Canadian Journal of Political Science* 31:263-283.

Rousseau, David L.; Christopher Gelpi; Dan Reiter; and Paul K. Huth. 1996. "Assessing the Dyadic Nature of the Democratic Peace, 1918-88," *The American Political Science Review* 90:512-533.

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

Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics," *American Political Science Review* 64:1033-1053.

_____. 1987. *The Theory of Democracy Revisited*. Chatham, NJ, Chatham House.

Singer, J. David, Stuart A. Bremer, and John Stuckey. 1972. "Capability Distribution, Uncertainty and Major Power War, 1820-1965," in *Peace, War and Numbers*, ed. B. M. Russett. Beverly Hills, CA: Sage Publications, 19-48.

Small, Melvin, and J. David Singer. 1982. *Resort to Arms: International and Civil Wars, 1816-1980*. Beverly Hills: Sage Publications.

Table 1. Replication of Russett and Oneal (2001) with Different Measures for Democracy

	1919-1992 Politically Relevant				1950-1992 Politically Relevant				
	Polity interval	Polity dichotomous	BNR	Doyle	Polity interval	Polity dichotomous	BNR	Doyle	ACLP
High Polity	0.012 [0.010]				0.024* [0.012]				
Low Polity	-0.068** [0.010]				-0.058** [0.013]				
Single Democracy/Liberal		-0.09 [0.149]	0.005 [0.140]	-0.096 [0.145]		-0.107 [0.192]	0.212 [0.167]	0.209 [0.177]	0.062 [0.172]
Joint Democracy/Liberal		-1.270** [0.221]	-0.600* [0.248]	-1.376** [0.214]		-1.032** [0.255]	-0.044 [0.287]	-0.834** [0.262]	-0.704** [0.240]
Lower dependence	-45.533* [18.422]	-53.822** [19.710]	67.333** [21.461]	-43.292* [18.285]	-41.486+ [23.158]	-39.182+ [21.522]	-43.727+ [22.391]	-64.545* [28.038]	-41.825+ [24.552]
International Organizations	-0.021** [0.005]	-0.018** [0.005]	-0.021** [0.005]	-0.018** [0.005]	-0.023** [0.006]	-0.021** [0.006]	-0.022** [0.006]	-0.025** [0.006]	-0.023** [0.006]
Power ratio	-0.365** [0.047]	-0.363** [0.047]	-0.378** [0.047]	-0.378** [0.046]	-0.346** [0.062]	-0.356** [0.063]	-0.359** [0.062]	-0.366** [0.062]	-0.369** [0.062]
Alliances	-0.511** [0.170]	-0.549** [0.170]	-0.541** [0.165]	-0.571** [0.171]	-0.642** [0.204]	-0.765** [0.201]	-0.733** [0.199]	-0.714** [0.195]	-0.679** [0.202]
Noncontiguity	-1.039** [0.183]	-1.015** [0.187]	-1.034** [0.187]	-1.048** [0.183]	-1.193** [0.259]	-1.146** [0.264]	-1.183** [0.259]	-1.155** [0.264]	-1.215** [0.257]
Log Distance	-0.416** [0.069]	-0.420** [0.070]	-0.421** [0.070]	-0.423** [0.071]	-0.442** [0.090]	-0.443** [0.090]	-0.439** [0.090]	-0.453** [0.088]	-0.441** [0.090]
Only Minor Powers	-0.874** [0.206]	-0.909** [0.201]	-0.903** [0.207]	-0.988** [0.206]	-0.706* [0.289]	-0.802** [0.279]	-0.769** [0.291]	-0.682* [0.286]	-0.807** [0.286]
Constant	0.235 [0.572]	0.743 [0.579]	0.799 [0.569]	0.8 [0.585]	0.535 [0.759]	1.037 [0.752]	1.089 [0.730]	0.919 [0.748]	0.959 [0.751]
Observations	35052	35052	35052	35052	27447	27447	27447	27447	27447
Model chi-square	250.82	244.88	201.8	232.92	192.76	189.44	163.32	174.76	165.91

Standard errors in brackets.

+ significant at 10%; * significant at 5%; ** significant at 1%

Table 2. Directed Dyad Analysis of Dispute Initiation, 1950-2000

	All Dyads					Politically Relevant Dyads				
	Polity Interval	Polity Dichotomous	ACLP	BNR	Doyle	Polity Interval	Polity Dichotomous	ACLP	BNR	Doyle
Both Democratic/Liberal	-0.449*	-1.409**	-2.049**	-2.109**	-2.179**	-1.014**	-1.707**	0.443	-1.807**	0.775**
	[0.205]	[0.281]	[0.267]	[0.254]	[0.252]	[0.237]	[0.328]	[0.293]	[0.308]	[0.248]
Initiator Polity	0.001					-0.005				
	[0.010]					[0.010]				
Target Polity	0.045**					0.034*				
	[0.011]					[0.013]				
Democratic/Liberal Initiator		0.702**	0.737**	0.826**	0.906**		0.391	0.968**	0.524*	1.130**
		[0.213]	[0.215]	[0.200]	[0.193]		[0.253]	[0.310]	[0.255]	[0.272]
Democratic/Liberal Target		1.310**	1.285**	1.340**	1.396**		1.083**	-1.648**	1.075**	-1.704**
		[0.198]	[0.208]	[0.191]	[0.194]		[0.249]	[0.346]	[0.275]	[0.299]
Initiator Relative Advantage	0.680**	0.726**	0.756**	0.741**	0.744**	0.465**	0.553**	0.520**	0.554**	0.418*
	[0.174]	[0.174]	[0.180]	[0.178]	[0.181]	[0.163]	[0.162]	[0.178]	[0.180]	[0.165]
Distance	-1.007**	-1.004**	-1.074**	-1.051**	-1.070**	-0.382**	-0.407**	-0.433**	-0.437**	-0.428**
	[0.094]	[0.093]	[0.096]	[0.096]	[0.095]	[0.104]	[0.099]	[0.101]	[0.103]	[0.102]
Peace Years	-0.445**	-0.448**	-0.447**	-0.448**	-0.452**	-0.479**	-0.477**	-0.476**	-0.474**	-0.478**
	[0.045]	[0.045]	[0.045]	[0.046]	[0.045]	[0.058]	[0.057]	[0.057]	[0.057]	[0.057]
Constant	2.633**	1.969**	2.557**	2.322**	2.445**	-0.021	-0.406	-0.141	-0.234	-0.35
	[0.655]	[0.693]	[0.713]	[0.703]	[0.701]	[0.682]	[0.713]	[0.724]	[0.710]	[0.743]
Observations	805714	805714	805714	805714	805714	74462	74462	74462	74462	74462
Log Likelihood	-4637.44	-4578.34	-4567.38	-4560.72	-4554.94	-2311.6	-2287.9	-2296.8	-2292.07	-2297.48
Model chi-square	685.72	719.36	672.75	705.75	700.36	378.16	375.7	393.54	406.73	420.27

Robust standard errors in brackets.

Splines not reported in table.

+ significant at 10%; * significant at 5%; ** significant at 1%