The Democratic Peace: Weighing the Evidence and Cautious Inference

Allan Dafoe
Yale University

John R. Oneal
University of Alabama

AND

Bruce Russett
Yale University

The greater peacefulness of jointly democratic pairs of states is an important finding in research on the causes of war. Here, we outline a set of criteria for evaluating critiques of such robust empirical associations and apply them to two recent articles. Mousseau (2013) claims that contract-intensive economies, as proxied by a measure of life insurance expenditure, account for the democratic peace. However, his research suffers from various problems of analysis and interpretation, including a miscoded dependent variable, a misleading specification for dyadic democracy, a suppression of heterogeneous associations, and a heavy dependence on imputation in which greater than 90% of the values of the central independent variable were (improperly) imputed. We estimate 144 specifications that build from Mousseau’s models and control for life insurance expenditures, finding substantial, robust support for the democratic peace. Gartzke & Weisiger (2013) claim that the importance of the democratic peace has declined as the proportion of democracies in the international system has increased; but their tests are misspecified and do not address the issue they raise. There are also serious errors in their data. When these problems are corrected, we find that the peacefulness of democratic pairs has actually increased as the proportion of democracies grew after 1816.

Democracies interact more peacefully with each other than do other pairs of states. This empirical association, known as the democratic peace, is highly robust and has been the foundation of a productive research program. Various political explanations have been offered. Skeptics have challenged the democratic peace on realist grounds and in the belief that some economic aspect of liberal societies accounts for the phenomenon. Here, we evaluate two such critiques: those of Mousseau (2013), who traces the greater peacefulness of democracies to economic roots, and Gartzke and Weisiger (2013), who offer a realist account.

Mousseau argues that peaceable societal norms arise from experience with a market economy and that these account for the democratic peace. In his analyses, per capita expenditures on life insurance seem to account for the peaceful behavior of democratic dyads. His results, however, depend on a major error in the coding of the dependent variable. In addition, there are problems in the construction of the independent variables, a misunderstanding of interaction effects, and an improper imputation procedure for his key variable—a serious issue since more than 90% of the values of Mousseau’s life insurance measure are imputed. We analyze 144 alternative specifications, correcting the two most serious errors and systematically modifying the other problematic aspects of Mousseau’s analyses. We find the democratic peace to be highly robust even to controlling for life insurance expenditures.

For G&W, the source of the democratic peace is essentially realist in nature: Cooperation depends upon the presence of a threat against which states define themselves and must align. Earlier, when democracies were scarce and autocracies plentiful, democracies were united by a common threat; but as the number of democracies increased, the threat from autocratic countries declined and differences among democracies are apt to have become more salient, undermining over time the separate peace they have enjoyed. G&W’s empirical reassessment of the democratic peace, however, is not consistent with their argument, and there are serious errors in their data.

The Democratic Peace Research Program

Credible causal inference requires strong theory or a strong design. By strong theory we mean that a scientific community has only a few viable contending theories, and each of these theories makes sharp competing predictions. By strong design we mean that causal factors of interest have been (as if) randomly assigned to a large population of independent units (Dunning 2012). The study of international relations, however, generally has neither strong theory nor many opportunities for strong design. Theories of international relations can rarely be subject to critical tests, and most data sets suffer from extensive confounding and dependence. Consequently, the study of international relations can rarely generate strong causal

1 Dafoe was responsible for the analysis of Mousseau (2013), Oneal for Gartzke & Weisiger (2013); all authors contributed to the writing of the paper. All replication files (data, Stata and R code, and log files) are on the Dataverse Network: http://hdl.handle.net/1902.1/19241.

We thank Charlotte Jande for suggestions on our figures.
inferences about important questions. In such a scientific domain, a robust empirical association involving important phenomena is of great value: It provides a key descriptive feature to be explained and an empirical foundation for theory. The democratic peace is such a foundation: It is robust and substantively important.

What specification (set of covariates, functional forms, estimators) should one use to analyze the democratic peace? The answer, increasingly appreciated by leading methodologists, is that we just do not know. Dependence in observations, myriad forms of endogeneity, measurement error, and unknown functional forms are some of the many threats to model-based causal inference. Bias in estimates from cross-national studies could be large. However, the analysis of cross-country data, done with care, can provide descriptive information about the size of conditional associations and how much they stand out from the unexplained variation (their statistical significance). Cautious causal inferences can then be drawn. Stronger conclusions can rarely be made in the study of international relations.

Our recommended approach to studying the democratic peace is thus to examine conditional associations under a broad set of reasonable conditioning strategies, with additional caution when interpreting models that include variables that may be causally affected by democracy. This approach suggests several criteria.

**Criteria for Evaluating Challenges to the Democratic Peace**

Challenges to the democratic peace generally have the following form: Including a variable $X$ leads the estimated coefficients on democracy to become statistically insignificant and the estimated coefficient on $X$ is associated with greater peacefulness. This implies that $C(X)$ is the cause of peace, where $C(X)$ is some concept well measured by $X$ and which is distinct from the political explanations previously offered. For example, Mousseau includes in his regression analyses an estimate of the logarithm of countries’ per capita life insurance expenditures on life insurance. Mousseau reports that including $\ln(\text{Life Insurance per Capita})$ renders the association between joint democracy and peace small and insignificant, that his preferred variable has a significant association with peace and that it proxies for processes distinct from the factors emphasized in other theories of the democratic peace. From this he concludes that it is commercial norms arising in contract-intensive economies, and not democracy, that make dyads more peaceful.

Before evaluating Mousseau’s work and that of Gartzke and Weisiger, we identify general criteria for evaluating challenges to robust descriptive associations such as the democratic peace.

**Criteria**

1. **Transparency and Replicability**
   
   Replication files should be publicly provided with as much detail as possible to automatically and precisely replicate all aspects of the analysis, beginning from primary data files.

2. **Principled Analysis**
   
   Analyses should follow a set of principles derived from precedents in the literature, simplicity, theory, and methodological best practice.
   
   a. Precedents in the literature.

Many analyses, especially of the democratic peace, examine data and address questions that have been studied extensively by others. That literature establishes important precedents and provides a useful guide to new studies. Analyses should be based wherever possible on standard models to minimize the chance that unrelated aspects of the analysis are driving the results. Deviations from standard practice should be scrutinized to see whether they account for novel findings.

b. **Simplicity and parsimony.**

Simple parsimonious methods are better understood, more transparent, and less likely to rest on subtle, inappropriate assumptions.

c. **Ex ante articulated and plausible theory.**

Plausible theory should guide analysis and be articulated prior to empirical testing.

d. **Best practice.**

The best methodological practices should be followed, though there may be trade-offs when following the principles set by precedence, simplicity, and theory.

3. **Robust Analysis**

The costs of running and reporting many analyses are today relatively small. Graphics can communicate the important statistics from a variety of specifications. Scholars should report extensive robustness tests for those dimensions of the analysis not constrained by strong principles.

4. **Cautious Causal Inference.**

Because of the many challenges to drawing causal inferences from observational data in fields like international relations. Causal inferences should be calibrated. Researchers should consider alternative explanations for their results and fairly represent the research of others.

Adhering to these criteria will improve inference in general. In particular, it will mitigate the problems of multiple-comparisons bias and publication bias. Multiple-comparisons bias is the tendency for results to be biased in the direction desired by the researcher through selective representation of results. A researcher has considerable freedom regarding the variables to include in his model, transformations of the variables, alternative dependent variables, and different data sets. Multiple-comparisons bias need not be conscious. As long as scholars are more likely to reevaluate their analyses when results are different than expected, they will be less likely to identify errors that disconfirm their theory.

Publication bias arises from the generally laudable institutions of peer review and editorial discretion in publishing. If reviewers and editors favor publication of “interesting” results, the distribution of results in published papers will then be biased toward those that are more “interesting.” If many scholars examine what happens to the democratic peace when capitalist variables are included in the model, those who choose a specification where the democratic peace is weakest will be overrepresented in the literature, even if most scholars found that capitalist variables account for only some (or none)
of the democratic peace, because it is a more interesting result.2

**Commercial Norms, Wealth, and the Democratic Peace**

In many publications, Michael Mousseau has argued that contract-intensive economies generate societal norms that promote peace. Since Mousseau (2009), he has also argued that the normative effect of widespread reliance on impartial contracts can account for the rise of democracy and the democratic peace. Mousseau claims to show that when controlling for the presence of a contract-intensive economy “there is no evidence of causation from democracy to peace, and contract-intensive economy is one of the most powerful nontrivial variables in international conflict” (2013, 194), that “the democratic peace is found to be spurious in analyses of [Militarized Interstate Disputes]” (Mousseau, Orsun, Ungerer and Mousseau 2013a; 93, citing Mousseau 2009), and that there is “precious little evidence that democracy is the cause of [the democratic peace]” (2013, 186).

In this section, we review the evidence for these claims. We find it to be highly fragile, with numerous flaws in the analyses. Using the general criteria articulated above, Mousseau’s research passes the first criterion since largely complete replication files are publicly available; but it fails on criteria 2, 3, and 4. His analyses deviate without adequate justification from precedent, are often ad hoc, and do not follow best practice (Criterion 2). The results are not robust to modest modifications in the statistical specification (Criterion 3). And the causal inferences are not cautious (Criterion 4).

Specifically, Mousseau’s earlier claims (2009) against the democratic peace, which he continues to cite without qualification, were based on a misinterpretation of interaction terms, as we have pointed out before (Dafoe 2011; Russett 2011). Dafoe and Russett (2013) discuss problems with Mousseau et al. (2013a). The work reviewed here (Mousseau 2013) involves a statistical analysis almost identical to that of Mousseau et al. (2013a), the primary difference being that Mousseau (2013) uses a different dependent variable. However, there is a major error in the construction of this dependent variable, the correction of which suffices to return significance to DemocracyLow in Mousseau’s primary specification. In addition, Mousseau (2013) inherits the flaws of Mousseau et al. (2013a), including an inappropriate specification for dyadic democracy, an unjustified suppression of heterogeneous conditional associations, and a central variable for which more than 90% of its values were (singly) imputed. To evaluate the effect of controlling for measures of life insurance on the democratic peace, we re-examine 144 variants of Mousseau’s models, systematically considering various modifications of problematic aspects of his work. Controlling for measures of life insurance expenditure, DemocracyLow (or its counterpart) is highly significant ($p < .01$) for 66% of the models and weakly significant ($p < .1$) for 89% of them.

We proceed by: (i) reviewing the theory linking liberal economic and political institutions to show the difficulty of disentangling their causal effects; (ii) reviewing the statistical analysis reported in Mousseau (2013); (iii) summarizing the most severe problems in Mousseau (2013); and (iv) reporting the estimated conditional association between DemocracyLow and peace for 144 analyses which systematically correct the problems.

**Theory**

How do liberal political institutions relate to liberal economic institutions? Scholarship on this thorny question is extensive and ongoing. Schumpeter (2003, 296) noted that “modern democracy arose along with capitalism, and in causal connection with it.” Olson (1993, 567) articulated what has become the modern view: “the conditions necessary for a lasting democracy are the same necessary for the security of property and contract rights that generates economic growth.” Acemoglu, Robinson, and coauthors expanded on Olson’s claim, showing through analysis of cross-country data (Robinson 2006; Acemoglu, Johnson, Robinson and Yared 2008), the use of creative instrumental variables (Acemoglu, Johnson and Robinson 2001), and detailed historical analysis (Acemoglu and Robinson 2012), that countries “embarked on divergent political-economic development paths at certain critical junctures” (2008, 808), and in particular that “extractive political and economic institutions support each other and tend to persist” (2012, 81). North, Wallis and Weingast (2009, 24) offer a similar conceptual framework for understanding “violence and social orders” that identifies a “double balance: open access and entry to organizations in the economy support open access in politics, and open access and entry in politics support open access in the economy” and that in open access orders “political control over violence combines with the rules governing the use of that violence to reduce and control access to violence” (115). Their approach is particularly resonant with Mousseau’s view since they highlight the role of “belief systems [in open access orders] that emphasize equality, sharing, and universal inclusion” (110) and that “all open access orders are, largely, impersonal” (113).

There are many other plausible connections between liberal economic institutions, economic growth, modernization, and liberal political institutions (for example, Teorell 2010).

From this perspective, Mousseau’s finding that expenditures on Life Insurance per Capita descriptively account for the democratic peace can be interpreted in many ways. Both the Polity index and measures of per capita life insurance expenditure might be serving as crude measures of the extent to which a country has “inclusive institutions” or is an “open access order.” Or life insurance expenditures could be picking up the achievement of particularly inclusive/open political-economic institutions, which the Polity index, for whatever reason, fails to do. Causal inference from cross-country data should be cautious.

**Mousseau’s (2013) Analyses**

Mousseau’s claims against the democratic peace are based on the results from six statistical models that he reports in his Table 1. The data include dyad-years from 1961 to 2001. The dependent variable for models 1,2,3,4, and 6 is (meant to be) the onset of fatal militarized interstate disputes (denoted here as FO for fatal onset), and for model 5 the onset of a militarized dispute (denoted O), for original disputants in the conflict, as coded by Maoz (2005). Logistic regression is used, with robust standard errors clustered on the dyad. All models have as covariates (denoted as Z) measures of: relative capability, whether one country is a major power, contiguity,
distance, number of states in the system, and an intercept. Different measures of DemocracyLow are used: the minimum Polity score in the dyad (denoted $D$), an indicator variable if both countries have a Polity score above 6 (denoted $BD$), an indicator for both countries having a Polity score of 10 (denoted $D10$), and the square of DemocracyLow (denoted $D^2$). Mousseau also includes Werner’s measure of the difference in the regime types of the two countries (called PoliticalDistance, denoted PD). Mousseau’s measure of life insurance expenditures is denoted as (denoted $LI$). All independent variables are lagged 1 year. To summarize, Mousseau’s analyses consist of the following models (where $\Lambda$ is the inverse logit function, and coefficients are suppressed):

- $M1: P(FO=1) = \Lambda(D + Z)$
- $M2: P(FO=1) = \Lambda(D + LI + Z)$
- $M3: P(FO=1) = \Lambda(D + LI + PD + Z)$
- $M4: P(FO=1) = \Lambda(BD + LI + PD + Z)$
- $M5: P(O=1) = \Lambda(D10 + LI + PD + Z)$
- $M6: P(FO=1) = \Lambda(D^2 + LI + PD + Z)$

Mousseau’s evidence involves comparing the estimated coefficient on DemocracyLow in M1, where it is negative and significant, and models M2–M6 where it is not (and in fact is often positive). The comparison of M1 to M3, M4, M5, or M6 is confounded by the introduction of PD, a control variable that is irrelevant to Mousseau’s theoretical argument. This distorts the meaning of the quantity being estimated. We therefore regard M2 as Mousseau’s primary specification, though we consider Mousseau’s alternative operationalizations of DemocracyLow used in M3–M5. We now discuss in detail the problems with Mousseau’s construction and selection of the dependent variable, his decision to control for PoliticalDistance, the decision to exclude an interaction between DemocracyLow and LI, and his measure of life insurance expenditure.

Operationalizing and Selecting the Dependent Variable

Studies of international conflict based on time-series cross-sectional data vary on at least two dimensions in specifying the dependent variable. One is whether to code each event in time unit (generally years) in which there is conflict (denoted here as ongoing MIDs) or to code only the onset of an event and set the remaining observations during which conflict continues to missing (Bennett and Stam 2004; 53–54, 229; Beck, Katz and Tucker 1998). Either approach can be justified. Mousseau (2013), however, incorrectly codes MID onsets and fatal MID onsets, setting dyad-years with an ongoing conflict to 0, rather than to missing. No scholar that we are aware of has advocated this approach. Under Mousseau’s setup, an increase in the duration of a conflict will make all variables correlated with that conflict appear to be more peaceful. Since democratic dyads have shorter conflicts (as well as fewer conflicts), Mousseau’s operationalization underestimates the pacific benefit of joint democracy. This one correction restores the evidence for the democratic peace in M2.

A second dimension over which analyses differ is the severity threshold used to code an event as a conflict. The most expansive category is a “militarized interstate dispute” (MID), defined as any event involving a threat, show, or use of force by one state against another. A more narrow category is a “fatal militarized interstate dispute,” which is a MID in which at least one combatant was killed. Mousseau uses the (misimplemented) onset of fatal MIDs in most of his models, though he examines the (misimplemented) onset of MIDs in one model (M5). It turns out that for M2, the estimated coefficient DemocracyLow would again be negative and significant when using the more expansive definition of conflict (MIDs). Following the criterion of robustness, in our analyses we systematically look at all four kinds of dependent variables: (corrected versions of) the onset of MIDs, ongoing MIDs, the onset of fatal MIDs, and ongoing fatal MIDs. To be consistent with Mousseau, we exclude states that join a dispute after the first day.

* * *

** Dyadic Specification of Democracy: What is the Counterfactual? **

Absent justification, statistical critiques of robust empirical associations should change as little as possible from standard analyses (Criterion 2a). Otherwise, it is hard to know what novel aspect of the specification accounts for the different results. If conditioning on a new covariate really accounts for the democratic peace—that “the democratic peace is found to be spurious” (Mousseau et al. 2013a; 95)—then such results should hold up under conventional models. Mousseau writes of his analyses that “most of the independent variables are conventional to the conflict studies literature” (2013, 191; Mousseau et al. 2013a). However, in four out of five models (M3–M6) Mousseau includes a problematic covariate—“PoliticalDistance”—which differs from standard practice (see Dafoe and Russett 2013, 113–16), is unrelated to Mousseau’s argument about economic norms, and distorts the meaning of the estimated association on DemocracyLow. For the purposes of our examination of Mousseau’s claims, it is sufficient to note that this modification is inappropriate since it is not relevant to Mousseau’s theoretical critique and it involves a fundamental reconceptualization of the quantity of interest. However, in the interests of evaluating this alternative specification, we explain the problems with this operationalization in more detail (see also Bennett 2006).

Onewe and Russett (1997) introduced an (almost) continuous operationalization of dyadic democracy involving two variables that has become common: One variable measures the democracy level of the less democratic country in a dyad (DemocracyLow) and the other variable the level of the more democratic country (DemocracyHigh). DemocracyLow tends to have a negative and significant association with conflict; this is typically regarded as the main empirical referent of the democratic peace. The estimated coefficient of DemocracyHigh is usually positive and significant; it captures the greater conflict propensity of dyads that have different political systems. Some scholars just include DemocracyLow. We denote models with DemocracyLow and DemocracyHigh as DLIH, and models with just DemocracyLow as DL. Mousseau et al. (2013a) use an alternative operationalization involving DemocracyLow and a measure of RegimeDifference = DemocracyHigh − DemocracyLow, denoted here as DLRD. Mousseau (2013) now uses Werner’s measure of PoliticalDistance (denoted as DLPD) instead of RegimeDifference, which is meant to proxy the same conditional associations and plays the same problematic role. Here, we focus on the effect of controlling for Regime-

---

3 We do not examine $D^2$ because the indefinable square functional form is not appropriate, and adding more terms would complicate our summary statistics.

4 Which implicitly assumes that the coefficient on DemocracyHigh is 0.
Difference since the conceptual issues and problems are the same as for PoliticalDistance, but more transparent.

The DLRD specification is econometrically equivalent to the DLDH specification: It involves estimating the same statistical model and simply apportions the estimated associations across coefficients in a different manner (Dafoe and Russett 2013). Nothing new is revealed about any underlying empirical associations, and any conclusions drawn from one model follow from the other. At issue, then, is what is the quantity of interest that we want the coefficient on DemocracyLow to represent.

Most quantities of interest are causal counterfactuals. A heuristic for clarifying the counterfactual of interest is to ask: What is the experiment that researchers would run if they could (Dorn 1953, 680; Sekhon 2009, 496)? The counterfactual implicit to most democratic peace research is that of increasing (or decreasing) the level of democracy of one of the countries in a dyad. Asking such a question directs us to Oneal and Russett’s DLDH specification, for which both counterfactuals have an estimator.

The DLRD specification used in Mousseau et al. (2013a) estimates a coefficient on DemocracyLow that represents the effect of increasing the level of democracy of both countries in a dyad simultaneously and by exactly the same amount. Controlling for RegimeDifference means to hold it constant. In the DLRD specification, the coefficient on DemocracyLow represents the change in conflict propensity associated with an increase in DemocracyLow by one unit, while also increasing DemocracyHigh by one unit, so as to hold RegimeDifference constant.

It is hard to imagine a policy manipulation or historical process that would generate this counterfactual. Further, even if we hypothetically entertain this counterfactual, due to the structure of dyadic data all other dyads involving these countries would experience the conventional counterfactual reported by the DLDH model in which the democracy level of only one country changes.

Thus, the DLRD specification implies a quantity of interest that is unrealistic, a minor causal feature even of hypothetical scenarios, and radically deviates from the conventional understanding of the democratic peace. The problem remains, but is obfuscated, by Werner’s (2000) PoliticalDistance, which is based on a complex function of the component variables in the Polity index. As before, the quantity of interest implied by the coefficient on DemocracyLow for the DLDP specification is the effect of increasing DemocracyLow by one unit, while simultaneously changing the more democratic country by the precise extent needed to hold PoliticalDistance constant.

More work is needed to understand just how political regime interacts with the regimes of other countries. At present, however, the DLDH specification maps more closely to the counterfactuals of interest to most scholars than do the DLRD or DLDP specifications.

**Estimating Conditional Associations with Interaction Terms**

In his earlier work, Mousseau reported a significant interaction between DemocracyLow and GDP/capita. For example, Mousseau, Hegre and Oneal (2003) found that the (conditional) association between DemocracyLow and peace is stronger among dyads with higher levels of GDP/capita. In their analyses, the interaction term was highly significant, but “… most democratic dyads in [their] sample, 91%, are in the zone of peace at usual thresholds of statistical significance…” (2003, 300).

Mousseau (2009) reported a set of regressions involving a binary measure of Life Insurance per Capita and an interaction between Life Insurance per Capita and DemocracyLow. The estimated coefficient on the constituent term for DemocracyLow is insignificant for all models involving Life Insurance per Capita and the interaction, and the interaction term is highly significant. Mousseau interpreted this as “compelling evidence that [democracy does not promote] peace among nations” and that the “zone of peace … among democratic nations … appears to be caused by economic rather than governing institutions” (Mousseau 2009: 54). He continues to cite Mousseau (2009) as evidence that “the democratic peace correlation is not significant once [controlling for] contract-intensive economy [in analyses of] fatal militarized interstate conflicts” (2013, 186) and as the study “when [he] first directly showed the democratic peace [to be] spurious” (2013, 190).

However, this inference is based on a misreading of the interaction term (Dafoe 2011; Russel 2011; Dafoe and Russett 2013), as Mousseau only interprets the coefficient on DemocracyLow without taking into account the interaction. When interpreted correctly, the results in Mousseau (2009) look similar to those in Mousseau et al. (2003). Consider all models in Mousseau’s (2009) Tables 1 and 2, except for Models 5 and 6, which drop close to 50% of their sample due to missing data. For those dyads that have low levels of Life Insurance per Capita for both countries, DemocracyLow is insignificant though still correctly signed (associated with peace) for all models. For those dyads in which at least one country has above-median levels of Life Insurance per Capita, the estimated association on DemocracyLow is highly significant ($p < .001$) in the predicted direction. Seventy-four percent of all democratic dyads (DemocracyLow > 6) and 95% of the most democratic dyads (DemocracyLow = 10) are in this subset of the data where the democratic peace is found to be strong. Without the interaction term, the coefficient on DemocracyLow is estimated to be highly significant ($p < .001$) in all models. In short, the entire evidential basis for the strong claims against the democratic peace in Mousseau (2009) rests on a basic misreading of interaction terms.

Mousseau’s most recent analyses (Mousseau 2013; Mousseau, Orsun and Lee Ungerer 2013b; Mousseau et al. 2013a) no longer include interaction terms between DemocracyLow and his measures of life insurance expenditure. Dafoe and Russett (2013), when examining the robustness of (Mousseau et al. 2013a), add this interaction term and again find a strongly significant interaction similar to that found in Mousseau (2009) and Mousseau et al. (2003). DemocracyLow is significantly associated with peace for the most developed dyads (dyads with high GNP/capita or high Life Insurance per Capita) and weakly (generally insignificant but correctly signed) associated for less developed dyads.

Mousseau now writes that including an interaction term lacks “theoretical justification” (contra Mousseau (2009) which depended on it), that it “appears to be no more than an ad hoc attempt to save a hypothesis” which is possible since “it is not always difficult to mine data and dredge up third variables that can help obtain the results one wants,” and “as such cannot be convincing and deserves no more of our attention.” (2013, 191)

How can we adjudicate whether it is appropriate to include an interaction term? In this case, the issue is rela-
tively straightforward. Suppose we believe that our model yields unbiased and consistent estimates because the zero conditional mean assumption, or a stronger assumption such as conditional independence, is thought to hold (see Wooldridge 2008, 54). Then, the inclusion of an interaction term of the variables already in the model cannot induce bias or inconsistency, since the zero conditional mean assumption will hold under the addition of any interaction or transformation of the independent variables. Put differently, the original model assumes that the interaction term is zero. If that model was correct or appropriate, then the estimated association on the interaction term should be close to zero. That the interaction is significantly different from zero provides strong evidence against the assumption of no interaction.

Even though including an interaction term cannot undermine the zero conditional mean assumption, the risk of multiple-comparisons bias remains. To adjudicate specification in cases such as this, we recommend (1) selecting a specification that captures the main contours of the data and (2) is consistent with some plausible theories. If there is reasonable doubt about the correct specification, we recommend (3) examining robustness under different specifications. Regarding these, (1) clearly weighs in favor of including the interaction. Regarding (2), it is not hard to articulate theories that are consistent with an interaction: Economic norms may express themselves more forcefully in liberal polities; moral concerns weigh more heavily when people are rich; the stability and bargaining credibility made possible by democracy (Lipson 2003) is more robust when governments are dependent on capital. Regarding (3), our analyses reported below consider models with and without this interaction term.

**Life Insurance per Capita and Its Imputation**

Mousseau’s analyses hinge on a single variable—*Life Insurance per Capita*—to proxy for commercial norms arising from contract-intensive economies. Ideally, a theory would be subject to a variety of empirical tests, and not be dependent on a single empirical domain (cross-country data) and especially a single operationalization of its central concept. This variable is especially problematic, however, because originally it was coded only at 5-year intervals since 1960 for 66 countries (about 330 country-years). Consequently, Mousseau has imputed a large proportion of his data. In fact, 98% of the dyad-year observations in Mousseau’s final data set are (singly) imputed, 90% of which involve extrapolation; in addition, Mousseau deletes 7% of the original sample (the years from 1950 to 1960) due to this missing data. This is highly problematic.

In Mousseau (2009), the measure of *Life Insurance per Capita* was a dichotomous variable set to 0 for values below the median of the continuous measure, and 1 above. Values were first linearly interpolated for 8% of the sample. The extensive remaining missing values were then imputed as 0, under the reasonable assumption that they are less than the observed median. The major innovation of Mousseau et al. (2013a) and Mousseau (2013) is imputing a continuous version of \( \ln(\text{Life Insurance per Capita}) \). Mousseau’s imputation model is not clearly documented, but from his replication files, we can discern that his imputation model consists of the following variables: energy consumption per capita and its natural logarithm, absolute private investment, absolute private consumption, private investment as share of trade, private consumption as a share of trade, indicator of oil production, indicator for communist countries, indicator for formerly communist countries, total population, indicator for countries with small populations, indicator for the Middle East, and a seemingly ad hoc set of 18 temporal dummies and piecewise linear variables. Mousseau imputes values of the natural logarithm of a variable coded by Beck and Webb (2003, 53–54) called *Life Insurance Density*, which Beck and Webb define as “premiums per capita” and is meant to show “how much each inhabitant of a country spends on insurance on average.”

Neither Mousseau (2013) nor Mousseau et al. (2013a) give sufficient discussion or justification of the details of the imputation procedure. Information is not given about what proportion of the data were imputed or about what imputation model is used. Imputation procedures rely on a set of assumptions, most critically that the data are “Missing at Random” (MAR), which requires that, conditional on the statistical model, missingness is independent of everything else. What evidence do we have that MAR holds in this case? If MAR does not hold, what kinds of biases might be generated for the greater than 90% of the data that are imputed? Why did Mousseau exclude some variables in the imputation model that were in the analysis model, violating another best practice (Meng 1994; Rubin 1996, 478; Honaker and King 2010, 564)?

Mousseau uses a single imputation command in Stata—*impute*—which is obsolete in current versions of Stata because of the recognized limitations of single imputation. Single imputation estimates only one value for each missing observation, which fails to represent the uncertainty in these estimates. Multiple imputation, by contrast, estimates multiple values for each missing observation based on the estimated uncertainty in the imputation model; this allows an analyst to preserve the uncertainty of the imputation procedure into the final analysis. For this reason, multiple imputation is superior to single imputation (Rubin 1996; Honaker and King 2010).

With greater than 90% of the values of *Life Insurance per Capita* imputed, minor details of the imputation procedure could exert a substantial effect on the results. This imputation procedure should be scrutinized and justified before future scholars rely on this variable in their analyses. For a subset of our analyses, we implement best practice in imputation: multiple imputation, including all analysis variables in the imputation model. We used *Amelia II* for multiple imputation. We performed \( m = 20 \) imputations, adding country-specific time trends, as well as a number of other variables that could plausibly predict life insurance expenditure per capita: counts of the occurrence of MIDs, fatal MIDs, and wars, as well as 1-year lags and leads; variables for Beck & Webb’s other measures of life insurance expenditure; lagged, contemporaneous, and lead values of \( \ln(\text{GDP/capita}) \), as well as its square and cube; squares and cubes of the Polity score; the components of the Correlates of War military capabilities index; and all other analysis variables. Full details are in the replication files. We opted to build from Mousseau’s imputation model for parsimony and comparability; this is not the place to devise and justify a wholly new imputation model. However, any future use of this variable should

---

5 Imputation is complicated by the constraint to impute data in a country-year format that will ultimately be analyzed in a dyad-year framework. As such, analysis variables such as the occurrence of MIDs and fatal MIDs were implemented as yearly counts for each country.
begin with a more carefully developed imputation model and should thoroughly examine the MAR assumption.\footnote{We initially devised our own statistical model for multiply imputing \texttt{ln(Life Insurance per Capita)} and \texttt{ln(Life Insurance Penetration)}. Analysis with these variables again showed a robust democratic peace. However, these life insurance variables did not have a robust association with peace. The sensitivity of these associations reinforces the importance to look carefully at the effects of different imputation models.}

To summarize, the central variable introduced by Mousseau is imputed to such a large degree that any inferences based on it will be heavily dependent on the imputation model and the validity of the MAR assumption. Basic maxims of best practice in imputation—to do multiple imputation, to include all analysis variables, to describe one’s imputation model—were neglected. In our analyses, the use of the multiply imputed versions of \texttt{ln(Life Insurance per Capita)} generally strengthened the democratic peace (approximately doubling the estimated effect, see Figure 1a–c); this multiply imputed version of \texttt{ln(Life Insurance per Capita)} was also more weakly associated with peace than the singly imputed version, though still generally significant (see Figure 2).

The authors who initially coded \textit{Life Insurance Density} (Beck and Webb 2003), on which Mousseau builds his \textit{Life Insurance per Capita}, also provided three other measures meant to capture different aspects of life insurance consumption. In particular, they note that the variable Mousseau used—\textit{Life Insurance Density}—“compares life insurance consumption across countries without adjusting for income.” This is potentially problematic: \texttt{ln(Life Insurance Density)} correlates at a high level ($r=0.86$ in Beck and Webb’s data set) with \texttt{ln(GDP/per capita)}. GDP/capita is a measure theorized to be associated with a variety of liberal processes, such as greater state capability, wealth, education. To the extent that one wants to draw relatively precise causal inferences from cross-national associations then it is prudent to address this confounding. In the interests of robustness, we also multiply impute and analyze Beck and Webb’s (2003, 53) first measure of life insurance consumption—\textit{Life Insurance Penetration}. This is defined as “the ratio of premium volume to GDP [], and [measures] insurance activity relative to the size of the economy.”

\section*{Analysis}

To examine the democratic peace when controlling for \texttt{ln(Life Insurance per Capita)}, we ran 144 different specifications; for comparison, we also estimated 24 specifications without \texttt{ln(Life Insurance per Capita)}. We consider three operationalizations of the dyadic lower democracy score: analyses using the continuous measure of \texttt{DemocracyLow} in Figure 1(a) and \texttt{BothDemocracy} (an indicator for \texttt{DemocracyLow=0}) in Figure 1(b), and \texttt{BothDemocracy10} (an indicator for \texttt{DemocracyLow=10}) in Figure 1(c). Mousseau’s primary specification, Model 2, from Table 1, involving the erroneous dependent variable in which ongoing conflicts are coded as peace, is plotted (+) and labeled “(Mousseau 2013) Base Model”. For the remaining specifications, the dependent variables are correctly implemented so that ongoing conflicts are set to either missing or 1.

We label the other models as follows. When the dependent variable involves fatal MIDs, as opposed to all MIDs, we include an \texttt{F}. When the dependent variable is an ongoing conflict (as opposed to just onset): \texttt{O}. When the specification has an interaction between \texttt{DemocracyLow} and \texttt{ln(Life Insurance per Capita)}: \texttt{I}.\footnote{For these models, the estimated effect and significance on \texttt{DemocracyLow} is based on setting \texttt{ln(Life Insurance per Capita)} to the 75th percentile. Models \texttt{I} are only considered for models \texttt{L}, that is, models including \texttt{ln(Life Insurance per Capita)}.} When a control for \texttt{DemocracyHigh} is included: \texttt{D}.\footnote{We do not include \texttt{RegionDifference} or \texttt{PoliticalDistance} for the reasons described above.} When the measure of \texttt{ln(Life Insurance per Capita)} is based on our multiple imputation procedure, rather than Mousseau’s imputation: \texttt{M}. When we substitute a multiply imputed version of the natural logarithm of Beck and Webb’s primary measure of life insurance expenditure (\textit{Life Insurance Penetration}) that is designed to be less correlated with Beck and Webb’s primary measure of life insurance expenditure (\textit{Life Insurance Penetration}) that is designed to be less correlated with \texttt{GDP/capita} \texttt{L}.

For models \texttt{M} and \texttt{L}, we redo the entire analysis, merging data files from original sources often using EUGene (Bennett and Stam 2012); this involves additional deviations from Mousseau’s analysis, such as which measure of Polity is used (Mousseau’s Polity measure does not correspond to Polity IV). Except for the changes described above, all other aspects of the models are based on Mousseau’s model specification and data. We also include 24 specifications, labeled with an \texttt{n} and represented using a square (\texttt{c}), that do not control for \texttt{ln(Life Insurance per Capita)}, as a comparison to evaluate how including \texttt{ln(Life Insurance per Capita)} changes the democratic peace.

The results from these analyses are presented in figure. Figure 1(a) plots the estimated effects\footnote{Specifically the reduction in predicted probability of conflict from one SD (6 units) change in \texttt{DemocracyLow} or one unit change in \texttt{BothDemocracy} or \texttt{BothDemocracy10}, from baseline probability of conflict of 0.5.} and $p$-values for all permutations of the above specifications using the continuous measure of \texttt{DemocracyLow}. Figure 1(b) does the same for the dichotomous measure of \texttt{BothDemocracy}. Figure 1(c) the same for \texttt{BothDemocracy10}. Note that Figure 1(c) is missing some models: These are models in which \texttt{BothDemocracy10} perfectly associated with peace and thus a coefficient could not be estimated.

Figures 1(a–c) reveal that the democratic peace is highly robust to controlling for measures of life insurance expenditure, especially after using the multiply imputed versions of \texttt{Life Insurance per Capita}. Democracy has a peaceful association in all (140) models that could be estimated; in 24 models, \texttt{BothDemocracy10} perfectly predicts peace and thus the coefficient cannot be estimated. Of the 120 estimable specifications controlling for \texttt{Life Insurance per Capita}, the coefficient for Democracy is significant at the $p < .1$ level in 107 of them (89%) and is highly significant ($p < .01$) in 79 (65%) of them. When only considering estimable models with multiply imputed versions of \texttt{Life Insurance per Capita}, Democracy has a significant peaceful association in 77 out of 80 (96%) of them, and these effects are especially pronounced for \texttt{BothDemocracy} (Fig. 1b) and \texttt{BothDemocracy10} (Fig. 1c).

In short, considering this large space of specifications, one must select the precise specification that Mousseau used, and operationalize the dependent variable incorrectly, to get results like his. All other specifications estimate a conditional association between \texttt{DemocracyLow} and peace, and the $p$-values on these associations tend to be small. Readers should examine these figures for the specifications that they find most informative.

Controlling for \texttt{ln(Life Insurance per Capita)} does weaken the association between democracy and peace. To see this, compare the specifications that control for \texttt{ln(Life Insurance per Capita)} (circles and triangles) with those that do not (squares). Specifications without such a control have an estimated effect of \texttt{DemocracyLow},
FIG 1. The (conditional) association between DemocracyLow and peace is strong and highly significant for a large set of specifications. The \( p \)-values on the \( x \)-axis are plotted on a square-root scale. (a) includes specifications with continuous DemocracyLow, (b) an indicator for DemocracyLow > 6, and (c) an indicator for DemocracyLow = 10. Y-axis: reduction in predicted probability of conflict from one SD (6 units) change in DemocracyLow, or one-unit change in BothDemocracy or BothDemocracy10, from baseline probability of conflict of 0.5. “Base Model” is Model 2 from Table 1 of (Mousseau 2013), with misimplemented dependent variable (DV) of Fatal MID onset. All remaining models use corrected DVs and vary aspects of specification labeled by letters. F: DV as fatal MIDs, not all MIDs. O: DV ongoing, not onset. I: DemocracyLow x ln(Life Insurance per Capita) interaction included. D: DemocracyHigh included. M: ln(Life Insurance per Capita) multiply imputed. L: alternative measure of life insurance expenditure, ln(Life Insurance Penetration), used.
Fig 2. The (conditional) association between imputed measures of life insurance expenditure and peace seems robust. Model notation the same as for previous figures, except ‘denotes that DemocracyLow > 6 and ‘denotes that DemocracyLow = 10, rather than continuous DemocracyLow. Effect estimates based on one SD change in independent variable, from a baseline probability of conflict of 0.5. Due to high degree of imputation (>90%), results may be sensitive to other imputation models.
Both Democracy, and Both Democracy 10 about twice as large. This demonstrates that the association between democracy and peace partly overlaps with the association between ln(Life Insurance per Capita) and peace. Nevertheless, democracy is still a significant predictor of peace. Gartzke and Weisiger’s Reevaluation of the Democratic Peace

Many explanations have been offered for the peacefulness of democratic pairs of states. We have suggested that democracies fight less because they have less to fight about than other dyads (Onew and Russett 1997). They agree on important principles of political economy, including the preference for pursuing one’s interests nonviolently. For G&W, the source of the democratic peace has realist roots: Cooperation depends upon the presence of a threat against which states define themselves and must align. When democracies were scarce and autocracies plentiful, democracies were united by a common threat; but as the number of democracies increased, the threat from autocratic countries declined and differences among democracies became more salient. G&W emphasize that the prospects for peace depend upon the interaction of dyadic and systemic political characteristics. H 1 is clear: The probability of conflict for democratic dyads is conditional on systemic democracy levels, in particular the proportion of democracies among the countries of the world (p. 177). “It is not the impact of either dyadic or systemic democracy alone that is anticipated by the theory, but the interaction of each with the other” (p. 177). G&W’s empirical reassessment of the democratic peace, however, is not consistent with their theory, violating Criterion 2c. The model they specify does not make the peacefulness of democratic pairs conditional on the political composition of the international system.

In their primary test (Table 2), they interact Dyadic Difference [Democracy (high) – Democracy (low)] with the Proportion of Democracies, which equals the number of democracies divided by the total number of states in the international system in a given year. Two thoroughly democratic states have the same Polity score of +10, so Dyadic Difference equals 0, as does the interactive term. The strength of the democratic peace is, therefore, by construction, not conditional on the proportion of democracies in the system. The political composition of the system only has an independent effect on the odds ratio of conflict for the most democratic pairs. Worse, in G&W’s model, the systemic effect is the same for any two countries with identical Polity scores, two autocracies as well as two democracies; but their theory predicts that the effects of the Proportion of Democracies on jointly democratic dyads and jointly autocratic dyads should be quite different. 10

Indeed, G&W give little consideration to the effect of the proportion of democracies on autocratic pairs of states, and surprisingly, they do not present empirical evidence on this point, though they emphasize the generality of their theory. As indicated by their illustrative numerical model on p. 176, a growing proportion of democracies in the system should have opposite effects on jointly autocratic and jointly democratic pairs. An autocratic peace should wax as the democratic peace wanes. Autocracies should refrain from fighting one another as democracies become more numerous and the

10 We consider only the proportion of democracies as a measure of Systemic Difference because it is the most appropriate measure for testing their theory and the one emphasized by G&W. It might be fruitful to explore a measure weighted by the power of states, but the standard deviation of Polity scores, a measure they also consider, is inappropriate because it would be the same whether 90% of the countries were democracies or 90% were autocracies.
common threat they present to their autocratic rivals grows.

G&W also claim that the probability of conflict for mixed dyads, one autocracy and one democracy, will decline as the number of democracies increases, though the theoretical reasons for this are unclear. Their expectation seems derived from their Figure 1(a) (p. 177) where, according to their numerical model, the risk of conflict for a mixed pair declines sharply at first and then levels off as the number of democracies grows. It is hard, however, to reconcile this decline in violence with their Table 1, where there is virtually no difference—0.037 vs. 0.038—in the risk of conflict for mixed pairs over the historical range of values for the Proportion of Democracies and the full range of possible values (0–100%). It is important to note that even at contemporary levels of systemic democracy, their numerical model indicates that mixed pairs will be an order of magnitude more conflictual than democratic dyads.

Properly Specifying the Conditional Effects of Political Regimes

The conditionality of the democratic peace on the political composition of the international system can be simply assessed by creating an indicator variable that identifies jointly democratic dyads and interacting it with the proportion of democracies in the system. As noted above, the behavior of autocratic pairs should also be affected if G&W are correct in believing that cooperation between states with similar political systems is dependent on the level of threat they face from dissimilar polities. When autocracies are scarce and democracies plentiful, autocracies should be united by the common threat they face and behave more peacefully toward one another. Consequently, in the tests below we also identify jointly autocratic pairs and interact that variable with the proportion of democracies system-wide. Historically, the "autocratic peace" should have grown with the increasing number of democratic states, if G&W are correct. Finally, we add an indicator for mixed dyads, those pairs that include one democracy and one autocracy, and again create an interactive term. This setup allows us to estimate the effects of the political composition of the system on the three types of dyads of greatest interest. In particular, we can assess the conditionality of the democratic peace and can determine whether it is meaningful to speak of an autocratic peace.

Like G&W, we estimate the consequences of the increasing number of democracies on the probability of interstate conflict using a logistic regression in which there are controls for important realist influences, but we make two changes. First, we replace the indicator (Major Power) identifying dyads that include at least one state that was a major power according to the Correlates of War (COW) project with a continuous measure of the size of the more powerful state in each dyad. To do this, we use COW’s Composite Index of National Capability (CINC). We justified this improvement in the liberal-realist model elsewhere (Hegre, Oneal and Russett 2010), noting that Bremer (1992:337) worried from the beginning that the major-power indicator was unduly subjective and biased by a posteriori knowledge about which states fought frequently, a behavior characteristic of major powers. A binary variable is also a crude indicator of a state’s ability to project military force. The CINC score is a continuous measure of states’ power that takes into account demographic, industrial, and military capabilities. Second, we include controls for temporal dependence (Beck et al. 1998) as G&W did, but add a measure of the number of the states in the international system to account for the real opportunities for interaction as the state system expands (Raknerud and Hegre 1997).

Our dependent variable is the onset of a militarized interstate dispute (MID). Bennett and Stam (2004) recommend considering only the originators of a dispute, those states involved on the first day of fighting, because the outbreak of fighting changes the situation for countries that might join (or be forced to join) subsequently. In particular, information regarding the correlation of force is affected. Excluding joiners may be appropriate if one is focusing on realist theories; but we are concerned with testing the democratic peace. The onset of a dispute provides new information about the balance of power but not states’ political characteristics. Consequently, we also report tests in which conflicts involving joiners are included. In addition, we consider two sets of MIDs: (1) all threats, displays, and uses of force including those that become wars and (2) only fatal disputes, where at least one combatant was killed. We have a long-standing preference for the latter (Oneal, Russett, and Berbaum 2003). Focusing on these more violent conflicts reduces the bias in the reporting of less severe military incidents. Examining fatal disputes also helps insure that our analyses are relevant to the violent international interactions of greatest concern. Bluffing is more common at low levels of conflict (Bueno de Mesquita 1981).

There are serious errors in G&W’s data, so we created our own to conduct tests outlined above. Most importantly, many of the Democracy scores [Democ–Autoc] of −9 in their data are in fact missing values in the Polity IV data (Marshall, Jaggers and Gurr 2011). For example, Polity does not report data for the Bahamas (1973–2000), Malta (1964–2000), or the Maldives (1965–2000) because these countries were below the 500,000 population threshold. Other countries have missing data because Polity’s definition of statehood differs from COW’s. Thus, G&W include tens of thousands of dyadic observations that are miscoded on the variable of primary interest. Of course, this also affects their measurement of the proportion of democracies in the international system. Second, G&W do not accurately identify contiguous states. They say their measure of geographic proximity identifies countries that share a border either directly—state to state—or through colonial possessions (p. 178); but the COW variable they use (Colcont) captures only the latter condition. Thus, in their data, Canada and the United States are contiguous only until 1959 when Alaska becomes a state. In the lib-

---

11 The liberal-realist model is consistent for all three dependent variables (all MIDs, fatal MIDs, and wars), as reported previously (Oneal and Russett 2005).
Table 1. Models of the Onset of Militarized Interstate Disputes and Fatal Disputes, 1816–2001, Coefficients and Standard Errors of Liberal Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>All MIDs Originators</th>
<th>Fatal MIDs Originators</th>
<th>All MIDs All Pairs</th>
<th>Fatal MIDs All Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joint Democracy</td>
<td>β</td>
<td>0.63</td>
<td>0.82</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>SEβ</td>
<td>0.67</td>
<td>1.24</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Proportion of Democracies</td>
<td>1.98***</td>
<td>2.07</td>
<td>1.80***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.71</td>
<td>1.57</td>
<td>0.65</td>
</tr>
<tr>
<td>JtDem * PropDem</td>
<td>−4.04*</td>
<td>−7.13*</td>
<td>−4.85***</td>
<td>−3.83***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.25</td>
<td>3.79</td>
<td>1.93</td>
</tr>
<tr>
<td>JtAut * PropDem</td>
<td>−0.69**</td>
<td>1.01</td>
<td>−0.38</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.33</td>
<td>0.64</td>
<td>0.27</td>
</tr>
<tr>
<td>Mixed Pair</td>
<td>0.63*</td>
<td>−0.38</td>
<td>0.54*</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.39</td>
<td>1.17</td>
<td>0.32</td>
</tr>
<tr>
<td>Mixed * PropDem</td>
<td>−0.33</td>
<td>3.51</td>
<td>0.77</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.31</td>
<td>3.95</td>
<td>1.09</td>
</tr>
<tr>
<td>Wald χ² (df)</td>
<td>1934.1 (17)</td>
<td>446.3 (17)</td>
<td>2577.1 (17)</td>
<td>1065.3 (17)</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>.14</td>
<td>.22</td>
<td>.14</td>
</tr>
<tr>
<td>N</td>
<td>537,603</td>
<td>537,520</td>
<td>538,336</td>
<td>538,161</td>
</tr>
</tbody>
</table>

* p < .10; ** p < .05; *** p < .01 (two-tailed test). Estimated coefficients of the realist variables and the statistical controls (the years-of-peace, its cubic splines, and the number of states in the international system) are not reported to save space. See the posted log-files for complete results.

Table 2. Probabilities of the Onset of a Militarized Interstate Dispute, in %

<table>
<thead>
<tr>
<th>Proportion of Democracies in System</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.032</td>
</tr>
</tbody>
</table>

All MIDs, originators only
- Democratic pair: 13.8% | 6.1 | 4.9 |
- Autocratic pair: 4.6 | 17.2 | 31.8 |
- Mixed pair: 13.8 | 21.5 | 27.7 |

Fatal MIDs, originators only
- Democratic pair: 2.2% | 0.2 | 0.1 |
- Autocratic pair: 2.1 | 0.3 | 0.3 |
- Mixed pair: 0.8 | 4.3 | 14.6 |

All MIDs, originators and joiners
- Democratic pair: 14.5% | 4.7 | 3.0 |
- Autocratic pair: 6.4 | 20.4 | 35.0 |
- Mixed pair: 13.8 | 29.7 | 41.9 |

Fatal MIDs, originators and joiners
- Democratic pair: 1.4% | 0.2 | 0.1 |
- Autocratic pair: 1.2 | 2.1 | 3.4 |
- Mixed pair: 1.2 | 7.7 | 20.2 |

Results

We present the estimated coefficients and standard errors for the indicators of jointly democratic, jointly autocratic, and mixed dyads; the proportion of democracies in the system; and their interactive terms for our four analyses of interstate conflict, 1816–2001, in Table 1. We also provide summary statistics for the models. We omit the realist variables and statistical controls from the table to save space. Those results are consistent in almost all cases with our previous work and the work of others, including G&W (Table 2, p. 178). Our results leave no doubt about the continuing importance of the democratic peace. Indeed, contrary to G&W’s claim, the separate peace among democracies has strengthened as the proportion of democracies has increased. That is the only interactive term that is statistically significant in all four tests. Contrary to G&W’s expectations, there are no consistent effects of the growth in the number of democratic polities on either jointly autocratic or mixed dyads. The interactive term Mixed * Proportion of Democracies is never significant, and the results for the corresponding variable for jointly autocratic pairs are inconsistent and in only one case statistically significant at even the .10 level. G&W’s results are not robust (Criterion 3). Finally, the estimated coefficient for Proportion of Democracies is positive and significant in all four models. Interstate violence for the reference group (all dyads that include one state that is neither a democracy nor an autocracy) has increased over time. Thus, we find no evidence that a growing number of democracies pacify the system as a whole (Russett and Oneal 2001; Crescenzi, Kadera, Mitchell and Thyne 2011).

Table 2 presents estimates of the probability of the onset of interstate conflict (either any MID or a fatal one, for originators only or originators and joiners), varying only the political character of the dyad and the proportion of democracies in the international system. We use the historical range for the latter variable—0.032 and 0.427—and extrapolate into the future by adding 50% to the higher value (0.641). In generating our estimates of the risk of a dispute, we focus on “dangerous dyads,” to use Bremer’s (1992) memorable phrase, setting the realist variables at levels conducive to violence: The two states are large, share a land border, have proximate capitals, are not allies, and recently fought. We used King, Tomz

14 G&W (p. 177) say the onsets of disputes are appropriate for testing their theory; but in creating their dependent variable, they specified that no cases be excluded so subsequent years of a dispute are coded 1, not missing. We follow Beck et al. (1998) and set subsequent years to missing.

15 Specifically, the states are at the 90th percentile in the system-wide CINC scores and at the 10th percentile in capital-to-capital distance, and there were 5 years since the last MID.
and Wittenberg’s (2000) statistical program to clarify our results. Many of the estimated probabilities reported in the table have large standard errors because the interactive terms are themselves insignificant, but the patterns in the results are clear.

The first thing to note is that in the early 1800s, when democracies constituted only a small proportion of the states in the system, jointly democratic pairs were not particularly peaceful (Farber and Gowa 1997; Russett and Oneal 2001). The United States had several low-level conflicts with other democracies in the nineteenth century as did France, but fatal disputes between democratic countries were rarer. By 1864, two democracies were more peaceful than two autocratic states if the less favorable of the analyses using fatal MIDs is considered, that with disputes involving originators only; the transition is early in the nineteenth century if both originators’ and joiners’ fatal disputes are analyzed. The greater peacefulness of democratic pairs is evident only later if all MIDs are analyzed, somewhere between 1912 and 1917. By the end of the twentieth century, when the proportion of democracies was over .40, the greater peacefulness of democratic pairs of states is clear and dramatic. The risk of a fatal dispute involving two democracies is about 2/3 of that for two autocracies if originators only are included; it is only 1/10 if all fatal MIDs are compared. Comparisons at the end of the twentieth century are also very favorable to democracies if all MIDs are studied, and our analyses indicate that the democratic advantage will continue to grow if the proportion of democracies in the system increases further. Thus, these analyses provide no reason to worry that the democratic peace is becoming attenuated in the post-Cold War era.

Is there an autocratic peace? Certainly not if this is taken to mean something comparable to the separate peace enjoyed by jointly democratic dyads over the last 100–150 years. This is hardly surprising. Democracies as a group are more homogeneous than autocracies, a political category that includes communist states, fascists, personalistic rulers, and theocracies (Peceny, Beer and Sanchez-Terry 2002). It is true that autocracies are less likely to fight one another than they are to fight a democracy. In almost all the comparisons reported in Table 2, mixed dyads are by far the most conflict-prone; autocracies and democracies have much over which to fight. It is only relative to these particularly violent dyads that autocracies look peaceful.

To summarize, we find no indication that the democratic peace diminished as the proportion of democracies in the international system increased during 1816–2001. Indeed, the peacefulness of jointly democratic pairs has grown over time, leaving a wide gap in behavior between democratic and autocratic dyads. Our results make clear that the separate peace among democracies cannot be subsumed under the general rule regarding the violent consequences of political differences that G&W have proposed. Nor do we find evidence of a meaningful “autocratic peace.” Autocratic pairs are peaceful only relative to mixed dyads, which are especially prone to fight. It is encouraging that the democratic peace strengthened as liberal institutions spread beyond the North Atlantic region where they originated. The Arab Spring may after all improve the prospects for peace in the Middle East.

Conclusion: From Monads to Complex Systems

We reviewed two challenges to the democratic peace and found their conclusions unpersuasive due to problems related to theory, data, and misspecification of their models. Detecting these errors was made possible by the laudable scientific practice—adhered to by the authors—of making their data and computations available for examination by other scholars (Buono de Mesquita et al. 2003). This leads to some thoughts about the development of our field.

In the beginning, quantitative analysis of international conflict with cross-country data was essentially monadic, trying to generalize about the behavior of individual states as autonomous actors or about a few attributes of international systems, for example bipolarity or multipolarity. This led to simple additive equations where the independent variables were largely realist attributes such as power, alliances, and distance.

Further developments came in at least three forms. First was the exploration of the role of domestic politics on international relations. This meant, for example, that effective cooperation rather than conflict derived heavily from the political and economic characteristics that define the different interests of states. This required creating new data sets. Not coincidentally, the new perspectives on the sources of cooperative behavior severely damaged the realist perspective of inevitable conflict in an individualist dog-eat-dog world. A second development was the expansion of dyadic analysis to triads, then to regions or other groupings, and most recently the blossoming of complex network analysis to the entire global system. A third development arose from increasing appreciation of the challenge that causal complexity and dependence poses to causal inference using observational data. Some scholars sought to address this challenge by developing model-based techniques, such as multiple equation models and novel estimators. These techniques, however, are prone to making statistical inference more opaque, and often depend on additional conditions that are harder to verify. Other scholars have begun to employ design-based strategies for improving causal inference, leveraging experiments, sources of as-if random variation in causes of interest, causal process observations, and other techniques.

These and future developments will no doubt reshape how we think about the causes of cooperation and conflict in international relations. At present, however, none of these developments has persuasively overturned the empirical regularity first identified almost 40 years ago. In particular, the two articles addressed here do not inform our understanding of the democratic peace, and they do not provide evidence against the theory that democracy operates as a powerful force for peace. In a subject of study where reliable insights are rare, the robust finding that democracies are more peaceful toward each other remains an important empirical regularity for future scholarship to build upon.

References


The Democratic Peace


