

An Analysis of the Militarized Interstate Dispute (MID) Dataset, 1816–2001

RESEARCH NOTE

DOUGLAS M. GIBLER

University of Alabama

STEVEN V. MILLER

Clemson University

AND

ERIN K. LITTLE

University of Alabama

This research note discusses a five-year examination of the original coding work of the Militarized Interstate Dispute (MID) project. After strictly applying MID coding rules, we recommend dropping 251 cases (or over 10% of the dataset), as either we were unable to find a militarized incident in the historical record or the dispute appeared elsewhere in the data. We found evidence linking 75 disputes to other cases, and we could not identify 19 cases in the historical record. Among the remaining disputes, we recommend major changes (changes in dispute year, fatality level, and participants) in 234 disputes and minor changes in 1,009 disputes. We use this article to examine the potential impact of our suggestions on existing studies. Though we identified several systematic problems with the original coding effort, we also find that these problems do not affect current understandings of what predicts the onset of interstate conflict. However, estimates in our replications of three recent studies of dispute escalation, dispute duration, and dispute reciprocation all witness substantial changes when using corrected data—to the point of reversing previous conclusions in some cases.

The Correlates of War (CoW) Militarized Interstate Dispute (MID) dataset provides scholars with some of the best available data on international conflict between states. Its value has been immeasurable since its first release by Gochman and Maoz (1984). The dataset undergirds the most influential arguments about the relationships between joint democracy, territorial issues, alliances, arms races, rivalry, and interstate conflict. Expanded and updated further by Jones, Bremer, and Singer (1996); Ghosn, Palmer, and Bremer (2004); and Palmer et al. (2015), no other interstate conflict dataset

available to conflict scholars featured MID's temporal reach, regional inclusiveness, or consideration of interstate disputes at lower levels of hostility than open warfare. No other conflict dataset allowed for thorough comparisons of which disputes escalate toward war and which do not. The scope of MID made it the standard dataset for evaluating important claims of war and peace in the international system.

There has never been a completed systematic analysis of the cases included in the dataset. In consequence, we do not know whether the amended dataset created by such a review would still support important inferences in studies that use MID.¹ We change that here. We recently concluded a five-year study that attempted to replicate the original coding work of MID for the entirety of the dataset through version 3.1 (disputes from 1816 to 2001). After using a strict reading of MID's coding rules, we find that approximately 68 percent of the dataset needs revision. We recommend dropping 251 cases from the dataset for not meeting MID coding rules for inclusion. We also recommend merging 72 other MIDs with disputes that are connected by militarized incidents in the same area over the same issue. We could not identify 19 MIDs either using the sources identified by MID or after our own searches of the historical record. Of the remaining disputes, we recommend what we believe are major changes

Douglas M. Gibler is Professor of Political Science at the University of Alabama. He is the author of *The Territorial Peace: Borders, State Development, and International Conflict* (Cambridge University Press, 2012).

Steven V. Miller is an assistant professor in the Department of Political Science at Clemson University. His recent publications appear in *Conflict Management and Peace Science* and *Journal of Peace Research*.

Erin K. Little is a doctoral candidate at the University of Alabama studying the strategy and consequences involved in civilian targeting during international conflicts. Her other work on the MID project was just published in *Political Science Research and Methods*.

Authors' note: Our thanks go to Zeev Maoz, Glenn Palmer, Scott Bennett, Faten Ghosn, Vito D'Orazio, Michael Kenwick, Matthew Lane, Aaron Shreve, Shareefa Al-Adwani, Tracy Quo Lin, and participants at a MID conference hosted by UC-Davis, January 24–25, 2014; Matthew Lane and Glenn Palmer also provided excellent feedback on an earlier version of this article. We thank Paul Hensel for additional help locating South American disputes. We also thank Jessica Weeks and Alex Braithwaite and Doug Lemke for making their data available for replication. Finally, Doug Gibler thanks the National Science Foundation for their generous support of this project (Awards #0923406 and #1260492). Please address correspondence to Douglas M. Gibler; e-mail: dmgybler@ua.edu.

¹Others have focused on trying to source the early disputes. For example, Jessica Weeks and Mike Tomz have begun such an effort for the post-1950 disputes, called MIDipedia.

to 234 disputes (for example, change of participants, start and end year, fatalities) and minor changes to 1,009 disputes (for example, change of start and end days).

The changes we suggest are obviously substantial. They are likely to affect many inferences found in previous studies. We document several of these with replications of recent works on dispute escalation (Braithwaite and Lemke 2011), dispute duration (Gibler and Miller 2013), and dispute reciprocation (Weeks 2008). Our reanalyses of these studies suggest that interpretation of several key relationships depends wholly on improperly coded cases in the original data, and we document how specific change recommendations affect inferences from these studies.²

The note proceeds as follows. First, we start with a summary of our findings regarding the dispute data and describe the changes we suggest by both temporal domain and region. Next, we discuss how our recommended changes may affect studies of interstate conflict; we include this section as a guide for those interested in revisiting existing studies. We then analyze the revised dataset and present replications of several recent studies. We conclude with a summary of our contribution to this important dataset for conflict scholars. While we laud the original coders for their work in compiling this important dataset, we believe our findings—and our cleaned data—will increase faith in MID-based analyses of interstate conflict.

Procedures for Confirming Original MID Data

CoW defines a militarized incident as “a single military action involving an explicit threat, display, or use of force by one system member state towards another system member state.” A MID or dispute is the aggregation of these incidents over time, space, and issue; or, more formally: “Militarized interstate disputes are united historical cases of conflict in which the threat, display or use of military force short of war by one member state is explicitly directed towards the government, official representatives, official forces, property, or territory of another state” (Jones, Bremer, and Singer 1996, 169).

Incident coding rules are extensively discussed in the data codebooks and also the published articles releasing the several iterations of the dataset. Similarly, each publication discusses how these incidents are aggregated into a dispute. For example, in Jones, Bremer, and Singer (1996, 174–77), these aggregation rules stipulate that incidents that form a dispute must (1) involve the same or overlapping parties, (2) involve the same issue(s), and (3) have identifiable start and end dates. The coding rules also provide information for how to treat wars, civil wars, and other concerns. Together, these rules describe how the dataset comes from (1) the identification of militarized incidents between and among states in the CoW

²Since we are suggesting substantial changes to one of the most important datasets in international relations research, we provide a great deal of background material on specific cases. These are included in the appendices to this manuscript. Appendix I documents several issues related to the sources the original coders used and how this source information likely led to systematic problems when coding the dataset. Appendix IIA provides information about the dispute cases we could not find. Appendix IIB details our drop recommendations by dispute number. Appendix IIC presents information about the cases that should be merged with other disputes. Appendix IID provides summary information for remaining change recommendations. Appendix IIE details our discussions with the CoW Project regarding our review, and, finally, Appendix IIF provides a comprehensive bibliography for all disputes, by dispute number.

international system and then (2) on the method of aggregation of these incidents into unified conflicts over time.

CoW provides incidents for all disputes between 1993 and 2010, but the incident data for disputes prior to 1993 has never been made available. CoW also provides a list of sources for all MID data, available at <http://www.correlatesofwar.org/>. The source data for MID3 (covering the years 1993 to 2001) and MID4 (2002–2010) are much more extensive than the original MID2 data; nevertheless, the source information for the cases between 1816 and 1992 was very helpful in assessing the majority of the cases during that time period.

We began the re-creation of the more than 2,000 cases of disputes between 1816 and 1992 by first finding the source information in the MID bibliography. In most cases this information proved inadequate to completely code the dispute, and, in a substantial number of cases, we could not confirm the existence of a dispute based on the source. We conducted extensive Internet, newspaper, and scholarly literature searches for information on any and all conflicts related to the participant and dispute data provided by CoW. We compiled an extensive bibliography of that source information, disaggregated by dispute number. This bibliography is contained in the final appendix to this manuscript.

The results from our re-creation of the dataset can be subdivided among five types of cases—disputes that we could not find in the historical record, cases that were found but for which information suggests no militarized dispute occurred, disputes for which major changes need to be made to the original codings, minor-change cases, and, finally, disputes for which the historical record matches the original coding from MID2. We focus first on the dataset cases that were most difficult to validate.

We identified a very small minority of disputes—19 in all—for which neither the original source nor our extensive searches provided information on any type of conflict between the states listed in the MID data on the dates provided or anytime temporally proximate to those dates.³ We recommend excluding these cases from analyses. We provide short narratives of the events in the dyad-dates listed by the MID project for these cases and include these as Appendix IIA to this manuscript.

After our review, we also recommend dropping from the dataset approximately 10 percent of the original cases. For each of these 240 cases, we first established the historical event identified by the original coders, and we then applied the MID3 coding rules to the source information we found. We were incredibly conservative in our re-creation efforts and accepted existing data whenever possible. However, we believe that the cases we list as drop recommendations cannot be substantiated as MIDs using the CoW coding rules for militarized incidents, and we provide descriptions for our reasoning for each of these cases in Appendix IIB of this manuscript.⁴

There are also numerous cases in which we discovered new information that linked previously separate disputes.

³We consulted the CoW-provided sources for each dispute, and, in many cases, these provided help identifying the likely conflict. Still, there is no mention of a dispute in the labeled sources for these 19 cases. Our meetings with CoW and their examination of dataset files also provided no information on these cases.

⁴We should note that we have had several discussions with MID personnel regarding our drop recommendations. They actually agree with the majority of our recommendations, though not all of them, and we provide an overview of the discussions we have had regarding these disagreements in Appendix IIE of this article.

The MID coding rules stipulate that six months of continued inaction ends a dispute, with the last recorded militarized incident coded as the end date (Jones, Bremer, and Singer 1996, 175–76). We purposely searched all same-dyad disputes to determine whether there were militarized incidents that were missed by the original coders. In all, we found 74 disputes that should be merged with existing disputes and dropped from the dataset. These cases are described in Appendix IIC of the manuscript.

Of the 1,238 cases for which we recommend changes, 234 involve major changes to the data. We define “major change” according to whether we believe the change may influence inferences in a substantial number of studies. These major changes include corrections to the number of participants in a dispute, the actual parties to the dispute, the dispute start year, or a change in fatalities from zero to one or more, from a fatal to a non-fatal dispute, or from missing data to fatal dispute. Many disputes had multiple major change recommendations. All other change recommendations were, we believe, relatively minor in comparison and include such things as day/month changes, end year changes, highest action, hostility level, outcome, and/or settlement changes. We found minor changes needed for 1,009 of the disputes cases we examined. Finally, we were able to re-create the original dispute coding perfectly in 758 cases of disputes, or approximately 32 percent of the cases between 1816 and 2001.

Summarizing Our Change Recommendations

Table 1 describes our review recommendations by region and time period and shows that we actually find no clear spatial or temporal pattern in these changes. The could-not-find cases were most numerous among European dyads, so these missing cases are probably not due to the scarcity of information available to CoW. The post-World War II time period is most numerous in the drop and merge categories, but this time period also has the highest number of disputes. Overall, there seems to be no clear area or time-period prediction for when we had difficulty re-creating the original data, and this lack of systematic bias may be good news for existing studies.

A few patterns do emerge when we divide our general recommendations according to the highest action level in the dispute, as we provide in Table 2. We found several issues related to disputes classified as seizures in the dataset, and these cases represent a substantial number of each category. Among other categories, we recommend drops of 35 cases that were originally coded as threats; in most of these cases, the threat was too vague to constitute a militarized incident. Finally, a majority of our merge and major change recommendations concern cases of attacks and clashes. These were cases in which the original coders found good information about the occurrence of militarized incidents, at intense levels of contestation, but our additional searches provided new information that more properly described the militarized dispute.

Appendix IID provides information about the types of changes we recommend for every dispute in the MID dataset. Our replication data for this manuscript also provides a clean set of confirmed disputes and dyad-years in dispute for the years 1816 to 2001.⁵

Analysis of Recommended Changes

Thus far we have described our review of the MID project and several issues we found with the data. The important point of this review, though, is determining whether the changes we recommend for the dataset are systematic enough to change existing conclusions on the causes of conflict. We do this in two parts. First, in this section we provide an analysis of several predictors of our categories—could-not-finds, drops, merges, and major and minor change recommendations—in order to determine the circumstances under which various studies will be affected. For example, if we find that a particular time period, region, or type of dispute was prone to being poorly coded and should be dropped from the data, then researchers whose work has focused on these cases may want to re-examine their previous studies.

Table 3 presents the results of five logistic regressions in which we used the issues we found as predictors of the various categories. Each of the models uses the entire dataset of dispute cases we analyzed, 2,330 in all, and each predictor was included in each of the models analyzed. Table 3 provides a + sign when the predictor is a positive predictor of the category and a – sign when the variable predicts fewer of the category; only predictors that are statistically significant from zero at $p < 0.05$ are included.

Several interesting patterns emerge from this analysis. In the first column, Table 3 shows that seizures and disputes that were coded as lasting for only one day were difficult for us to find. Eleven of the nineteen cases in this category were coded as lasting one day, and there were six seizures, two of which lasted longer than one day. Again, we consulted the original sources from CoW and conducted extensive searches of the historical record, but we still found no information regarding these events.

The second column suggests that there may be systematic patterns in the cases we recommend dropping. Seizures and threats to use force are highest action levels that suggest a drop recommendation was more likely; disputes coded as clashes were unlikely to be dropped. Among the clustering variables, those cases that were related to the Tanker War and during the 1900 to 1945 time period were more likely to be dropped, and, regionally, cases in the Middle East had a higher drop rate. The original source information was also a consistent predictor for this category. Cases using Langer (1972), Facts on File, or Keesing’s were much less likely to be dropped; this is most likely due to the fact that these sources were used for coding major international events over time.

We found only positive predictors of the cases that should be merged. Attacks and clashes were likely separated by the six-month rule in the original data, but, even controlling for those cases, we were able to find continuous action in many instances. Our searches were also more likely to find information that merge disputes in Asia and the Middle East as well as events that were sourced using newspapers or daily and weekly bulletins. We believe this last finding underscores our argument in Appendix I that caution should be used when aggregating incidents based solely on news reports.

The predictors of major and minor changes were quite different, as would be expected. We found major changes necessary in attacks, clashes, and “joins ongoing war” cases, and the Middle East proved to be difficult to recreate without substantial coding changes. Cases coded using Keesing’s or Dupuy and Dupuy (1986) also were likely to need major recoding. The pattern for seizures is striking here when

⁵This data is available at the following Dataverse page: <https://dataverse.harvard.edu/dataverse/dmgibler>. Script and data for non-directed and directed dyad-year data can be found at <https://github.com/svmiller/gmlmid-data>.

Table 1. Significant changes recommended, by region and time period

		<i>North America</i>	<i>South America</i>	<i>Europe</i>	<i>Africa</i>	<i>Middle East</i>	<i>Asia</i>	<i>Total by period</i>	<i>Total by Type</i>
Could not find	1816–1899		1	3				4	
	1900–1945		1	6	3			10	
	1946–1992			2	1	2		5	
	1993–								19
Drop	1816–1899	10	7	4		2	2	25	
	1900–1945	3	19	40	1	3	9	75	
	1946–1992	5	25	25	23	29	24	130	
	1993–2001		2	1	3	2	6	14	245
Merge	1816–1899		4					4	
	1900–1945	2	2	3			5	12	
	1946–1992	2	1		4	20	24	51	
	1993–2001		1	2			2	5	72
Major change	1816–1899	5	7	7		4	2	25	
	1900–1945	7	3	22		7	5	44	
	1946–1992	16	3	19	25	39	41	143	
	1993–2001	2			4	7	4	17	229
Total by region	52	76	134	63	115	124			

Table 2. Significant changes recommended, by highest action level

	<i>Threats</i>	<i>Shows/Alerts</i>	<i>Seizures</i>	<i>Attacks/Clashes</i>	<i>Declare/Join War</i>	<i>Other</i>	<i>Total</i>
Could not find	3	2	6	6	1	1	19
Drop	35	28	53	85	8	36	245
Merge	2	5	11	44	2	8	72
Major change	6	24	19	143	16	21	229

Table 3. Predictors of recommended change

	<i>Could Not Find</i>	<i>Drop</i>	<i>Merge</i>	<i>Major Change</i>	<i>Minor Change</i>	<i>No Change</i>
<u>Dispute Characteristics</u>						
Seizure	+	+		–		–
Threat		+			–	
Attack			+	+		
Clash		–	+	+	–	
Join ongoing war				+		
One-day dispute	+	+			+	+
Six-month rule			+			
All democracies						
<u>Temporal and/or Spatial Clustering</u>						
Tanker War 1816 to 1899		+			+	–
1900 to 1945		+			+	–
1946 to 1992					+	–
Africa (sub-Saharan)					–	
Asia			+			
Middle East		+	+	+		
North America						
South America						
World War I		+				
World War II						–
<u>Source Information</u>						
Dupuy and Dupuy				+	–	
Langer		–				
New York Times			+		+	–
Facts on File		–	+			
Keesing's		–		+		
Number of sources					+	

compared to the other categories: these cases are less likely to be found and more likely to be dropped, but, if found and correctly coded, no major changes are likely. Minor changes were likely in all time periods except those cases covered by MID3, the base category, which again makes sense considering the data and funding available to the project coding 1993 to 2001. Minor changes were also less likely for threats and clashes and for cases in Africa, but changes were more likely for disputes sourced with only the *New York Times*. Very strangely, we found that the number of sources given by the CoW project for the 1816 to 1992 disputes were a predictor of minor changes needed. Perhaps this results from the tendency to name several standard sources when a certain region or time period is coded.

When Our Recommended Changes May Affect Conflict Studies

We have identified several patterns among the change suggestions we offer, and we use this section to examine whether—or, more properly, when—these changes will matter to a given study. We begin by returning to the “dangerous dyads” framework first offered by Bremer (1992) and analyze the predictors of dispute onset at the dyadic level. We then reanalyze three recent studies that focus on events that occur within a dispute. As we demonstrate, there are substantial inferential differences in the latter type of analyses when using our revised data.

Dangerous Dyads and Dispute Onset

We use several known predictors of interstate conflict in our analysis of conflict onset. First, the presence of land contiguity in the dyad signals an opportunity for conflict, increased interactions, or the presence of territorial issues. We use the Stinnett et al. (2002) data for this measure. We measure parity as the smaller CINC score divided by the total CINC scores in the dyad, using data from Singer, Bremer, and Stuckey (1972). We include a measure for the presence of a major state in the dyad, and we code whether the states were in a formal alliance (Gibler and Sarkees 2004). Joint democracy is included in the models, with democracy defined as a Polity IV score between 6 and 10 (Marshall and Jaggers 2002). Each of these independent variables is lagged by one year from the dyad-year observation. Finally, we estimate each model using general estimation equations (GEE) with a binomial distribution and a one-year time lag to control for autocorrelation, and we use all non-directed dyad-years for which independent variable data was available for the years 1816 to 2001.⁶

Table 4 presents the results of these analyses. The dependent variable of dispute onset in each analysis is the presence of two dispute originators in the first year of conflict, with all subsequent years of conflict set to zero. The first column uses the currently available MID data from the CoW project.⁷ This estimation demonstrates that our basic model of conflict conforms well to general expectations from the literature. Contiguity, parity, and the presence of a major state are each associated with an increased likelihood of dispute onset; joint democracy and the presence of an alliance in the dyad decrease the risk of dispute onset, though the presence of an alliance is not statistically significant. This is interesting given previous

efforts to connect interests and democracy as competing explanations of conflict (see, for example, Gowa 1999).

Columns two and three introduce our recommended changes to the data. Column two applies the corrected dispute years to the CoW MID onsets, and column three applies our drop recommendations. The fourth column applies all changes and treats as peaceful the dyad-years that had a seizure, joined an ongoing war, or had a dispute related to the Tanker War between Iraq and Iran.

Evident across these models is that our suggested changes make little difference in the estimation of conflict onset. There are no changes in the direction of relationships, there are no other changes in statistical significance, and there is very little change in the relative strength of relationships between the independent variables and dispute onset. Not even the omission of several large groups of dispute cases affects these findings. Seizures and war joins are in many ways dissimilar to the rest of the dataset, as we detail in Appendix I, and the often unreciprocated shipping strikes of the Tanker War are so numerous and strongly clustered they could bias several relationships. Nevertheless, these omissions produce estimates that are quite similar to the other models.

The lack of change across models is most likely due to the large number of dyad-years in the analysis and the large number of event observations (over 650,000 and over 2,000, respectively). These sample and event sizes make it difficult for any of our recommended changes to substantially alter found relationships in the data. This changes, however, when analyses concern mostly what took place once fighting began in the dyad.

Replication 1: Dispute Escalation

A recent study by Braithwaite and Lemke (2011) provides an analysis of all bilateral MID to determine the correlates of dispute escalation. They define escalation based upon the level of severity that a dispute reached, measured in terms of reciprocation by the target, uses of force by one or both sides, and various fatality levels. The predictors of escalation include the presence of joint democracy, whether the dispute was a territorial issue, the share of capabilities held by the initiator, the presence of a defense pact, and the status quo evaluation of both states. Braithwaite and Lemke structure the escalation process with common predictors of dispute onset and use a censored probit to jointly estimate onset and escalation. The dependent variables in the outcome equation (escalation) include whether the dispute was reciprocated by the target, whether the dispute reached a hostility level that included the use of force, whether both sides reached use of force, and three different dispute fatality levels (any military deaths, more than 250 military deaths, and 1,000 or more military deaths).⁸

Fatality and hostility levels are primary variables of interest in the MID project, and the escalation study provides an excellent vehicle for exploring whether our recommended changes are likely to alter findings related to these variables.

⁶The independent variables used in the estimates are well described in Braithwaite and Lemke (2011). Dispute onset is predicted by the presence of contiguity (Stinnett et al. 2002), the presence of a rivalry (Thompson 2001), joint minor status in the dyad, the presence of a territorial claim (Huth and Allee 2002), the share of capabilities of the initiator (Singer, Bremer, and Stuckey 1972), joint satisfaction with the status quo (Signorino and Ritter 1999), the presence of a defense pact (Gibler and Sarkees 2004), and temporal controls for duration dependence (Carter and Signorino 2010). The escalation equation omits several of these variables but adds whether one of the states in the dyadic dispute is fighting over a territorial issue (Ghosn, Palmer, and Bremer 2004).

⁶The most recent iteration of the MID dataset, from 2002 to 2010, was released as we were finishing our review. We have not yet examined this data.

⁷Downloaded on October 28, 2014, from <http://correlatesofwar.org/datasets/MIDs>.

Table 4. Analysis of MID predictors following dispute sample changes, 1816–2001

	<i>Current CoWMID</i>	<i>After Year Corrections</i>	<i>After Drop Recommendations</i>	<i>Omissions for Robustness[†]</i>
Contiguity	3.073*** (0.168)	3.069*** (0.167)	3.130*** (0.168)	3.339*** (0.171)
Joint Democracy	-0.861*** (0.147)	-0.862*** (0.147)	-0.919*** (0.158)	-0.888*** (0.169)
Capability Ratio	1.891*** (0.406)	1.867*** (0.403)	1.866*** (0.409)	1.742*** (0.413)
Allied	-0.189 (0.141)	-0.180 (0.141)	-0.203 (0.144)	-0.234 (0.146)
Major State Present	1.763*** (0.175)	1.760*** (0.174)	1.727*** (0.176)	1.640*** (0.181)
Constant	-6.794*** (0.105)	-6.795*** (0.105)	-6.941*** (0.106)	-7.157*** (0.108)
<i>N</i>	654,513	654,513	654,513	654,513

[†]This model excludes cases of seizures, joins ongoing war, and disputes related to the Tanker War. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In Table 5 we list the estimates of the same models as Braithwaite and Lemke (2011, Table 1) using the dispute data with our recommended changes. We highlight with shaded boxes any estimates that changed sign or gained or lost statistical significance (at $p < 0.05$) in the reanalysis.

Consistent with our earlier findings, there were no substantive changes from the original models predicting bilateral dispute onset. However, there were numerous changes across the outcome equations. Joint democracy was a statistically significant promoter of the use of force, mutual use of force, and the escalation to war in Braithwaite and Lemke (2011, Table 1). That changes when using our revisions—the new estimates show no statistically significant result in either revised model. In the war equation the standard error is now almost as large as the coefficient. Very little theory suggests why joint democracy would escalate disputes to war, and our analyses are consistent with a selection effect in which democracies are unlikely to fight but have no consequential effect should fighting begin (Reed 2000).

Joint satisfaction is no longer statistically significant in three of the models—reciprocation, use of force by the initiator, and fatalities of 250 or more.⁹ Now, only mutual uses of force and fatal disputes are less likely when both states are satisfied with the status quo (see also Werner 1999). Our changes also eliminate the statistical significance of power preponderance in one model (250+) but give defense pacts statistical significance in another model (mutual force).

We find important changes in the joint estimation of two models—the initial use of force and war outcome variables—and believe both changes are due to the elimination of several disputes from the dataset. First, the ρ estimate in Table 5 suggests that initiators are more likely to use force if the state does decide to start a dispute. We eliminated many cases that were not disputes but were originally coded as low-level threats and displays of force. Historical coverage was originally poor for many of these cases, and our expanded searches corrected the cases as lacking any militarized action. With these low-level cases omitted from the analyses, dispute initiations are relatively more likely to involve the use of force should disputes occur.

The second changed ρ also results from altering the treatment of several cases. The original data had nine cases

of fatalities between democracies that reached 1,000 or more deaths; however, our bilateral data suggest only four cases should be present. Two changes were made by our data—MID#1293 between Poland and Lithuania should be merged with their earlier conflict because of continuous action, and MID#1786 double-counts a British declaration of war on Finland and should be dropped. Two cases involve 1974 disputes between Greece and Turkey that did not end in war, though the Cyprus War is already in the data for that year. Similarly, Hungary declared war on Romania in 1944, but the fatalities are missing data in this dispute. Both were fighting in World War II, and that dispute was coded as having fatalities that crossed the war threshold. Changing the treatment of these cases eliminates the effect of joint democracy on escalation.

Our recommended changes alter the empirical predictors of escalation in a manner consistent with existing theories of conflict. Though dyadic predictions of onset remain unchanged, we have evidence that relationships involving several different dispute characteristics may be altered, even among primary variables like fatality and hostility level. We explore the effects of our changes on other variables in the next replication when we re-examine the predictors of dispute duration.

Replication 2: Dispute Duration

Dispute end dates are one of the variables we found most likely to be either missing or miscoded, followed closely by start dates, so it makes sense to assess whether these changes affect our understanding of dispute duration. We do that in this section with a reanalysis of the Gibler and Miller (2013) study that examined the effects of contiguity, territory, and democracy on the length of a dispute. Their study was a response to the many arguments suggesting that democracies are good at selecting the disputes they escalate and fight shorter conflicts overall (see, for example, Reiter and Stam 2002).

Their estimation strategy used Ordinary Least Squares estimates of directed dispute dyads from 1816 to 2001. The dependent variable was the length of each dispute, in days, and the predictors included dummy variables for the presence of a democratic initiator, a democratic target, contiguity, an alliance in the dyad, and whether the issue was territorial. They also included a continuous measure of capability share held by the initiator and two interaction terms—an interaction of

⁹Our replication of Braithwaite and Lemke (2011), using the available replication data, is marginally different from the published results, so we highlight differences from the published results but report the replication models.

Table 5. Reanalysis of Braithwaite and Lemke (2011, Table 1)

	Reciprocation		Use of Force		Mutual Force		Fatalities > 0		Fatalities > 250		Fatalities War	
	(Original MID)	(Revised MID)	(Original MID)	(Revised MID)	(Original MID)	(Revised MID)	(Original MID)	(Revised MID)	(Original MID)	(Revised MID)	(Original MID)	(Revised MID)
Joint democracy	0.016 (0.127)	0.24 (0.141)	-0.029 (0.162)	0.244 (0.134)	0.296 (0.161)	0.196 (0.138)	0.104 (0.156)	0.13 (0.151)	0.382 (0.255)	0.376 (0.192)	0.365 (0.261)	0.365 (0.261)
Territorial MID	0.615*** (0.106)	0.549*** (0.115)	-0.105 (0.103)	0.517*** (0.112)	0.544*** (0.132)	0.478*** (0.110)	0.413*** (0.116)	0.253* (0.124)	0.682*** (0.195)	0.406** (0.142)	0.635** (0.198)	0.635** (0.198)
Joint satisfaction	-0.219 (0.138)	-0.043 (0.146)	-0.274 (0.142)	-0.388** (0.148)	-0.488** (0.182)	-0.602*** (0.156)	-0.489** (0.181)	-0.346* (0.159)	-0.055 (0.358)	0.174 (0.198)	0.265 (0.357)	0.265 (0.357)
Power preponderance	-0.256 (0.292)	-0.345 (0.315)	-0.031 (0.301)	0.003 (0.325)	0.127 (0.355)	0.117 (0.297)	-0.118 (0.356)	0.664* (0.312)	-0.257 (0.575)	0.436 (0.439)	0.016 (0.735)	0.016 (0.735)
Defense pact	0.054 (0.129)	0 (0.124)	-0.02 (0.127)	0.105 (0.127)	0.321* (0.137)	0.199 (0.13)	0.171 (0.141)	0.11 (0.137)	-0.121 (0.369)	-0.309 (0.216)	-0.475 (0.469)	-0.475 (0.469)
Constant	0.716* (0.281)	0.881** (0.322)	0.744** (0.276)	-0.012 (0.336)	-0.236 (0.353)	-0.117 (0.278)	-0.325 (0.324)	-1.150*** (0.272)	-1.304** (0.46)	-1.673*** (0.428)	-1.751** (0.591)	-1.751** (0.591)
Contiguous	1.069*** (0.058)	1.080*** (0.060)	1.068*** (0.058)	1.069*** (0.058)	1.077*** (0.06)	1.068*** (0.058)	1.076*** (0.06)	1.068*** (0.058)	1.072*** (0.06)	1.063*** (0.059)	1.072*** (0.06)	1.072*** (0.06)
Rivalry	0.815*** (0.08)	0.733*** (0.083)	0.809*** (0.081)	0.815*** (0.08)	0.737*** (0.083)	0.814*** (0.08)	0.737*** (0.083)	0.809*** (0.081)	0.739*** (0.083)	0.811*** (0.081)	0.736*** (0.084)	0.736*** (0.084)
Minor-minor	-0.700*** (0.047)	-0.632*** (0.051)	-0.707*** (0.047)	-0.698*** (0.047)	-0.635*** (0.051)	-0.700*** (0.047)	-0.637*** (0.051)	-0.708*** (0.047)	-0.645*** (0.051)	-0.712*** (0.047)	-0.645*** (0.051)	-0.645*** (0.051)
Joint democracy	-0.203*** (0.055)	-0.238*** (0.062)	-0.204*** (0.055)	-0.203*** (0.055)	-0.238*** (0.063)	-0.203*** (0.055)	-0.238*** (0.063)	-0.204*** (0.055)	-0.238*** (0.063)	-0.204*** (0.055)	-0.238*** (0.063)	-0.238*** (0.063)
Joint satisfaction	-0.007 (0.055)	-0.014 (0.056)	-0.009 (0.055)	-0.007 (0.055)	-0.015 (0.056)	-0.007 (0.055)	-0.015 (0.056)	-0.008 (0.055)	-0.017 (0.056)	-0.01 (0.055)	-0.017 (0.056)	-0.017 (0.056)
Territorial claim	0.505*** (0.068)	0.476*** (0.071)	0.510*** (0.069)	0.508*** (0.068)	0.477*** (0.071)	0.508*** (0.068)	0.477*** (0.071)	0.509*** (0.068)	0.476*** (0.071)	0.512*** (0.068)	0.479*** (0.071)	0.479*** (0.071)
Power preponderance	-0.281* (0.142)	-0.326* (0.152)	-0.295* (0.143)	-0.278* (0.142)	-0.329* (0.153)	-0.283* (0.142)	-0.332* (0.153)	-0.295* (0.142)	-0.342* (0.153)	-0.301* (0.143)	-0.343* (0.153)	-0.343* (0.153)
Defense pact	-0.001 (0.06)	0.006 (0.063)	-0.001 (0.061)	-0.002 (0.064)	0.007 (0.064)	-0.002 (0.061)	0.007 (0.064)	-0.001 (0.061)	0.009 (0.064)	0.001 (0.061)	0.009 (0.064)	0.009 (0.064)
Peace years	-0.026*** (0.003)	-0.026*** (0.003)	-0.026*** (0.003)	-0.026*** (0.003)	-0.026*** (0.003)	-0.026*** (0.003)	-0.026*** (0.003)	-0.026*** (0.003)	-0.026*** (0.003)	-0.025*** (0.003)	-0.026*** (0.003)	-0.026*** (0.003)
Constant	-2.050*** (0.146)	-2.150*** (0.157)	-2.035*** (0.147)	-2.056*** (0.146)	-2.146*** (0.158)	-2.050*** (0.147)	-2.142*** (0.158)	-2.035*** (0.147)	-2.128*** (0.158)	-2.027*** (0.147)	-2.127*** (0.158)	-2.127*** (0.158)
Rho	-0.331*** (0.062)	-0.373*** (0.071)	0.061 (0.06)	-0.286*** (0.07)	-0.241** (0.076)	-0.284*** (0.063)	-0.214** (0.071)	-0.214** (0.067)	-0.345* (0.145)	-0.281** (0.092)	-0.306 (0.181)	-0.306 (0.181)
Stage 1 N	499,185	499,185	499,185	499,185	499,185	499,185	499,185	499,185	499,185	499,185	499,185	499,185
Stage 2 N	1,392	1,392	1,392	1,392	1,392	1,076	1,076	1,076	1,076	1,076	1,076	1,076

Standard errors in parentheses; peace splines omitted from table. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 6. Reanalysis of Gibler and Miller (2013, Table 3)

	Duration (days) of MID Initiated by State A							
	Original MID Data				Revised MID Data			
	(Model 1)	(Model 2)	(Model 3)	(Model 4)	(Model 1)	(Model 2)	(Model 3)	(Model 4)
State A is democracy	-3.052 (16.324)	-4.316 (16.44)	-32.118 (22.089)	-34.235 (22.263)	7.158 (17.406)	6.214 (22.339)	53.043 (34.845)	53.059 (38.091)
State B is democracy	-46.640** (15.347)	-46.993** (15.359)	-49.201** (15.394)	-49.671** (15.408)	-67.323*** (16.252)	-67.441*** (16.348)	-68.067*** (16.256)	-68.065*** (16.349)
Territorial dispute	54.209** (16.870)	53.787** (16.884)	55.323** (16.869)	54.855** (16.882)	69.812*** (17.343)	69.788*** (17.350)	69.777*** (17.339)	69.777*** (17.346)
Contiguous dyad	34.563* (15.446)	34.149* (15.461)	18.561 (17.479)	17.724 (17.514)	-46.734** (16.292)	-47.312* (18.411)	-48.941** (16.353)	-48.932** (18.438)
Allied dyad	-80.523*** (18.256)	-80.678*** (18.26)	-81.069*** (18.248)	-81.262*** (18.251)	-83.280*** (20.032)	-83.266*** (20.036)	-82.195*** (20.04)	-82.195*** (20.044)
Initiator's capability share	-0.011 (0.023)	-0.059 (0.077)	-0.009 (0.023)	-0.066 (0.077)	18.048 (22.612)	18.149 (22.665)	36.945 (25.8)	36.944 (25.822)
Initiator is hegemon	11.389 (28.442)	11.113 (28.449)	25.709 (29.357)	25.701 (29.359)	-105.282** (34.615)	-104.809** (35.325)	-95.491** (35.202)	-95.498** (35.846)
Contiguity X	0.053 (0.081)		0.062 (0.081)		2.413 (35.779)		-0.037 (35.808)	
Initiator capabilities X			63.742 (32.657)	65.143* (32.711)			-78.702 (51.779)	-78.705 (51.84)
State A is democracy								
Constant	153.491*** (13.219)	154.856*** (13.384)	162.466*** (13.988)	164.261*** (14.184)	239.611*** (17.603)	239.857*** (17.981)	230.903*** (18.509)	230.899*** (18.921)
N	2,366	2,366	2,366	2,366	3,163	3,163	3,163	3,163

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

democratic initiator and capabilities and an interaction of democratic initiator and contiguity.

We should note that we made one significant set of changes in how dates were treated in our reanalysis. Disputes with missing dates were omitted in the original study, but, based on our analysis of the dispute data, we were able to infer likely durations in these missing-day cases. There are 653 dyadic-dispute cases with at least one missing day: 234 cases have a missing start day, 187 have a missing end day, and 232 have both days listed as missing in disputes lasting less than one month. Based on our searches of historical records, we are confident that the cases in which both days are missing are cases in which an exact day could not be determined for when the militarized action occurred; we coded all these cases as one-day disputes. We coded disputes with missing start days as beginning on the first of the month, and we coded cases with missing end days as ending on the 30th of each month. As in the original study, we continued to omit the eight cases that had missing start or end months.¹⁰ Our reanalysis of this study is presented in Table 6.

Consistent with the primary findings of Gibler and Miller (2013), our new estimates suggest that territorial disputes last much longer—an average of 70 days or 30 percent longer than other disputes, all else equal. This is actually a stronger result than what was found in the original study, which had a territorial dispute average of approximately 54 days. Again, too, the presence of a democratic initiator in the dispute has no effect. However, we do find three important changes in the new estimates. First, contiguity reverses

¹⁰This treatment of missing dates re-introduces almost 650 additional dyadic-dispute cases into the analysis. Two hundred thirty-two were already coded by CoW as being less than one month, and their exclusion can introduce important biases in the analysis, as we demonstrate below. The remaining cases will, at worst, only be slightly altered (30 or fewer days' duration) from the actual duration of the event, and we can find no correlation between this treatment and any variables of interest in the original study.

signs (in two models) and is now a statistically significant—and negative—predictor of dispute length in all four models. Second, a challenge by a hegemon is now statistically significant and also negative; in Gibler and Miller (2013) this variable was not significant at any conventional level. Third, the contiguity X democratic initiator interaction term is no longer statistically significant in the fourth model.

Our revised treatment of missing start and end days added 645 directed dispute dyads for analysis that were of very short duration. The vast majority of these cases, too, were disputes involving contiguous states, often in regions and time periods that had spotty contemporary coverage. Including these cases is important for understanding the effects of contiguity on dispute duration since their omission inflated the length of time contiguous states seemingly fought in the data.

None of the cases involving a hegemon had missing days. Instead, the changes resulted directly from our recommended changes to several US-initiated cases that proved highly influential in the Gibler and Miller (2013) study. The United States was on Side A for ninety-four disputes since 1945; Britain was also Side A for ninety-four disputes between 1816 and 1939. Our date changes actually increased the duration in twenty-eight of the US cases and decreased the duration of only nine US cases. However, the changes for the latter nine were more substantial, including an end-date change of three years (MID#3551).¹¹ Meanwhile, twenty-four British dispute durations increased for the period when they

¹¹CoW codes the end date for MID#3551 (a response to Yugoslav involvement in the Bosnian Civil War) as the 1996 passage of UN Resolution 1074, which declared an end to the blockade of Yugoslavia. However, blockades should be coded as six-month counts according to Jones, Bremer, and Singer (1996, 176). A separate blockade by the United States started on May 1, 1993, and is coded by CoW as ending by a six-month count on November 1, 1993. That date should be the end date for the dispute. (Also, CoW erred when coding the year of the blockade in the incident data; the blockade began on November 21, 1992, not 1993.)

Table 7. Reanalysis of Weeks (2008, Table 4)

	<i>Original MID Data</i>				<i>Revised MID Data</i>			
	<i>(Model 1)</i> <i>Nondemocracies</i> <i>base category</i>	<i>(Model 2)</i> <i>Democracies</i> <i>base category</i>	<i>(Model 3)</i> <i>Bilateral</i> <i>disputes only</i>	<i>(Model 4)</i> <i>Nondemocracies;</i> <i>personalists base</i>	<i>(Model 1)</i> <i>Nondemocracies</i> <i>base category</i>	<i>(Model 2)</i> <i>Democracies</i> <i>base category</i>	<i>(Model 3)</i> <i>Bilateral</i> <i>disputes only</i>	<i>(Model 4)</i> <i>Nondemocracies;</i> <i>personalists base</i>
Democratic (<2 years)	-0.400*				-0.018			
	(0.171)				(0.210)			
Personalist		0.911***	0.707**			0.407	0.441	
		(0.209)	(0.219)			(0.249)	(0.232)	
Single-party		-0.016	0.224	-0.959***		-0.179	0.130	-0.729**
		(0.240)	(0.261)	(0.244)		(0.273)	(0.289)	(0.268)
Military		0.475	0.340	-0.401		0.254	0.373	-0.159
		(0.345)	(0.354)	(0.349)		(0.395)	(0.400)	(0.386)
Hybrid		0.292	0.584	-0.647		0.137	0.371	-0.281
		(0.974)	(1.059)	(1.020)		(1.281)	(0.978)	(1.324)
Mixed nondemocratic		0.023	-0.063	-0.746**		-0.430	-0.323	-0.782**
		(0.242)	(0.244)	(0.231)		(0.285)	(0.256)	(0.247)
Dynastic Monarchy		-0.043	-0.160	-0.977		-0.052	0.075	-0.535
		(0.596)	(0.623)	(0.579)		(0.642)	(0.660)	(0.625)
Nondynastic Monarchy		0.420	0.431	-0.532		0.359	0.670	-0.124
		(0.399)	(0.418)	(0.389)		(0.453)	(0.462)	(0.429)
Nondemocratic interregna		0.492*	0.555*	-0.354		0.092	0.154	-0.344
		(0.206)	(0.221)	(0.184)		(0.248)	(0.236)	(0.208)
New democracy		0.772*	0.660			0.519	0.545	
		(0.340)	(0.356)			(0.364)	(0.370)	
Polity score				-0.022				-0.008
				(0.015)				(0.016)
Major-Major	-0.180	0.254	-0.114	0.377	-0.102	0.165	0.048	0.394
	(0.288)	(0.307)	(0.352)	(0.388)	(0.325)	(0.355)	(0.397)	(0.431)
Minor-Major	-0.006	0.078	0.101	0.245	-0.243	-0.168	0.162	0.075
	(0.229)	(0.231)	(0.244)	(0.269)	(0.259)	(0.263)	(0.268)	(0.290)
Major-Minor	0.197	0.532*	0.355	0.429	0.224	0.414	0.266	0.475
	(0.202)	(0.228)	(0.257)	(0.294)	(0.225)	(0.259)	(0.277)	(0.321)
Initiator capabilities share	-0.192	-0.196	-0.254	-0.341	-0.297	-0.264	-0.312	-0.335
	(0.230)	(0.234)	(0.254)	(0.274)	(0.255)	(0.260)	(0.282)	(0.306)
Contiguity	0.530***	0.584***	0.488**	0.770***	0.421**	0.442**	0.683***	0.710***
	(0.136)	(0.142)	(0.165)	(0.173)	(0.158)	(0.167)	(0.178)	(0.191)
Ally	-0.114	-0.133	-0.084	-0.286	0.035	-0.009	0.014	-0.088
	(0.161)	(0.164)	(0.179)	(0.183)	(0.175)	(0.177)	(0.188)	(0.196)
Alliance portfolio similarity	0.257	0.339	0.534*	0.340	0.266	0.335	0.474	0.338
	(0.211)	(0.211)	(0.237)	(0.268)	(0.241)	(0.241)	(0.265)	(0.306)
SQ evaluation initiator	0.135	0.037	0.238	0.018	0.416	0.448	0.560	0.316
	(0.280)	(0.303)	(0.335)	(0.412)	(0.300)	(0.329)	(0.362)	(0.451)
SQ evaluation target	-0.175	-0.153	-0.307	-0.255	-0.224	-0.179	-0.362	-0.224
	(0.274)	(0.276)	(0.312)	(0.351)	(0.290)	(0.290)	(0.330)	(0.363)
Territory	0.278	0.242	0.034	0.123	0.475*	0.452*	0.179	0.291
	(0.170)	(0.171)	(0.180)	(0.196)	(0.189)	(0.189)	(0.194)	(0.215)
Government/regime	0.037	-0.006	-0.376	0.142	0.606	0.575	-0.008	0.618
	(0.277)	(0.275)	(0.353)	(0.311)	(0.372)	(0.368)	(0.404)	(0.397)
Policy	-1.305***	-1.291***	-1.450***	-1.344***	-0.973***	-0.930***	-1.210***	-1.109***
	(0.150)	(0.150)	(0.159)	(0.175)	(0.174)	(0.177)	(0.171)	(0.187)
Other	-1.346***	-1.350***	-1.580***	-1.251***	-1.082**	-1.054**	-1.313***	-0.937*
	(0.328)	(0.333)	(0.357)	(0.371)	(0.346)	(0.359)	(0.365)	(0.397)
Constant	0.189	-0.358	-0.332	0.460	0.162	-0.046	-0.392	0.333
	(0.264)	(0.306)	(0.326)	(0.351)	(0.293)	(0.363)	(0.357)	(0.377)
<i>N</i>	1,582	1,582	1,276	1,222	1,361	1,361	1,086	1,064

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

were considered hegemon (1816–1939), and twenty-five disputes shortened. However, the changes were very similar in both types of cases, so we believe the sign change for the hegemon variable is due primarily to the changes made to the American dispute initiations, especially the revision of MID#3551 to be consistent with CoW MID coding rules.

These results demonstrate well the impact that changes in just a few cases can have on our inferences. We have also shown that how the data is treated matters greatly as well. Our final replication takes these points one step further by examining the studies that use dispute reciprocity as a dependent variable.

Replication 3: Reciprocation

During our research we consistently found the reciprocation variable to be poorly coded in the original data. This is important because several studies use reciprocation as a dependent variable to identify the potential for audience costs, which is one of the leading explanations for why democracies do not fight each other. In this final replication we revisit the Weeks (2008) study, which itself is an extension of prior work by Schultz (2001a). Both of these studies argue that (1) disputes initiated by democracies are less likely to be reciprocated because they can credibly signal their intentions. Weeks takes this argument a step further and claims that (2) certain forms of non-democratic governments can credibly signal as well, and their disputes are also unlikely to be reciprocated.

The dependent variable in Weeks (2008) is reciprocation by the target of a dispute in a directed-dyad sample, and the period analyzed includes the years 1946 to 1999. The base predictor is the presence of a challenger that has been a democracy for at least two years. This variable is then compared to different types of autocratic regimes, as defined by Geddes (2004). The control variables include dummy variables for various status differentials in the dyad (major-major, minor-major, and major-minor), the share of capabilities held by the initiator, contiguity, an alliance in the dyad, the similarity of alliance portfolios between the two states, and the status quo evaluations of the challenger and target. Dummy variables for the revision type of the dispute are also included. The overall sample, predictors, and estimation strategy closely mirror the research design developed by Schultz (2001a), with the exception of the dummy variables for the many types of non-democratic governments. Weeks (2008) argues that comparing the reciprocation results for each of these categories to the base category of democracy, which is statistically significant, can establish whether other forms of government also have audience costs. We replicated Weeks (2008) perfectly using her data, and above, in Table 7, we present the reanalysis using our suggested MID changes.

Our MID changes alter three important results. First, the presence of a democratic challenger no longer has an effect. If audience costs are identified well by reciprocation, then democracies would seemingly have no signaling advantage over other types of regimes. The standard error of the democratic challenger variable is actually twenty times greater than the coefficient. Similarly, personalist regimes and the governments labeled by Polity IV as non-democratic interregna both exhibit no differences when compared to the baseline of democracy in the second model. In fact, no regime type consistently predicts the likelihood of reciprocation by the target. Finally, contrary to the original study, the presence of a territorial issue is now statistically significant in both of the full-dispute models. Territorial disputes are more likely to be reciprocated by the target than other types of disputes, which is consistent with the numerous studies that establish the domestic salience of territorial issues.

We estimated additional models that constrained the original Weeks (2008) data to only those dispute cases in both datasets and found that our drop suggestions cause the change in the democratic initiator variable (the standard error becomes the same size as the coefficient). This finding echoes recent work by Gibler and Hutchison (2013) that found democratic audience costs in the reciprocation model were not consistent across dispute type. Clashes are more likely among less developed non-

democracies and are reciprocated by definition; meanwhile, seizures are often initiated by strong naval powers and are more often unreciprocated. Controlling for these two dispute types eliminated the effect of democracy in Gibler and Hutchison (2013, Table 1). Our research found that many cases originally coded as seizures were never protested. Clashes also tend to be information-poor cases, and our expanded research suggests a number of drops for these cases. Thus, either controlling for these cases or correctly applying MID coding rules produces the same effect, and democratic challengers have no reciprocation advantage.

While other regime-type variables also lost statistical significance due to the drop suggestions, personalist regimes remained more likely to be reciprocated in our smaller sample of disputes using original reciprocation coding. There are a total of thirteen cases of personalist challengers with reciprocation coding changes in our sample; we suggest changing eight cases of no reciprocation and five cases of reciprocation. However, there is no ready category that easily describes these cases. One dispute was originally coded incorrectly as an unreciprocated clash (MID#1382),¹² but we found no evidence of reciprocation in two other cases coded originally as clashes and change their highest action to shows of force. Overall, the thirteen cases seem to all be cases of incorrect original coding of information-poor cases. These are the cases that alter the inferences we drew from the original study.

Finally, for the territorial dispute variable, there are twenty three reciprocation changes in the sample. Eighteen no-reciprocation cases change, and five reciprocation cases actually had no response from the challenger. Most numerous among these changes are shows of force and threats that were actually part of larger conflicts that included responses from the target—this accounts for nine cases recoded as reciprocated. Two cases again had clashes that were coded as unreciprocated, but there were no other patterns across changes. There were also no patterns among the action categories that can describe the changes to no reciprocation.

Audience cost is an elusive concept to identify empirically (Schultz 2001b), and both Schultz (2001a) and Weeks (2008) were incredibly creative to use dispute reciprocation as a possible measure. However, our reanalysis suggests that there are no regime-based differences in dispute reciprocation, and prior findings may be based largely on poorly coded data. Reciprocation never seemed to be a primary variable for the MID project, and these results confirm that. Nevertheless, we want to still emphasize that our results do not necessarily invalidate the work using other measures of audience costs. Weeks (2014), for example, demonstrated in numerous ways that there may be differences across regime types in how domestic politics affects the ability of leaders to challenge and signal internationally. Instead, our results suggest that reciprocation is governed mostly by issue type and the types of challenges that are made.

Summary and Next Steps

We remain amazed that the original coders developed such an incredibly useful dataset of interstate conflict in

¹²As per Jones, Bremer, and Singer (1996, 173), clashes are “outbreak(s) of military hostilities between regular armed forces of two or more system members, in which the initiator may or may not be clearly identified.” That the initiator cannot be determined implies that both states are fighting, and the dispute is reciprocated.

an era without access to Internet searches, electronic newspaper databases, or powerful personal computers. The original coders systematically captured over 2,000 instances of militarized conflict since 1816, most often providing enough detailed and extensive information for us to deconstruct their coding decisions. Nevertheless, this replication process revealed several noteworthy issues with the dataset.

Our replication attempts compel us to recommend coding changes to a large majority of the dispute cases in the dataset. Approximately a fifth of these require major changes to the data. These changes may have substantial effects on existing studies, as we have documented. The predictors of interstate conflict seem to be unchanged after introducing our many revisions. This is due to the structure of conflict that is largely determined by the number of observations; CoW also used great care when identifying its primary variable—the existence of some type of interstate conflict. What does change, however, is our understanding of what goes on once a dispute begins. We found substantial changes in studies that examine dispute escalation, dispute duration, and dispute reciprocity; in the last case, our revisions actually eliminate the statistical significance of the primary variables of interest—regime differences in the likelihood of dispute challenges being reciprocated.

For the sake of transparency, we summarize all recommended changes and disaggregate the most severe changes concerning conflict intensity and dispute-ending measures in Appendix IID to this article. We provide detailed narratives for all cases we could not find (Appendix IIA), cases we recommend to drop (Appendix IIB), and cases for which new information substantiated a merge of two or more disputes (Appendix IIC). We also provide, in Appendix I, a discussion of several issues we identified when replicating and coding the original dataset. All of these are available as part of this manuscript, and we also provide a new dataset that incorporates all of our suggested changes.

We also provide, with release of this work, several new variables for interested users to more easily account for potentially problematic groups of disputes in the data and to gain more utility from the dataset. We include whether the dispute involved a seizure, had a six-month count to end the dispute, or involved the Tanker War. Our data also disaggregates each dispute according to which states were actively fighting each other. Finally, we have compiled a comprehensive bibliography for the dataset, with entries for each dispute case (Appendix IIF). We provide this data with the hope that scholars will carefully use the dataset and consider the different types of cases that meet the definition of militarized incidents according to CoW coding rules.

Cumulatively, our work suggests that we know well the structural factors that predict whether conflict occurs. The original coders of the MID dataset did an excellent job of identifying the bulk of militarized disputes between pairs of states since 1815, and the very nature of large-*N* conflict studies demonstrates that structural patterns of conflict are entrenched among dyads that are contiguous, non-democratic, and the like. Nevertheless we also found that what develops after onset is not well identified in the data, and any examination of variables beyond the existence of dyadic conflict should be closely examined. We urge caution in these cases and, at the very least, robustness checks using our data. We believe our review will, ultimately, strengthen the validity of using MID data to examine the causes, evolution, and conflict between states in the international system.

Supplemental Information

Supplementary information can be found at ISA Online.

References

- BRAITHWAITE, ALEX, AND DOUGLAS LEMKE. 2011. "Unpacking Escalation." *Conflict Management and Peace Science* 28 (2): 111–23.
- BREMER, STUART A. 1992. "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816–1965." *Journal of Conflict Resolution* 36 (2): 309–41.
- CARTER, DAVID B., AND CURTIS S. SIGNORINO. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18 (3): 271–92.
- DUPUY, RICHARD ERNEST, AND TREVOR NEVITT DUPUY. 1986. *The Encyclopedia of Military History: from 3500 B.C. to the Present*. London: Jane's.
- GEDDES, BARBARA, WRIGHT JOSEPH, AND FRANTZ, ERICA "Autocratic breakdown and regime transitions: A new data set." *Perspectives on Politics* 12 (02) (2014): 313–331.
- GEDDES, BARBARA. 2004. "Authoritarian Breakdown." Unpublished Manuscript. Department of Political Science, UCLA.
- GHOSH, FATEEN, GLENN PALMER, AND STUART BREMER. 2004. "The MID3 Data Set, 1993–2001: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 21 (2): 133–54.
- GIBLER, DOUGLAS M., AND MARC L. HUTCHISON. 2013. "Territorial Issues, Audience Costs, and the Democratic Peace: The Importance of Issue Saliency." *Journal of Politics* 75 (4): 879–93.
- GIBLER, DOUGLAS M., AND STEVEN V. MILLER. 2013. "Quick Victories? Territory, Democracies, and Their Disputes." *Journal of Conflict Resolution* 57 (2): 258–84.
- GIBLER, DOUGLAS M., AND MEREDITH REID SARKEES. 2004. "Measuring Alliances: The Correlates of War Formal Interstate Alliance Dataset, 1816–2000." *Journal of Peace Research* 41 (2): 211–22.
- GOCHMAN, CHARLES S., AND ZEEV MAOZ. 1984. "Militarized Interstate Disputes, 1816–1976 Procedures, Patterns, and Insights." *Journal of Conflict Resolution* 28 (4): 585–616.
- GOWA, JOANNE. 1999. *Ballots and Bullets: The Elusive Democratic Peace*. Princeton, NJ: Princeton University Press.
- HUTH, PAUL K., AND TODD L. ALLEE. 2002. *The Democratic Peace and Territorial Conflict in the Twentieth Century*. New York: Cambridge University Press.
- JONES, DANIEL M., STUART A. BREMER, AND J. DAVID SINGER. 1996. "Militarized Interstate Disputes, 1816–1992: Rationale, Coding Rules, and Empirical Patterns." *Conflict Management and Peace Science* 15 (2): 163–213.
- LANGER, WILLIAM. 1972. *An Encyclopedia of World History; Ancient, Medieval, and Modern, Chronologically Arranged*. Boston: Houghton, Mifflin.
- MARSHALL, MONTY G., AND KEITH JAGGERS. 2002. "Polity IV Project: Political Regime Characteristics and Transitions, 1800–1999." University of Maryland, Center for International Development and Conflict Management, College Park, MD.
- PALMER, GLENN, VITO D'ORAZIO, MICHAEL KENWIC, AND MATTHEW LANE. 2015. "The MID4 Data Set, 2002–2010: Procedures, Coding Rules and Description." *Conflict Management and Peace Science* 32 (2): 222–42.
- REED, WILLIAM. 2000. "A Unified Statistical Model of Conflict Onset and Escalation." *American Journal of Political Science* 44 (1): 84–93.
- REITER, DAN, AND ALLAN C. STAM. 2002. *Democracies at War*. Princeton, NJ: Princeton University Press.
- SCHULTZ, KENNETH A. 2001a. *Democracy and Coercive Diplomacy*. Cambridge: Cambridge University Press.
- . 2001b. "Looking for Audience Costs." *Journal of Conflict Resolution* 45 (1): 32–60.
- SIGNORINO, CURTIS S., AND JEFFREY M. RITTER. 1999. "Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions." *International Studies Quarterly* 43 (1): 115–44.
- SINGER, J. DAVID, STUART A. BREMER, AND JOHN STUCKEY. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820–1965." In *Peace, War and Numbers*, edited by Bruce Russett, 19–48. Beverly Hills, CA: Sage Publications.
- STINNETT, DOUGLAS M., JAROSLAV TIR, PHILIP SCHAFER, PAUL F. DIEHL, AND CHARLES GOCHMAN. 2002. "The Correlates of War Project Direct

- Contiguity Data, Version 3." *Conflict Management and Peace Science* 19 (2): 59–67.
- THOMPSON, WILLIAM R. 2001. "Identifying Rivals and Rivalries in World Politics." *International Studies Quarterly* 45 (4): 557–86.
- WEEKS, JESSICA L. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62 (1): 35–64.
- WEEKS, JESSICA L. P. 2014. *Dictators at War and Peace*. Ithaca, NY: Cornell University Press.
- WERNER, SUZANNE. 1999. "The Precarious Nature of Peace: Resolving the Issues, Enforcing the Settlement, and Renegotiating the Terms." *American Journal of Political Science* 43 (3): 912–34.