Hot under the collar: A latent measure of interstate hostility

Zhanna Terechshenko [ORCID]

Center for Social Media and Politics (CSMaP), New York University

Abstract

The majority of studies on international conflict escalation use a variety of measures of hostility including the use of force, reciprocity, and the number of fatalities. The use of different measures, however, leads to different empirical results and creates difficulties when testing existing theories of interstate conflict. Furthermore, hostility measures currently used in the conflict literature are ill suited to the task of identifying consistent predictors of international conflict escalation. This article presents a new dyadic latent measure of interstate hostility, created using a Bayesian item-response theory model and conflict data from the Militarized Interstate Dispute (MID) and Phoenix political event datasets. This model (1) provides a more granular, conceptually precise, and validated measure of hostility, which incorporates the uncertainty inherent in the latent variable; and (2) solves the problem of temporal variation in event data using a varying-intercept structure and human-coded data as a benchmark against which biases in machine-coded data are corrected. In addition, this measurement model allows for the systematic evaluation of how existing measures relate to the construct of hostility. The presented model will therefore enhance the ability of researchers to understand factors affecting conflict dynamics, including escalation and de-escalation processes.

Keywords

international conflict escalation, interstate hostility, measurement, latent variable model

Introduction

Despite the existence of a relatively large body of theoretical and empirical research on conflict escalation, there is still no consensus among international relations scholars on why some interstate disputes lead to war while others do not. Part of the explanation for this issue is that the measures currently used in the quantitative interstate conflict literature are ill suited to the task of identifying consistent predictors of conflict escalation. In this article, I provide a theoretically motivated measurement model that enhances researchers' ability to explain conflict processes. In particular, I define interstate conflict escalation as an increase in the level of hostility between countries involved in a militarized conflict. This approach requires a granular and validated measure of hostility – a variable we cannot observe directly, but manifestations of which we can. In this article, I create a latent measure of interstate hostility by constructing a Bayesian ordinal item-response theory model using conflict events data, including the Dyadic Militarized Interstate Disputes

(MID) (Maoz et al., 2019; Palmer et al., 2015) and Phoenix political event datasets (Althaus et al., 2017).

This project makes several contributions to the literature. First, it introduces a more precise and granular measure of hostility as the model combines the accuracy of the expert-coded MIDs and the granularity of the Phoenix event datasets. By capturing the underlying tension in the relationship between states, this novel measure can help to answer theoretical questions related to conflict dynamic processes including both escalation and de-escalation of interstate conflcts. In addition, the model employed in this article allows for the systematic evaluation of how existing measures relate to the construct of hostility. By integrating some of these measures, the model enables one to make inferences about their quality. In addition, my new measure of hostility incorporates

Corresponding author:
zt10@nyu.edu

uncertainty inherent in the latent variable, which has been largely ignored by existing measures of conflict.

Second, the presented model addresses the problem of reporting bias in the machine-coded event data through the use of a time-varying intercept structure and the use of human-coded data as a benchmark against which biases in machine-coded data are corrected. The proposed solution can be applied to almost any measurement problem that involves the use of human- and machine-coded event datasets.

Conflict escalation: The need for a new measure

The concept of escalation is central to numerous theories of conflict including explanations for the democratic peace (Maoz & Abdolali, 1989; Senese, 1997; Dixon & Senese, 2002), deterrence theory (Huth & Russett, 1988; Geller, 1990; Brams & Kilgour, 1987), bargaining theory (Fearon, 1994; Schelling, 1960; Snyder & Diesing, 1977), the steps-to-war explanation (Senese & Vasquez, 2008; Vasquez, 1987; Vasquez & Henehan, 2001; Gibler, 1997), and power transition theory (Organski & Kugler, 1981; Lemke & Reed, 1996). Despite its theoretical centrality, relatively few studies have endeavored to empirically measure or analyze escalation, choosing instead to focus on the onset of militarized hostilities. As a result, little is known about what distinguishes the conflicts that escalate to wars from those that stay at the same level of hostility or de-escalate.

One reason for this is a lack of consensus among international conflict scholars on how to properly measure conflict escalation. The consequences of this disagreement were illustrated in a study conducted by Braithwaite & Lemke (2011), who compare a variety of measures in their test of five correlates associated with conflict escalation: regime type, issue at stake, satisfaction with status quo, power preponderance, and joint alliance membership. They use six measures of conflict escalation, which are based on the information from the Militarized Interstate Dispute (MID) dataset and include: reciprocation of hostilities; use of force; mutual use of force; and the number of fatalities with thresholds at 0, 250, and 1,000 battle-deaths. They found that among the five possible causes of escalation, only territory has been a consistent predictor of the dependent variable. Their findings underscore the absence of a conceptually precise and validated measure of conflict escalation as a significant barrier to our understanding of escalation processes.

In order to address this issue, I introduce a more granular measure of hostility that can be used as a foundation for more precise analyses of escalation processes, which can be tracked by identifying changes in the level of hostility over time. The idea that escalation is an increase in the level of hostility is not new in the conflict literature. In fact, numerous studies operationalize escalation using the level of hostility in a dispute (Palmer, London & Regan, 2004; Bueno de Mesquita & Lalman, 2008; Schultz, 2001). The majority of existing measures are based on states' behavior once militarized conflict has already started (and often when the use of force has already occured) thus ignoring states' interaction before militarization of the dispute. As a result, we are not aware of the processes that lead from a non-militarized threat or accusation to armed conflict. This can be important, for example, in the case of rivalry, when states perceive each other as enemies and the source of a threat that is likely to be militarized in the future (Thompson, 2001). Hostile interactions short of the use of force can be useful in identifying the reasons why some militarized disputes lead to war, but also why some conflicts become militarized while others stay at the same level or get resolved. Furthermore, current measures of hostility lack the granularity necessary for analyzing precise changes in conflict dynamics. The reason for this is the source of the data. As the current hostility measures are constructed using highly aggregated human-coded datasets, they miss the information, which is likely to be captured in machine-coded data. Machine-coded datasets, however, tend to be noisy and to have temporal and spatial biases. The model presented here provides a solution for this trade-off between granularity and bias by combining the two types of data using human-coded data as a benchmark against which biases in machine-coded data are corrected.

Current measures of hostility

The most widely used measure of hostility from the COW Militarized Interstate Disputes (MID) dataset (Palmer et al., 2015) contains a five-level scale of hostility, starting from the absence of militarized action to full-scale war (Table I). One of the major concerns related to this scale, however, is a lack of theoretical and empirical evidence for its ordinal character (Diehl & Goertz, 2001; Maoz, 1982). For example, does a blockade necessarily indicate more hostile behavior than the explicit threat to use force? Is seizure a more hostile act than the mobilization of troops? Furthermore, in this five-level scale, a blockade and occupation of territory are both coded in the same category (use of force with a hostility level equal to 4). It means that they are treated

Table I. MID hostility scale

Hostility level	Action
1 = No militarized action	No militarized action
2 = Threat to use force	Threat to use force
	Threat to blockade
	Threat to occupy territory
	Threat to declare war
	Threat to use CBR weapons
	Threat to join war
3 = Display of force	Show of force
	Alert
	Nuclear alert
	Mobilization
	Fortify border
	Border violation
4 = Use of force	Blockade
	Occupation of territory
	Seizure
	Attack
	Clash
	Declaration of war
	Use of CBR weapons
5 = War	Begin interstate war
	Join interstate war

by researchers as equally informative, which might not be the case in the real world.

Maoz (1982) converted the COW ordinal hostility scale into a 14-category scale interval measure of dispute severity, with the threat to blockade as the least severe and war as the most severe actions. In their attempt to construct more precise indicators of dispute severity that would capture low-level conflicts, Diehl & Goertz (2001) came up with their own interval-level 200-point measure of dispute severity, based on the level of hostility and number of fatalities provided by the MID (Jones, Bremer & Singer, 1996) and COW datasets (Sarkees & Wayman, 2010). The distinct feature of this measure is the fact that wars and non-war MIDs are scaled together, thus providing a range of levels of severity even among wars (Diehl & Goertz, 2001). Finally, the concept of crisis severity is also embedded in the ICB dataset, in which crises are described using seven dimensions: (1) source or trigger mechanism, (2) gravity, (3) complexity, (4) intensity, (5) duration, (6) communication pattern, and (7) outcome (Brecher, 1977).

Diehl & Goertz (2001) acknowledge, however, that the existing measures of hostility/dispute severity are often crude and do not allow researchers to make any inferences about relatively small changes in conflict dynamics. The example they discuss is rivalry behavior. As the militarized dispute data provide information about rivalry only at the time of armed conflict, the use of finer-grained data can provide a more precise picture of rivalry (Diehl & Goertz, 2001: 265).

A new measure of hostility

In order to introduce my measure of hostility, I deploy Goertz's (2006) framework, according to which a concept can be described at theoretical, ontological, and operationalization levels. At the theoretical level, I define hostility as the enmity directed from one state to another and the level of hostility as the intensity of this enmity. For example, a high level of hostility is likely to be reflected in the relatively high frequency and/or intensity of aggressive interactions between states, while low levels of hostility might be reflected in the absence of any conflicts.

At the ontological level, I decompose the concept into its constitutive elements. Given that hostility represents the nature of states' relationships, it can be observed only through the expressions of this relationship. Therefore, the elements are hostile rhetoric, such as threats to use force, or hostile behavior (e.g. shows of force or attacks). Finally, at the operationalization level or, in other words, data/indicator level, I operationalize hostility as a latent trait manifested through conflict events.

My conceptualization of hostility is thus different from the most widely used measure adopted by the Correlates of War (COW) project. In its common use, scholars deploy the MID data to identify threats, displays, and uses of military force, assuming that each corresponds to a greater degree of hostility. An implicit assumption of the COW hostility scale is that certain types of actions are inherently more hostile, or more reflective of greater degrees of enmity than others. However, while higher levels of hostility are more likely to be reflected in the use of force, such as an attack or the beginning of war, the relationship between the actual level of hostility and a conflictual action might be context dependent. While the assumption that material actions are more hostile than verbal ones may generally be accurate, there are also contexts in which this is unlikely to be the case – a threat to deploy chemical, biological, or radiological weapons is likely to reflect greater enmity than a display of military force through the deployment of patrol boats in disputed waters, the current MID hostility scale would rank the latter action more highly than it would the former. I make no assumptions about the ranking of conflict actions in developing my measure. More importantly, however, I believe there is variation in the level of hostility among

the actions that involve the use of force and this variation is omitted in the COW's hostility measure.

To summarize, the assumptions I make in developing my measure of hostility are the following. First, as hostility is a part of an underlying relationship between states, it is a continuous trait, which we can observe only at certain points of time through 'hostile' actions, such as a blockade or an attack on one state by another. Second, while hostility is directed from one state to another, it does not need to be reciprocated. In this sense it is dyadic without necessarily being symmetric. In order to capture this asymmetry, I focus on directed dyads. In other words, I allow the level of hostility of the United States to Russia to be different from the level of hostility of Russia to the United States. Third, I assume that hostility is a unidimensional trait, which means it has a single underlying dimension and can be captured using a single measure. For instance, at any given time a dyad can be either more hostile or less hostile. Fourth, hostility is a latent trait, meaning that we cannot observe or measure it directly but we can infer its level based on its manifestations – conflictual behavior. While the idea that disputes are manifestations of states' hostility is not new in the conflict literature (Klein, Goertz & Diehl, 2006; Zinnes & Muncaster, 1984), this project is among the first that directly incorporates this theoretical assumption into an empirical model. Finally, as hostility is a latent trait, there is uncertainty associated with it. The model of hostility presented here allows researchers to incorporate this uncertainty into their statistical analysis. The model proposed in this article does not look at the effect of hostility on the probability of conflict but rather focuses on the information which states' conflict behavior or lack thereof provides. The goal of the model is to use states' actions, including those taken in the context of a conflict, in order to infer the level of hostility within a dyad at different points of time and to facilitate the analysis of conflict dynamics.

Machine-coded data and temporal bias

A potential solution to the problem outlined in the previous section is the use of machine-coded political event datasets. One of the most prominent examples is the Phoenix dataset (Althaus et al., 2017), which provides information on states' cooperative and conflict behavior, including conflicts short of the threshold for militarized interstate disputes. However, using this dataset has several possible limitations. First, machine-coded data are often noisy. For instance, D'Orazio et al. (2016) illustrate that the ability of
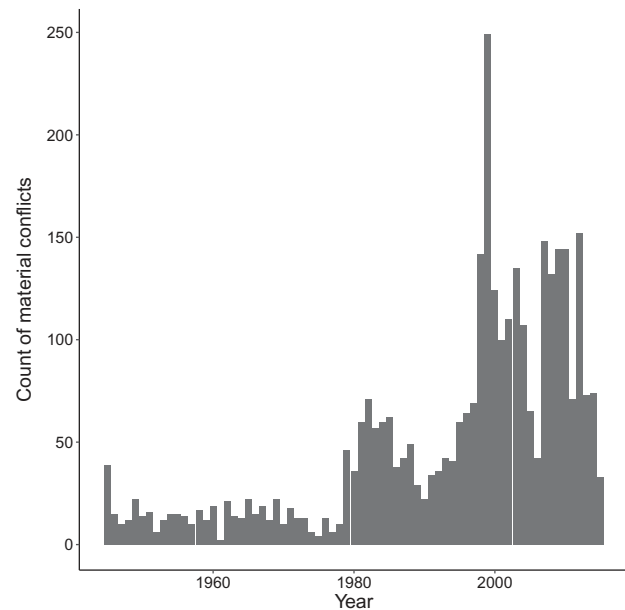

Figure 1. The number of material conflicts as reported in Phoenix dataset

machine coders to reproduce human-coded militarized interstate incidents data is relatively low. Second, machine-coded datasets are likely to suffer from temporal bias caused by the increase in the number of news stories over time. While this issue of volume increase has been illustrated in the analysis of the GDELT data (Ward et al., 2013), even datasets that rely on a limited number of sources, such as Phoenix, are subject to this bias.[1] This increase in the number of stories, if ignored, can translate into systematic measurement error and lead to incorrect inferences. For instance, Figure 1 shows that there has been an increase in the number of interstate material conflict events over time, which contradicts Pinker's (2012) argument on the global decline of interstate conflicts and is likely to be explained by an increase in the number of news stories.

In this article I present a solution (described in detail in the next section) to the temporal bias problem through the use of a varying-intercept model structure with human-coded data as a benchmark against which machine-coded data are integrated. In particular, in order to account for reporting bias, I allow the baseline probability of observing a material conflict event (variable from machine-coded Phoenix data) to vary as a function of time and combine it with the variables from

---

[1] In the Online appendix B, I provide empirical evidence for the presence of temporal bias in the Phoenix dataset.

Table II. Data variables and description

Data source	Variable name	Description
Militarized Interstate Disputes (1816–2010)	Show of force	Public demonstration by a state of its military forces, not involving combat operations
	Alert	Reported increase in the military readiness of a state's regular armed forces
	Nuclear alert	Reported increase in the military readiness of a state's nuclear forces
	Mobilization	Activation by a state of all or part of its previously inactive forces
	Border fortification	Explicit attempt to demonstrate control over a border area through the construction/reinforcement of military outposts
	Border violation	Crossing of a recognized land, sea or air boundary for a period of less than 24 hours by official forces of one state, without any use of force being used
	Blockade	Use of ships, planes or troops by one state to seal off the territory of another state so as to prevent entry or exit of goods or personnel
	Occupation of territory	Use of military force by one state to occupy the whole or part of another state's territory for a period of more than 24 hours
	Seizure	Capture of material/official forces from another state, or the detention of private citizens operating within contested territory. Lasts at least 24 hours
	Attack	Use of regular armed forces of a state to fire upon the armed forces/population/territory of another state
	Clash	Outbreak of military hostilities between regular armed forces of two or more system members, in which the initiator may or may not be clearly identified
	Beginning of interstate war	Beginning of interstate war between two or more states as defined by the COW project
	Joining interstate war	Joining of an ongoing COW interstate war by a state in opposition to another state that is currently involved in that war
Phoenix (1946–2010)	Material conflict	Physical acts of a conflictual nature, including armed attacks, destruction of property, assassination, etc.

the MID dataset.[2] This framework allows me to combine the accuracy of human-coded data and granularity of the machine-coded event dataset.

Model

My model assumes that hostility is a unidimensional trait that can be measured using observed outcomes. I employ an item-response theory (IRT) model, which is a type of latent variable model used to generate estimates of a latent trait of interest (hostility) by combining information from observable items or manifest variables (conflict events). It has been used increasingly in political science, enabling researchers to estimate a number of unobservable concepts (Jackman, 2009; Reuning, Kenwick & Fariss, 2019). Furthermore, this model allows me to directly model the uncertainty associated with the latent trait.

Data

I use the directed dyad-quarter as the unit of analysis. As one of the main goals of this project is to create a more

granular measure of hostility that can be used as a measure of conflict escalation and thus is able to capture even the smallest changes in states' interactions, I focus on the annual quarter instead of more traditional year-level unit of analysis. My sample consists of politically relevant dyad-quarters from 1950 to 2010. Politically relevant dyads are dyads where either states are contiguous or at least one of the states is a major power. I reduce my sample to politically relevant dyads both for computational tractability and following common practice in the international conflict literature which has shown that this sample is generalizable to the population of dyads (Lemke & Reed, 2001). As a result, I have 648,368 directed dyad-quarter observations and 3,388 directed dyads.

In order to construct the latent measure of hostility I merge variables from two conflict event datasets: the COW Dyadic Militarized Interstate Disputes (MID) dataset (Maoz et al., 2019; Palmer et al., 2015) and the Phoenix event dataset (Althaus et al., 2017) (summarized in Table II). For the Phoenix project, I combine data from three sources: New York Times Corpora events from 1945 to 2005, FBIS Corpora events from 1945 to 2005, BBC Summary of World Broadcast Corpora events from 1979 to 2015 in a way such that, if any of

---

[2] A similar framework has been used by Fariss (2014).

Table III. Combination of fatalities and variables indicating the use of force

	Attack	Clash
Level 0	No attack	No clash
Level 1	Attack, no fatalities	Clash, no fatalities
Level 2	Attack, data on fatalities are missing	Clash, data on fatalities are missing
Level 3	Attack with fatalities	Clash with fatalities

these sources capture an event, it is recorded in my dataset. The choice of the data is driven primarily by the prominence of the datasets and their temporal coverage. The MID dataset is the most commonly used human-coded conflict event dataset in the international conflict literature. While the Phoenix dataset is relatively new, with all its datasets combined, it allows me to capture conflict events in the post-World War II period.

The variables from the MID dataset include: *Show of force, Alert, Nuclear alert, Mobilization, Border fortication, Border violation, Blockade, Occupation of territory, Seizure, Attack, Clash, Beginning of war*, and *Joining war*. All of these variables except *Attack* and *Clash* are dichotomous. *Attack* and *Clash* variables are ordinal. As attacks and clashes are more likely to involve fatalities than other conflict actions short of war, I incorporate the fatality level into these variables. The assumption I am making is that attacks and clashes with fatalities reflect higher levels of hostility than the same actions that do not lead to fatalities. As a result, 0 indicates the absence of attack/clash, 1 – attack/clash without fatalities, 2 – attack/clash with missing information on fatalities, and 3 – attack/clash with fatalities (Table III).

Material conflict is a quad category from the Phoenix dataset. Quad categories represent a high level of aggregation of the Conflict and Mediation Event Observations (CAMEO) framework. The *Material conflict* category includes exhibition of force posture, reduction in relations, coercion, assault, fight, and engagement in unconventional mass violence (Gerner et al., 2002). 0 indicates the absence of the material conflict, 1 indicates one material conflict, 2 corresponds to two material conflicts, and I assigned 3 to the variable if the number of events is equal to or exceeds three. The assumption here is that higher hostility is likely to be reflected in the higher number of material conflicts within the dyad.

Model

In order to construct a latent measure of hostility, I use an ordinal item-response (O-IRT) model. The goal of this model is to estimate hostility, where θ_{it} is the hostility estimate for the directed dyad i at time t using the observable manifestations of hostility outlined in the previous section, y_{itj}.

Each item is indexed $j = 1, \ldots, J$ and is observed at the dyad-quarter level, where dyads are indexed $i = 1, \ldots, N$ and time is indexed $t = 1, \ldots, T$. Each item has two parameters: the 'item discrimination' parameter β_j and the 'item difficulty' parameter α_j. β_j shows how informative the item is. α_j indicates the probability of the item at the average level of hostility (Jackman, 2009). K_j is the total number of values that each item y_j can take on. If item j is binary, $K_j = 2$ and it has only one intercept and, if item j is ordinal, $K_j > 2$ and I use $K_j – 1$ cutpoints.

To account for the temporal bias in the machine-coded dataset variable (*Material conflict*), I make the following modification to the model. I assume that the difficulty parameter for the human-coded MID variables is constant over time and leave it as α_j. For the *Material conflict* variable, however, I allow the difficulty parameter to vary as a function of time, which means that I estimate it for each year and parameterize it as α_{tj}.

Following Fariss's (2014) framework, I specify the probability distribution as:

$$P[y_{itj} = k] = [F(\alpha_{tjk} - \beta_j\theta_{it}) - F(\alpha_{tjk-1} - \beta_j\theta_{it})]^{(v_j)}$$
$$* [F(\alpha_{jk} - \beta_j\theta_{it}) - F(\alpha_{jk-1} - \beta_j\theta_{it})]^{(1-v_j)} \quad (1)$$

where $v_j = 1$ when the $_j$ indicator is one of the Phoenix variables and $v_j = 0$ when it is one of the MID variables.

As θ_{it} cannot be fully observed and all of the parameters of interest must be estimated simultaneously, the Bayesian approach has been increasingly used to estimate this type of model (Jackman, 2009). In Bayesian inference, the researcher starts with specification priors that reflect the uncertainty or belief about the parameters before the data have been observed. I set the prior for latent trait to $\theta_{it} \sim$ Normal(0, 1). This weakly informative prior reflects an assumption that the population of dyads is roughly normally distributed across the spectrum of hostility. I assign the slightly informative prior $\beta_j \sim$ Gamma(4, 3) that restricts the value of the item discrimination parameter to be positive and reflects an assumption that all indicators contribute (and in the same direction) to the latent variable. Under this model specification, increases in the values of each indicator y_j correspond to the higher values of the latent trait. Finally, I assign the very weak priors to α ($\alpha_j \sim$
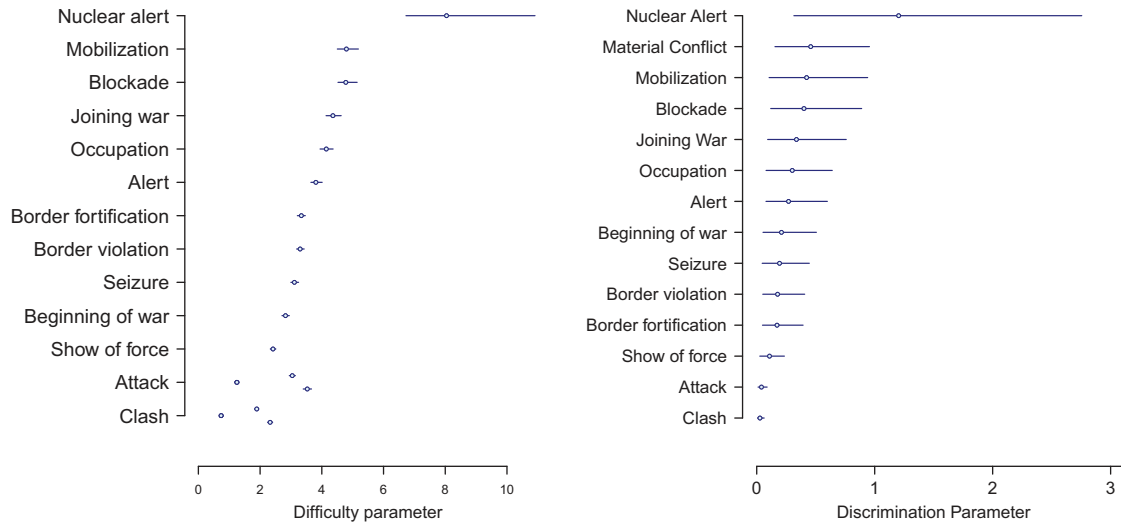
Figure 2. Model parameters

This figure displays the difficulty and discrimination parameters for each item. Dots indicate means of the cutpoints. Lines indicated 95% credible intervals. *Mobilization*, *Blockade*, and *Nuclear alert* are more likely to be in the higher category at the average level of hostility. The change in the level of hostility is reflected the most in *Material conflict*, *Beginning of war*, and *Nuclear alert* variables.

Normal(0, 10)). By assigning this prior I impose no information on the hostility measure and allow for wide variation in observations. In general, the priors I use in the model are either consistent with or less informative than those used in several other studies using dynamic IRT models (Fariss, 2014; Reuning, Kenwick & Fariss, 2019; Schnakenberg & Fariss, 2014; Treier & Jackman, 2008).[3]

In order to assess convergence, I look primarily to the R̂ statistic, which compares the variance over chains to that within chains for a given parameter (Gelman & Hill, 2007). All R̂ values are less than 1.1. To check for autocorrelation, I look at the effective sample size (ESS), which is sample size adjusted for autocorrelation. The majority of the parameters have ESS higher than 2,000, thus there is no problem with autocorrelation. Posterior predictive checks are reported in the Online appendix and suggest the model fits the data well.

Model output

My final hostility estimate has an average value of 0, with the highest estimate of 1.5 for the China–Taiwan dyad and the lowest estimate of –0.17 for the United Kingdom–Kyrgyzstan dyad. Given the fact that the China–Taiwan dyad has emerged out of conflict, while the

United Kingdom and Kyrgyzstan have few within-dyad conflict interactions, this quick examination of the estimates provides some evidence for face validity of my measure.

The contribution of each of the variables in the model is reflected in the difficulty and discrimination parameters (left panel of Figure 2). For a given hostility level, the probability of conflict event increases as the item difficulty decreases. At the average level of hostility, the probability of a nuclear alert is lower than the probability of any other conflict event, which can be explained by the fact that a nuclear alert is a relatively rare event and we would expect a nuclear alert only at a high level of hostility.

The discrimination parameter reflects the extent to which a change in the level of hostility corresponds to the change in each of the manifest variables. For instance, Figure 2, right panel, shows that if the level of hostility within a dyad increases, we are more likely to observe *Alert* than *Attack* or *Clash*. This is important, as it indicates that, for example, *Alert* is a more hostile action than *Seizure*. This observation, however, contradicts the MID hostility level scale. It should be noted that while the MID hostility scale is based on theoretical assumptions, the ranking presented here is generated by the conflict data. Thus, the model output suggests that the use of the ordinal scale for measuring hostility and conflict escalation might bias the results as the measure would not necessarily reflect reality.

---

[3] The model is estimated using Stan, a C++ program, which performs Bayesian inference using a No-U-Turn sampler (Hoffman & Gelman, 2014), an adaptive form of Hamiltonian Monte Carlo sampling (Neal et al., 2011).

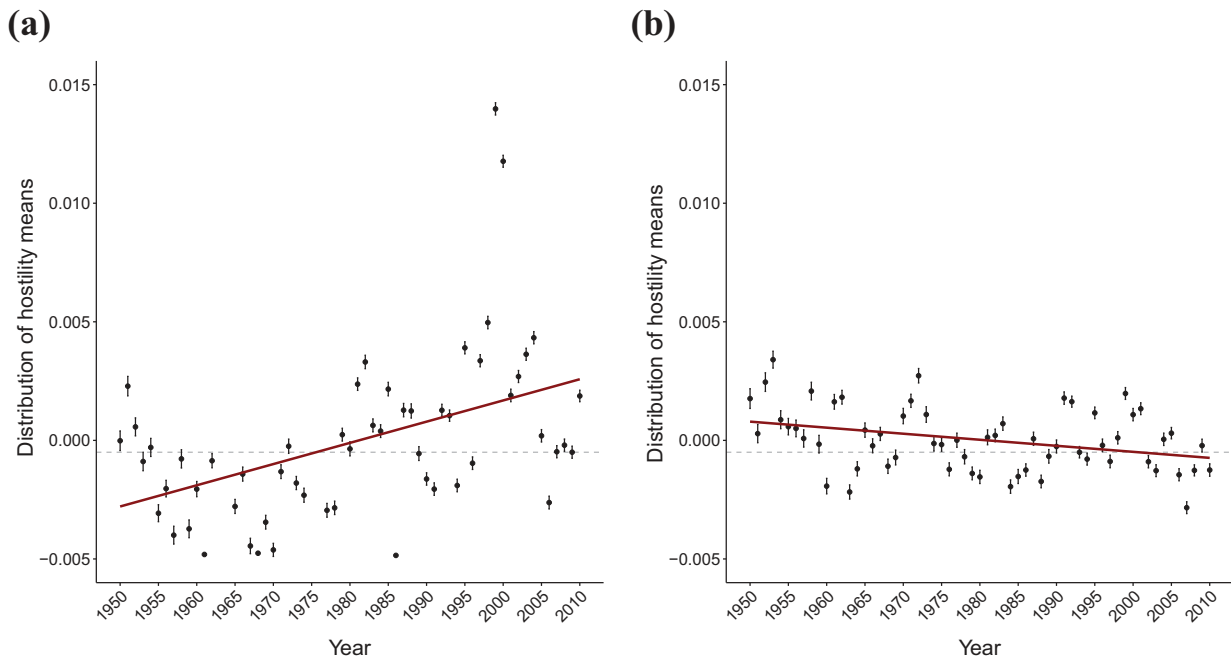**(a)**   **(b)**

Figure 3. Temporal variation in the level of global hostility

(a) Fixed-intercept (b) Varying-intercept

This figure shows temporal variation in the level of global hostility in models with the fixed intercept and time-varying intercept models. Dots indicate the mean posterior estimates across all politically relevant dyads over time. Bars represent 95% credible intervals. Solid lines show the linear relationship between years and the mean hostility values. Dashed lines indicate the mean values of hostility, 1950–2010.

Concurrent validity

As a concurrent validity check, I look at variation of the level of global hostility across time. As the number of interstate wars has been decreasing over time (Pinker, 2012), I would expect the level of global hostility to be decreasing as well. I take 1,000 draws from the posterior estimates of θ_{it} and calculate the annual means of hostility across these draws. I then take the mean values across all dyads in my sample for each year (Figure 3). According to Figure 3, the time-varying intercept model (b) shows the decrease in the level of hostility and, therefore, is more consistent with this theoretical prediction than the fixed-intercept model (a), which shows a substantial increase.

I also look at the levels of hostility in several dyads. For instance, according to Diehl & Goertz (2001), the China–South Korea rivalry is an example of an enduring rivalry with a pattern of decreasing severity over time. Therefore, we would expect hostility in this dyad to be decreasing over time as well. Figure 4 shows that varying-intercept model (b) is more consistent with our expectation than fixed-effect model (a). Thus, the varying-intercept model passes the concurrent validity check. The fixed-effect model also shows increasing levels of hostility in the

United States–Soviet Union/Russia (Figure 5) dyad, which provides additional evidence for the presence of the temporal bias in the model. In both cases, however, the varying-intercept model is aligned well with what we observe in the real world.

Predictive validity check

In this section, I assess the predictive validity of my hostility measure. If this measure is valid, it should be able to predict the processes to which it is theoretically linked (Trochim, 2006). In particular, I extend Findley, Piazza & Young's (2012) study on terrorism in international rivalries.

Findley, Piazza & Young (2012) argue that states hostile toward each other are more likely to abet transnational terrorists than other states. Sponsorship of terrorism gives the states involved tactical advantages such as plausible deniability and disproportionate effectiveness. Furthermore, the increasing costs of conventional warfare makes terrorism a viable option for states that try to avoid direct confrontation with their rivals (Jenkins, 1975; Conrad, 2011).

In order to test their theory, Findley, Piazza & Young (2012) use two measures of rivalry. The first measure is
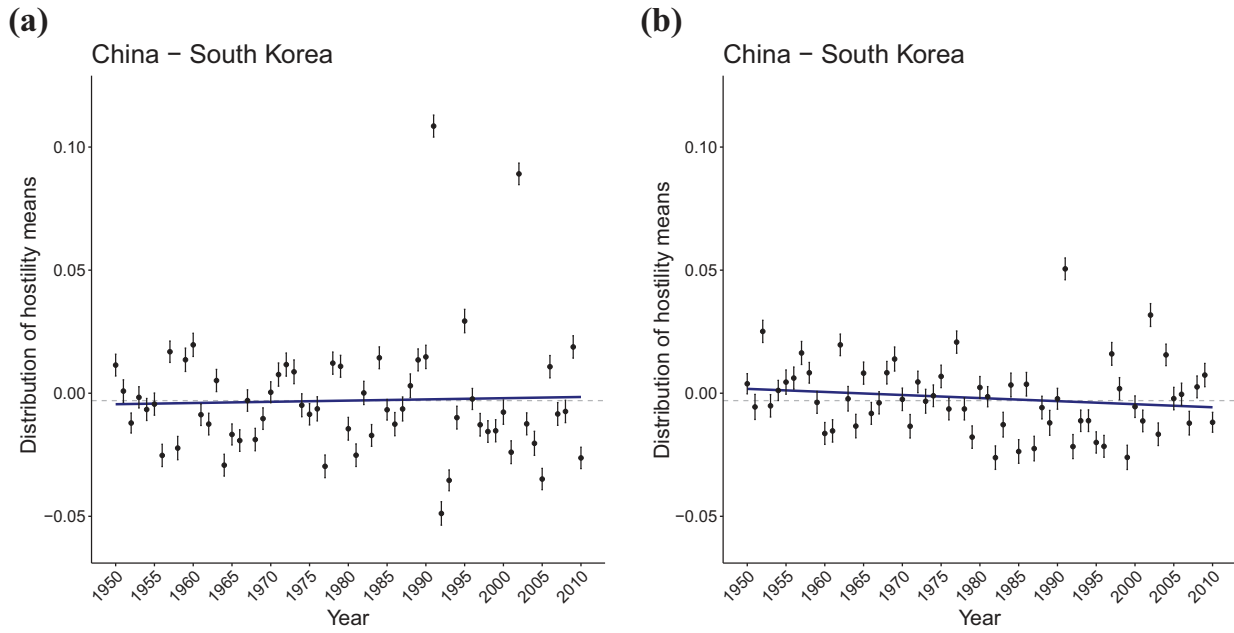
Figure 4. Temporal variation in the level of hostility in the China–South Korea dyad

(a) Fixed-intercept (b) Varying-intercept

This figure shows mean posterior estimates for the China–South Korea dyad over time in models with fixed intercept and time-varying intercept. Bars represent 95% credible intervals. Solid lines show the linear relationship between years and the mean hostility values. Dashed lines indicate the mean values of hostility within the dyad across time.
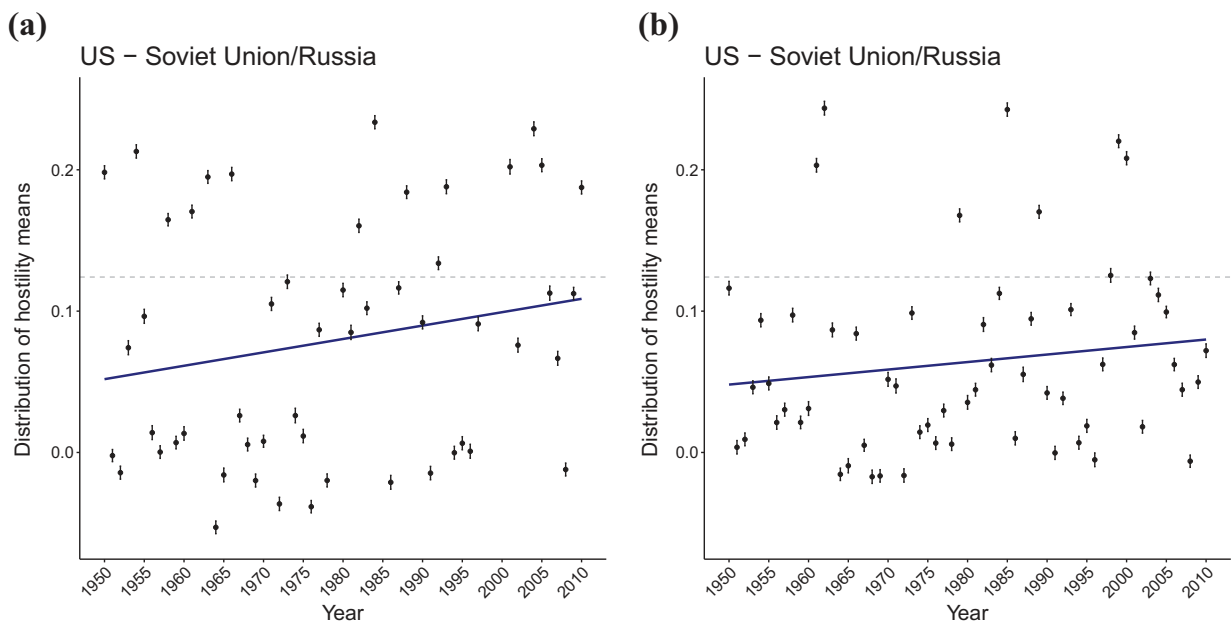

Figure 5. Temporal variation in the level of hostility in the United States–Soviet Union/Russia dyad

(a) Fixed-intercept (b) Varying-intercept

This figure shows mean posterior estimates for the United States–Soviet Union/Russia dyad over time in models with fixed intercept and time-varying intercept. Bars represent 95% credible intervals. Solid lines show the linear relationship between years and the mean hostility values. Dashed lines indicate the mean values of hostility within the dyad across time.

Table IV. Negative binomial models of transnational terrorist attacks using dyads 1968–2002 (without rivals)

	(1) Terror counts	(2) Terror counts 2	(3) Terror counts	(4) Terror counts 2
Hostility	7.080*	7.822**	5.288*	4.642**
	(2.471)	(1.588)	(1.854)	(1.413)
Joint democracy	1.223**	0.925**	0.271	–0.164
	(0.211)	(0.140)	(0.224)	(0.136)
Log (Capability ratio)	–0.055	–0.329**	0.121	–0.597**
	(0.053)	(0.027)	(0.108)	(0.055)
Contiguity	1.126**	0.083	1.841**	0.915**
	(0.221)	(0.167)	(0.208)	(0.173)
Past terror (Origin)			0.331**	0.724**
			(0.064)	(0.039)
Past terror (Target)			0.687**	0.669**
			(0.065)	(0.047)
Cold War			–0.900**	–0.361**
			(0.139)	(0.086)
Interstate war (Origin)			0.504†	–0.059
			(0.224)	(0.159)
Interstate war (Target)			–0.014	0.472**
			(0.223)	(0.105)
Civil war (Origin)			1.429**	1.041**
			(0.218)	(0.141)
Civil war (Target)			–0.237	–0.210
			(0.220)	(0.129)
Constant	–5.317**	–3.615**	–6.450**	–5.516**
	(0.172)	(0.114)	(0.224)	(0.166)
Number of observations	60,026	60,026	60,026	60,026

Standard errors in parentheses. $^{\dagger}p < 0.1$, $*p < .02$, $**p < 0.002$.

from Klein, Goertz & Diehl (2006), who identify a rivalry through issue linkages in repeated militarized conflicts between two states. The second is the operationalization used by Rasler & Thompson (2006: 151), where rivalries are defined as 'relationships among actors who mutually perceive their adversary to be a competitor'. The dependent variable is the number of transnational terrorist attacks based on the ITERATE dataset (Mickolous, Sandler & Murdock, 2007). The unit of analysis is the directed dyad-year. The sample is limited to politically relevant dyads from 1968 to 2002. Regardless of the operationalization of the independent variable, Findley, Piazza & Young (2012) find that transnational terrorist attacks are more likely to occur in the presence of interstate rivalry.

I extend Findley, Piazza & Young's (2012) study by looking at the effect of hostility on the number of terrorist attacks. Based on these theoretical and empirical studies, if my measure of interstate hostility is valid, the level of hostility should be able to reproduce a strong positive effect on the count of transnational terrorist attacks.

The model specification is very similar to the one used by Findley, Piazza & Young (2012). The unit of analysis is the directed dyad-year. The analysis includes only politically relevant directed dyads from 1968 to 2002. Following Findley, Piazza & Young (2012), I use two approaches in estimating the effect of hostility on the dependent variable. In both approaches, the origin country is defined as the nationality of the terrorists. However, in the first approach, the target country is defined as the country in which the terrorist event occurred, while in the second approach, the target country is the nationality of the victims. Finally, I preserve the set of control variables used in Findley, Piazza & Young's (2012) paper. The small set includes only dyadic variables: rivalry, hostility, joint democracy, contiguity, and capability ratio. In addition to these covariates, the fully specified model includes the history of terrorism, interstate war, Cold War, and civil war in both the origin and target states. First, I reproduce the analysis on the full sample and found that hostility in fact has a significant positive effect on the number of transnational attacks. The results are shown in Online appendix D. Second, I

designed a harder test for my measure by excluding rivals from the dataset. This test provides a challenge for my measure as it reduces the sample in size and limits it to the dyads that are not engaged in a rivalry and, thus, are likely to be on the lower end of hostility scale. Moreover, this test can illustrate the utility of my hostility measure over the dichotomous rivalry approaches, showing that states' interactions are likely to be more nuanced and thus require a fine-grained measure.

Table IV presents the results.[4] The coefficient for interstate hostility is positive and significant across all models, which provides support for my hypothesis, suggesting that dyads with higher levels of hostility are more likely to experience transnational terrorist attacks than other dyads. In addition to demonstrating the predictive validity of my measure, this analysis suggests that the theoretical expectations of Findley, Piazza & Young (2012) extend into important new grounds – I find that there is a link between hostility and transnational terrorism even in non-rivalrous dyads.

Conclusion

Current discussions of interstate conflict escalation are hindered by the lack of a single framework, which could be used for studying conflict dynamics. Understanding the process of conflict escalation, however, is important for numerous theories on international conflict. A significant part of the problem is the absence of a valid measure of conflict escalation. In this article, I address this issue by developing a new measure of interstate hostility that can serve as a foundation for measuring escalation. One of the major contributions of this project is a solution to the trade-off between granularity and bias in measuring hostility. For instance, it is well known that while human-coded datasets, such as MID, are often criticized for being too highly aggregated, events data struggle with irregular reporting. The measure presented in this article mitigates these issues through the use of a time-varying intercept model structure and combination of human-coded and machine-coded data. Furthermore, it incorporates uncertainty in the underlying measure which is hitherto ignored in virtually every measure of hostility to date. In addition, the presented model suggests the necessity for a re-evaluation of existing measures of dispute severity and hostility in terms of their relation toward the concept. Finally, the new measure of interstate hostility not only allows for testing of existing

theories, but also can motivate new theories on escalation and de-escalation processes as well as foreign policy in general. For example, the replication of Findley, Piazza & Young's (2012) study with the hostility measure suggests that the level of interstate hostility can potentially explain states' sponsorship of transnational terrorism. The fact that this finding holds even for non-rivalrious dyads suggests that Findley, Piazza & Young's (2012) argument can be extended to a larger set of dyads and illustrates the utility of the finer-grained measure of hostility.

This project constitutes a first step in the analysis of conflict escalation. Moving forward, incorporating more complex dependencies in the international system into the study of escalation seems to be a fruitful area of research. Potential lines of inquiry include the analysis of the impact of the conflict escalation process within one dyad on the conflict/cooperation in other dyads or the relationship between the alliance network structure and probability of the conflict escalation.

Replication data

The dataset, codebook, and do-files for the empirical analysis in this article, as well as the Online appendix, can be found at http://www.prio.org/jpr/datasets.

ORCID iD

Zhanna Terechshenko ⓘ https://orcid.org/0000-0001-7585-1369

References

Althaus, Scott; Joseph Bajjalieh, John F Carter, Buddy Peyton & Dan A Shalmon (2017) Cline Center Historical Phoenix Event Data. Distributed by Cline Center for Advanced Social Research.

Braithwaite, Alex & Douglas Lemke (2011) Unpacking escalation. *Conflict Management and Peace Science* 28(2): 111–123.

Brams, Steven J & D Marc Kilgour (1987) Threat escalation and crisis stability: A game-theoretic analysis. *American Political Science Review* 81(3): 833–850.

Brecher, Michael (1977) Toward a theory of international crisis behavior. *International Studies Quarterly* 21(1): 39–74.

---

[4] I used the posterior mean as a point estimate of interstate hostility in each dyad-year.

Bueno de Mesquita, Bruce & David Lalman (2008) *War and Reason: Domestic and International Imperatives*. New Haven, CT: Yale University Press.

Conrad, Justin (2011) Interstate rivalry and terrorism: An unprobed link. *Journal of Conflict Resolution* 55(4): 529–555.

Diehl, Paul F & Gary Goertz (2001) *War and Peace in International Rivalry*. Ann Arbor, MI: University of Michigan Press.

Dixon, William J & Paul D Senese (2002) Democracy, disputes, and negotiated settlements. *Journal of Conflict Resolution* 46(4): 547–571.

D'Orazio, Vito; Michael Kenwick, Matthew Lane, Glenn Palmer & David Reitter (2016) Crowdsourcing the measurement of interstate conflict. *PloS One* 11(6): e0156527.

Fariss, Christopher J (2014) Respect for human rights has improved over time: Modeling the changing standard of accountability. *American Political Science Review* 108(2): 297–318.

Fearon, James D (1994) Domestic political audiences and the escalation of international disputes. *American Political Science Review* 88(3): 577–592.

Findley, Michael G; James A Piazza & Joseph K Young (2012) Games rivals play: Terrorism in international rivalries. *Journal of Politics* 74(1): 235–248.

Geller, Daniel S (1990) Nuclear weapons, deterrence, and crisis escalation. *Journal of Conflict Resolution* 34(2): 291–310.

Gelman, Andrew & Jennifer Hill (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*, volume 1. New York: Cambridge University Press.

Gerner, Deborah J; Philip A Schrodt, Omür Yilmaz & Rajaa Abu-Jabr (2002) Conflict and Mediation Event Observations (CAMEO): A new event data framework for the analysis of foreign policy interactions. International Studies Association, New Orleans.

Gibler, Douglas M (1997) Control the issues, control the conflict: The effects of alliances that settle territorial issues on interstate rivalries. *International Interactions* 22(4): 341–368.

Goertz, Gary (2006) *Social Science Concepts: A User's Guide*. Princeton, NJ: Princeton University Press.

Hoffman, Matthew D & Andrew Gelman (2014) The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 15(1): 1593–1623.

Huth, Paul & Bruce Russett (1988) Deterrence failure and crisis escalation. *International Studies Quarterly* 32(1): 29–45.

Jackman, Simon (2009) *Bayesian Analysis for the Social Sciences*, volume 846. Hoboken, NJ: John Wiley & Sons.

Jenkins, Brian (1975) International terrorism: A balance sheet. *Survival* 17(4): 158–164.

Jones, Daniel M; Stuart A Bremer & J David Singer (1996) Militarized interstate disputes, 1816–1992: Rationale, coding rules, and empirical patterns. *Conflict Management and Peace Science* 15(2): 163–213.

Klein, James P; Gary Goertz & Paul F Diehl (2006) The new rivalry dataset: Procedures and patterns. *Journal of Peace Research* 43(3): 331–348.

Lemke, Douglas & William Reed (1996) Regime types and status quo evaluations: Power transition theory and the democratic peace. *International Interactions* 22(2): 143–164.

Lemke, Douglas & William Reed (2001) The relevance of politically relevant dyads. *Journal of Conflict Resolution* 45(1): 126–144.

Maoz, Zeev (1982) *Paths to Conflict: International Dispute Initiation, 1816–1976*. Boulder, CO: Westview.

Maoz, Zeev & Nasrin Abdolali (1989) Regime types and international conflict, 1816–1976. *Journal of Conflict Resolution* 33(1): 3–35.

Maoz, Zeev; Paul L Johnson, Jasper Kaplan, Fiona Ogunkoya & Aaron P Shreve (2019) The Dyadic Militarized Interstate Disputes (MIDs) dataset version 3.0: Logic, characteristics, and comparisons to alternative datasets. *Journal of Conflict Resolution* 63(3): 811–835.

Mickolous, Edward F; Todd Sandler & Jean M Murdock (2007) International Terrorism: Attributes of Terrorist Events, 1968–2006 (ITERATE). Vinyard Software.

Neal, Radford M, et al. (2011) MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo* 2: 113–162.

Organski, Abramo F & Jacek Kugler (1981) *The War Ledger*. Chicago, IL: University of Chicago Press.

Palmer, Glenn; Vito d'Orazio Michael Kenwick & Matthew Lane (2015) The MID4 dataset, 2002–2010: Procedures, coding rules and description. *Conflict Management and Peace Science* 32(2): 222–242.

Palmer, Glenn; Tamar London & Patrick Regan (2004) What's stopping you? The sources of political constraints on international conflict behavior in parliamentary democracies. *International Interactions* 30(1): 1–24.

Pinker, Steven (2012) *The Better Angels of Our Nature*. New York: Penguin.

Rasler, Karen A & William R Thompson (2006) Contested territory, strategic rivalries, and conflict escalation. *International Studies Quarterly* 50(1): 145–168.

Reuning, Kevin; Michael R Kenwick & Christopher J Fariss (2019) Exploring the dynamics of latent variable models. *Political Analysis* 27(4): 503–517.

Sarkees, Meredith Reid & Frank Whelon Wayman (2010) *Resort to War (Correlates of War): A Data Guide to Inter-State, Extra-State, Intra-State, and Non-State Wars, 1816–2007*. Washington, DC: CQ.

Schelling, Thomas C (1960) *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Schnakenberg, Keith E & Christopher J Fariss (2014) Dynamic patterns of human rights practices. *Political Science Research and Methods* 2(1): 1–31.

Schultz, Kenneth A (2001) *Democracy and Coercive Diplomacy*, volume 76. Cambridge: Cambridge University Press.

Senese, Paul D (1997) Between dispute and war: The effect of joint democracy on interstate conflict escalation. *Journal of Politics* 59(1): 1–27.

Senese, Paul D & John A Vasquez (2008) *The Steps to War: An Empirical Study*. Princeton, NJ: Princeton University Press.

Snyder, Glenn & Paul Diesing (1977) *Conflict Among Nations: Bargaining, Decision Making, and System Structure in International Crises*. Princeton, NJ: Princeton University Press.

Thompson, William R (2001) Identifying rivals and rivalries in world politics. *International Studies Quarterly* 45(4): 557–586.

Treier, Shawn & Simon Jackman (2008) Democracy as a latent variable. *American Journal of Political Science* 52(1): 201–217.

Trochim, William M (2006) *The Research Methods Knowledge Base*. Cincinnati, OH: Atomic Dog.

Vasquez, John & Marie T Henehan (2001) Territorial disputes and the probability of war, 1816–1992. *Journal of Peace Research* 38(2): 123–138.

Vasquez, John A (1987) The steps to war: Toward a scientific explanation of Correlates of War findings. *World Politics* 40(1): 108–145.

Ward, Michael D; Andreas Beger, Josh Cutler, Matt Dickenson, Cassy Dorff & Ben Radford (2013) Comparing GDELT and ICEWS event data. *Analysis* 21(1): 267–297.

Zinnes, Dina A & Robert G Muncaster (1984) The dynamics of hostile activity and the prediction of war. *Journal of Conflict Resolution* 28(2): 187–229.

ZHANNA TERECHSHENKO, b. 1992, PhD in Political Science and Social Data Analytics (The Pennsylvania State University, 2019); Postdoctoral Fellow, New York University (2019–).