

# Identifying Threats: Using Machine Learning in International Relations

Azusa Katagiri and Eric Min\*

Department of Political Science  
Stanford University

Draft  
September 15, 2015

## Abstract

International relations (IR) scholars have always placed great emphasis on the importance of threats in explaining interstate behavior—especially in the midst of crises. However, few studies have evaluated threats in a systematic way, mainly due to lack of substantively and methodologically sound data. Our study uses machine learning techniques on digitized U.S. diplomatic documents from the *Foreign Relations of the United States 1945-1980* collection. We utilize a statistical learning model to classify documents during the Cold War as expressing perceived threat or not, and then analyze this using event data on the United States and Soviet Union. Time-series analysis shows that threats are perceived prior to actually threatening events, and reveal far greater detail on interstate reciprocity. We conclude by suggesting that IR’s current empirical approaches to threats may be problematic, and that our methods may lay the foundations for an ambitious research agenda that can test many long-standing theories in IR.

## 1 Introduction

The concept of threat has played a critical role in international relations (IR), particularly given the sub-field’s historical roots in the Cold War. Despite the Soviet Union’s demise, IR scholars continue to highlight the importance of threats in explaining interstate behavior—particularly in the midst of crises that may precipitate armed conflict. Countless theoretical, historical, and formal analyses have been devoted to the study of threat perception. Besides crisis scenarios, perceived threats also affect alliance formation, trade, defense policies, and the like.<sup>1</sup>

---

\*Please do not distribute without authors’ permission. Authors’ e-mails: [azusak@stanford.edu](mailto:azusak@stanford.edu) and [ericmin@stanford.edu](mailto:ericmin@stanford.edu). We thank John Freeman, Justin Grimmer, Ken Schultz, Curt Signorino, Gaurav Sood, and participants of the Annual Meetings of the Society of Political Methodology in Atlanta (July 2014) and Political Science Association in Washington, D.C. (August 2014) for their helpful comments. Eric Min gratefully acknowledges the financial support of the National Science Foundation’s Graduate Research Fellowship #DGE-114747. All errors are our own.

<sup>1</sup>For notable examples, see Jervis 1976, 1986; Walt 1987; Fearon 1994; Sartori 2002; Stein 2013.

However, most of these works are limited to case studies of individual crises, broad game-theoretical models, or quantitative methods that use “black box” data with wars or disputes as the unit of analysis (which feature no intra-event information). This is no coincidence, as little systematic data exists to test these theories otherwise.

In this note, we demonstrate a manner in which to create such data. By applying machine learning and text analysis techniques on digitized U.S. documents from the *Foreign Relations of the United States 1945-1980* (FRUS) collection, we demonstrate a straightforward method in which to generate time-series data on elite threat perception.<sup>2</sup> The general process can also be applied to create measures of other concepts of interest on other corpora of text. These novel data and approaches, which help identify a key IR concept, will prove crucial in substantiating long-standing IR theories that have not been evaluated with more statistical techniques.

The note proceeds as follows. Section 2 briefly discusses the importance of threats in IR theory, and previous efforts made to measure the concept. Section 3 outlines the data and machine learning techniques used, presents the generated data, and demonstrates the validity of the newly predicted data. Section 4 uses time series analysis to revisit the topic of reciprocity—a seminal IR topic that faded due to lack of actionable data. Section 5 concludes and proposes several research agendas based on the paper’s methods and findings.

## 2 Appreciating Threat Perception

Threats exist in at least two forms. The first, and more prevalent, is as a latent and uncertain state (Schelling 1960, 1966). Walt’s (1987) theory of alliance formation heavily engages in this language, as does Buzan’s (1983) discussion of national security policy. Attempts to gauge this include military capabilities/expenditures or the more generalized CINC score, which ostensibly capture a state’s potential to be threatening at any given year;<sup>3</sup> or Kendall’s  $\tau_b$  (see Signorino and Ritter 1999) and Gartzke’s Affinity index (see Gartzke 2006), which loosely capture intentions by gauging policy similarities between countries.

The second conception of threat perception is in the context of crisis bargaining, where actors make explicit statements with the intent of coercion. Such a conception is at the heart of audience cost theory (Fearon 1994; Schultz 2001; Leventoğlu and Tarar 2005; Press 2005; Tomz 2007; Weeks 2008; Downes and Sechser 2012), as well as most formal models of diplomacy during crisis bargaining (Sartori 2002, 2005; Kurizaki 2007; Trager 2010; Ramsay 2011). Event-based conflict data is the norm in these studies. The Correlates of War Militarized Interstate Dispute (MID), the International Crisis Behavior (ICB), and the Military Compellent Threat (MCT; Sechser 2011)

<sup>2</sup>Most documents included in the FRUS were formerly classified and released decades later. As such, these memos are one of the best and only channels to see elites’ genuine and contemporary sentiments about international affairs.

<sup>3</sup>For an example using military expenditures over time, see Williams and McGinnis (1988). For CINC scores, see Jo and Gartzke (2007).

datasets are all attempts to track incidents of threats and their aftermaths.

These efforts feature at least three shortcomings to various extents. First, the data are likely to leave out less public or more non-militaristic incidents such as economic sanctions or tariffs that may cause elites to feel endangered. Second, and relatedly, the data do not look at actual elites' perceptions of threat and instead assumes events to indicate them. There is thus a substantial bias toward highlighting historically momentous events but bypassing incidents that are either less notable or never realized. Third, highly aggregated "black box" data, which either lists an entire crisis or year as a single unit, cannot capture changes in perceptions over time.<sup>4</sup>

### 3 An Alternative Approach

We would ideally have a measure of perceived threats as seen by elite decisionmakers that must directly deal with them. The data should also pick up on less conspicuous incidents or even concerns about events that never come to fruition.

Statistical learning methods applied to text data are a useful tool to generate such a new set of measures. In contrast to predominant statistical methods in political science, which are based on causal or explanatory inference (that is, identifying and quantifying the specific mechanisms that link a set of explanatory variables  $X$  to a dependent variable  $Y$  through  $Y = f(X) + \varepsilon$ ), machine learning approaches are more focused on predictive inference. Instead of attempting to understand the precise mechanisms connecting  $X$  to  $Y$ , learning methods simply seek to make the best predictions and leave the nature of  $f$  aside. Therefore, these methods tend to use more flexible approaches and computational approaches that do not necessarily rely on parametric models. Their main objective is to generate predictions for out-of-sample observations, while causal/explanatory analysis is more geared toward understanding relationships within the given data.

A comprehensive and well-established source of documents on diplomacy is United States Department of State's Office of the Historian's *Foreign Relations of the United States 1945-1980* (FRUS) collection, which archives tens of thousands of intra-government documents from 1945 to 1980.<sup>5</sup> We scraped this collection, in which the vast majority of the documents are internal memoranda sent within the cabinet and with American diplomatic outposts. Almost every document included in the FRUS was classified for at least twenty-five to thirty years before being published. Therefore, memos in the FRUS are a compelling manner in which to determine the true sentiments and

---

<sup>4</sup>One recent work worth mention is Yarhi-Milo (2013), who tries to use qualitative evidence to uncover how the United States policymakers perceived the intentions of the Soviet Union during several administrations. Nonetheless, her historical and archival analysis of 30,000 documents, which is focused on *how* intentions are interpreted, does not produce a systematic manner in which to capture the way in which American elites ultimately did interpret or foresee ominous Soviet behavior—that is, the actual "output" of these interpretive mechanisms. In what follows, we present a method that can overcome this and the other previously mentioned issues.

<sup>5</sup>The collection is available at <http://history.state.gov/historicaldocuments>. The FRUS begins in 1861, but the State Department has thus far only digitized documents post-1945.

thoughts of policymaking elite in the White House at that time, unadulterated by public opinion, the media, or memoir-like retrospection. In order to ensure coherent, consistent, and substantively interesting results, we limit ourselves to volumes of the FRUS that are centered on the Soviet Union and the Eastern Bloc during the Cold War—a natural and substantively important choice given the FRUS’s time coverage.<sup>6</sup>

This results in a total of 8,474 documents with 31,102 unique tokens. When sparse terms not occurring in at least 80% of documents are removed, we are left with 308 unique tokens.<sup>7</sup> This data is represented by a document-term matrix, where each row represents a document, and each column is a count of the number of occurrences of a token. As such, our document-term matrix has 8,474 rows and 308 columns.

To predict the dependent variable for each of the 8,474 documents we collected, the models must be provided with some data in which the actual values of the dependent variable are given. This constitutes *learning* or *training* data.<sup>8</sup> To this end, we draw a random sample of 664 documents and hand-code them for expressing perceived threat (1) or not (0).<sup>9</sup>

“Threat” is a somewhat vague term. Here, given the nature of the FRUS, we narrow the search to only *elite* perceptions—how decisionmakers in government interpret possibilities of threat, rather than the general public and/or media.<sup>10</sup> Within that context, we classify memos as expressing perceived threat if they convey a real and/or imminent potential for either (1) armed hostilities involving the United States and/or (2) a substantial risk in the bipolar balance of power, usually via encroachment on allied nations or direct Soviet intervention into other countries.

Despite the potential imprecision of this definition, it appears to be a useful set of criteria that permits reasonable classifications. We provide a couple examples below, which are excerpted paragraphs from documents. Sentences that seem to especially indicate perceived threat are italicized and highlighted in red for illustrative purposes.

*”The military are worried about the general attitude in Europe. There has been a let up on the ground that the threat is not as great as it was assumed to be. From our viewpoint the USSR’s capability is still there. We have no way of knowing what Soviet intentions are. Initially, our problem was one of encouragement and of laying out the general outline of a plan for European defense.”*<sup>11</sup>

---

<sup>6</sup>A full listing of the volumes is in Appendix 1.

<sup>7</sup>The high degree of sparse term removal also led to stronger predictive results for our models.

<sup>8</sup>We use *supervised* learning models, where the values of the dependent variable are incorporated into a model. In *unsupervised* learning models, values of the dependent variable are unavailable, and observations are instead categorized or reduced based on other features. Clustering methods and principal component analysis are perhaps two of the most prevalent forms of unsupervised learning.

<sup>9</sup>It is plausible to code 8,474 documents by hand; many scholars have done far more in their studies. However, part of our goal is to create a scalable approach that can be used for far larger sets of data, which we hope to collect for future studies (see the Conclusion).

<sup>10</sup>Of course, negative discourse by the public and/or media can exacerbate policymakers’ perceptions of threat, but we would be more interested in how/whether such negative discourse has an effect on elite perceptions.

<sup>11</sup>Memorandum of Discussion of State–Mutual Security Agency–Joint Chiefs of Staff Meet-

*”We are alive to problems presented by GDR harassment at crossing point and on autobahn. Latter particularly disturbing since touches Allied vital interest, and manner dealing with possible pattern such encroachments being considered in Quadripartite contingency planning. We shall look to you and our Mission in Berlin to detect any such pattern as it begins to emerge. We shall also welcome any suggestions for particular measures to forestall or retaliate for harassment.”*<sup>12</sup>

The first example contains the word “threat.” We emphasize, however, that our coding does not automatically classify memos involving this word to expressed perceived threat. For example, consider the following passage from a letter sent by President Eisenhower to Yugoslavia’s President Yosip Broz Tito:

I know that we both firmly share a common desire to see an evolution and improvement of the international atmosphere take place so that the world will be freed from the threat of war and human and material resources now used for military purposes can be diverted to improving living standards. The United States Government is now engaged in working on the many diplomatic problems involved, some of which will be discussed at the forthcoming conference of Foreign Ministers in Geneva.<sup>13</sup>

The first sentence contains the words “threat” and “war”, but we classify this document as exhibiting no direct perceived threat, since both concepts are mentioned in very generic terms. Our manual classification is therefore contextual and not motivated by instinctual reactions based on some set of predetermined trigger words. Coding was also performed without seeing specific dates in order to ensure that classifications were not directly based on knowledge of events occurring at a particular point in time.

Of the 664 hand-coded documents, 92 (14%) exhibit perceived threat, while the remaining 572 (86%) do not.<sup>14</sup>

### 3.1 Model Selection and Predicting Perceived Threat

Predictive models are assessed by how well they make predictions for out-of-sample observations not used to develop the model itself — that is, how well they extrapolate. Recall the classic bias-variance trade-off: A model that perfectly fits the data used to create the model will have zero bias but likely perform poorly on new observations. Minimization of out-of-sample error is therefore the

---

ing, Held at the Pentagon Building, January 28, 1953, 10:30 p.m. Available at <http://history.state.gov/historicaldocuments/frus1952-54v05p1/d374>.

<sup>12</sup>Telegram From the Department of State to the Mission at Berlin, October 3, 1961, 8:53 p.m. Available at <http://history.state.gov/historicaldocuments/frus1961-63v14/d167>.

<sup>13</sup>Letter From President Eisenhower to President Tito, September 19, 1955. Available at <http://history.state.gov/historicaldocuments/frus1955-57v26/d256>.

<sup>14</sup>Classifications were ultimately done in a binary manner in order to ensure enough of each class existed for effective model training. An attempt to create a low-moderate-severe classification system resulted in fewer than 10 severe cases and, consequently, poor predictions.

goal. However, since out-of-sample error is unknowable; we cannot compare predicted and actual values, since the point of using statistical learning models is to generate predictions on data for which actual values are unknown.

A process called cross-validation is thus used to overcome this problem in both choosing and evaluating predictive models. A random sample of the training data is left aside to emulate out-of-sample data (but for which the real values of the outcome are known) and is called the validation set. The remaining data is used to train the model, which then generates predictions on the validation set. These predictions are compared to the known values to create an estimate of the model’s performance on out-of-sample data.

Choosing the best model is a two-step process. Suppose we seek to find the best model out of many, where  $\mathcal{M} = \{A, B, C, \dots\}$  are the models of interest. Numerous predictive models exist; common ones include Naive Bayes, support vector machines, and  $k$  nearest neighbors. Each model has various attributes (called tuning parameters) that can be adjusted to tweak performance. For instance, a key parameter in  $k$  nearest neighbors is  $k$  itself — how many nearby observations around some new data point  $x'_i$  should be used to make predictions? This means that each model has different variants (i.e.  $A_1, A_2, A_3, \dots$ ) from which to choose. The first step is therefore model optimization, or finding the best version of  $A, B, C$ , etc. We could label each as  $A^*, B^*, C^*$ , and so on. The second step is model evaluation/selection, where the performance of  $A^*, B^*, C^*$ , and the rest are compared to one another to find the best overall model,  $\mathcal{M}^*$ . Each step requires cross-validation.

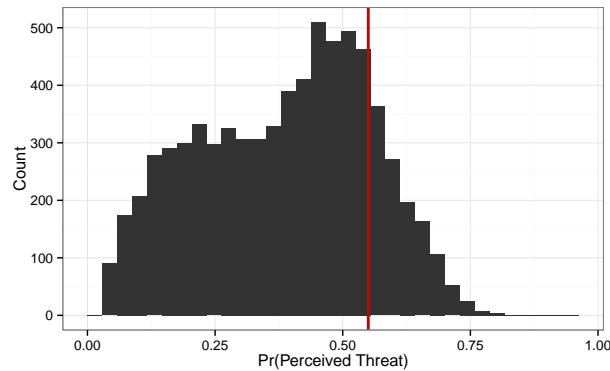
To this end, only 75% of the 664 hand-coded memos are randomly chosen to be a training set (498), and the remainder a test set (166). To find the best specification of each model ( $A^*, B^*, C^*$ , etc.), five-fold cross-validation is used. The 498 memos are randomly split into five nearly-equal subsets of about 100 each. In each of five iterations, a model was trained using four of these subsets (about 400 documents) and then tested on the remaining one (about 100 documents), which differed each time. The average error rate across these five trials is the model’s cross-validation error.<sup>15</sup> Each optimal model specification is then used to make predictions on the test set of 164 memos, which are compared to the known values to determine test error. The one with lowest test error is the optimal model,  $\mathcal{M}^*$ .

Of many models attempted, one was consistently the most effective for our particular task: the balanced random forest (Breiman 2001b; Chen et al. 2004). More details on random forest models are available in Appendix 2 for those interested. However, the particular model chosen is only important in that it exhibited the best levels of performance compared to many other alternatives. Although we leave more descriptive diagnostics of all models including the balanced random forest to Appendix 3, we note that both did a commendable job of identifying a somewhat elusive term such as perceived threat.

---

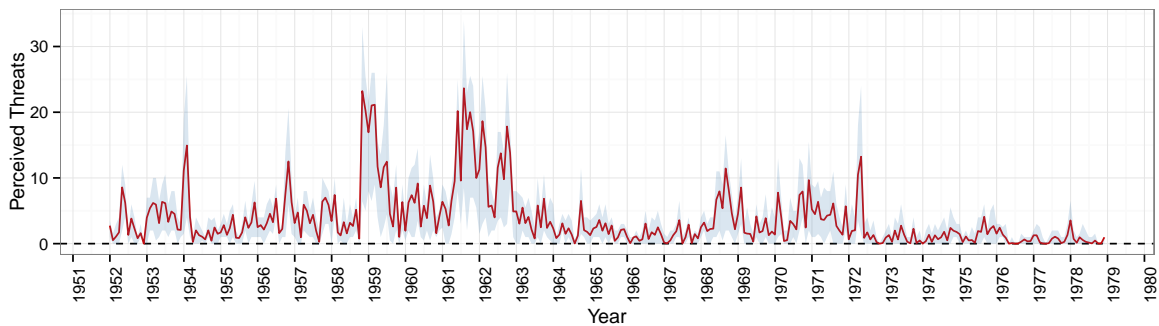
<sup>15</sup> $n$ -fold cross-validation is used to alleviate data loss from validation techniques, and to decrease the variance in estimates that may coincidentally arise from random splits in the data.

Once the balanced random forest model was trained and optimized using the training set of 664 memos, it was used to predict perceived threat on the remaining 7,810 documents based on the word counts in each. The balanced random forest predicts the probability of each document expressing perceived threat, which is then dichotomized with 0.55 as the cut-point. This threshold was also determined using cross-validation techniques. The distribution of the data across predicted probabilities is shown in Figure 1. These predictions (as well as the original training data) are aggregated to the weekly level to create data of elite perceived threat.



**Figure 1:** Distributions of predicted probabilities of expressing perceived threat for the test data. 1,219 documents are predicted to express perceived threats, and 6,589 documents are not.

The monthly raw data is illustrated in Figure 2. Confidence intervals were made by bootstrapping the entire balanced random forest 100 times. Each iteration uses a different sub-samples for training and testing, which also addresses concerns that any results are based on a “lucky” split in the data. More bootstrap results are available in Appendix 4.



**Figure 2:** Bootstrapped raw predicted data. The main red line indicates the mean level of threat perception from the 100 iterations.

This plot features some notable spikes in perceived threat. In order to assess their validity, we compare the highest eighth levels of perceived threat with events in the International Crisis Behavior dataset where the Soviet Union was the instigator and/or in which the Soviet Union was at least

economically involved. This generates a list of twelve events between 1952 and 1978.<sup>16</sup> Table 1 indicates that the machine-generated data performs quite well.

Dates	Event	ICB #	Description	In Data
6/16/52 - 7/52	Catalina Af-fair	137	Two Soviet fighters shoot down a Catalina flying boat of Sweden, a historically neutral party.	N
6/17/53 - 7/11/53	East German Uprising	141	East German workers go on strike; USSR sends in troops to forcibly and violently quell riots.	Y
10/23/56 - 1/57	Hungarian Uprising	155	A massive uprising demands withdrawal of Soviet troops; the USSR further mobilizes and installs a puppet government.	Y
11/27/58 - 9/15/59	Berlin Dead-line	168	Khrushchev issues an ultimatum to the U.S, U.K., and France to withdraw from West Berlin.	Y
8/61 - 10/28/61	Berlin Wall	185	The Berlin Wall is erected and a no-man's land is established, stopping the flow of refugees from East Germany.	Y
10/30/61 - 11/24/61	Soviet Note to Finland	189	USSR requests talks with Finland out of concerns that elections would end Finland's neutrality.	Y
10/16/62 - 11/20/62	Cuban Mis-sile Crisis	196	The US identifies Soviet missiles in Cuba and initiates a blockade of Cuba in response.	Y
4/9/68 - 10/18/68	Prague Spring	227	Eastern Bloc fears Prague democratization efforts and forcibly takes power of the country.	Y
3/2/69 - 10/20/69	Ussuri River*	231	Chinese ambush of Soviet troops sparks clashes and tensions between both nations.	N
3/8/69 - 8/7/70	War of Attrition*	232	Egypt (with Soviet support) launches a sustained offensive against Israel (which had American support).	N
9/16/70 - 10/23/70	Cienfuegos Base	239	US intelligence finds Soviets building a submarine base in Cienfuegos, Cuba.	Y

**Table 1:** Crisis events involving the Soviet Union, as identified by the International Crisis Behavior dataset. The two events marked with asterisks do not take place in Europe, so we do not expect to noticeable changes in perceived threat in our data.

<sup>16</sup>Using the systemic level version of the ICB, we identified cases in which `trigent==365` and/or where `suinv > 2`.



Our data passes an important validity check with considerable success: it finds threat perception in the midst of explicit threats. As we stated in the end of Section 2, however, datasets such as the ICB only record highly visible events, overlooking any potential for perceived threats in the absence of very conspicuous events. The generated data appears to address this, finding other spikes in threat perception outside of the ICB’s list. We take caution in providing ex post rationalizations, yet we find them to be reasonable given the strength of the data in identifying ICB incidents. Indeed, these additional points are encouraging, demonstrating that the data confirms previous expectations while also going further. These events are:

1. The Berlin Conference of January and February 1954, where the foreign ministers of the United States, Great Britain, France, and the Soviet Union (the “Big Four”) discussed the resolution of the Korean War, continuation of the Indochina War, and the fate of occupied Germany and Austria
2. The U-2 incident, where an American spy plane was shot down over Soviet territory, in May 1960
3. The aftermath of the failed Bay of Pigs invasion, the Soviet Union’s recognition of the new Cuban government, as well as the Vienna Summit between Kennedy and Khrushchev, between May and June 1961
4. Increasing tensions with Cuba in 1962, including the US embargo in February 1962, and growing suspicions about Soviet missiles in August 1962
5. The concluding stages of the Strategic Arms Limitation Talks (SALT I) in early 1972

As one last manner to assess the data’s validity, we present an excerpt of a memo that the balanced random forest determined to have a moderately high probability (58%) of expressing perceived threat.

Soviet allegations of supposedly impending Israeli military actions are particularly mischievous since they could easily trigger a chain of hasty Arab and Israeli military actions. (The Arab-Israeli war of 1967 began in large measure because of Soviet rumor-mongering.) The allegation of Turkish invasion plans against Cyprus can only hinder the delicate political and diplomatic efforts to resolve the crisis. Soviet misinformation concerning our role on Cyprus and on the next stage of Middle East diplomacy is obviously designed to compromise our efforts at mediation on both situations.<sup>17</sup>

This document comes from November 1974, as the United States and Great Britain attempted to mediate in the aftermath of the Turkish invasion of Cyprus.<sup>18</sup> Rather than being part of one of the major crises identified in Table 1, this memo exemplifies perceived threat in a less prominent but equally important context. The ability to identify and quantify these “blips” is arguably the most significant merit to our methods and data. We now turn to its application in this light.

<sup>17</sup>Memorandum From Secretary of State Kissinger to President Ford. November 20, 1974. Available at <http://history.state.gov/historicaldocuments/frus1969-76v16/d89>.

<sup>18</sup>Cyprus hosted radar and spy stations that both nations used to scrutinize the Soviet Union.

## 4 Time-Series Analysis: Revisiting Reciprocity

The newly generated data does an impressive job of identifying perceived threats from the Soviet Union and/or the general specter of Communism. As we noted earlier, very few fine-grained time-series datasets exist in the IR literature. While creating more data is an obvious longer-term goal, we want to immediately demonstrate the utility of this method and data by speaking to a classic IR topic: reciprocity.

Works concerning inter-state reciprocity (Goldstein and Freeman 1990; Goldstein 1991; Goldstein et al. 2001; Brandt et al. 2006) have largely been untested with the concept of threat placed in the analysis, even when expected threats are an explicit part of a theory (see Williams and McGinnis 1988 and McGinnis and Williams 1989). Furthermore, important studies of reciprocity use years or months as the unit of time, making the studies very coarse. In critiquing Williams and McGinnis’s (1988) annual-level study of superpower rivalry, Freeman (1989) demonstrates that highly aggregated temporal data can result in severely faulty inference. Shellman (2004) and Thomas (2014) more recently identify the same issue. While the end of the Cold War accelerated the decline of reciprocity research in the early 1990’s, lack of better data also played a major role. Our threat perception measure can aptly address this problem.

A considerable amount of reciprocity research utilizes event data such as the Conflict and Peace Data Bank 1948-1978 (COPDAB; see Azar 1980 and Schrodtt 1995).<sup>19</sup> The COPDAB is a longitudinal study that catalogs international and domestic events or interactions for about 135 countries. Each event is classified on nine different variables. Two of primary importance to us are the target of the event, and a scale/score for whether the event was cooperative or conflictual in nature.<sup>20</sup>

The original scale runs from 1 to 15, where 1 represents “voluntary unification into one nation,” 8 represents “neutral or non-significant acts for the inter-nation situation,” and 15 represents “extensive war acts causing deaths, dislocation or high strategic costs.” This scale is then reweighted according to scholars’ interpretations of how favorable or unfavorable events seemed in relation to the neutral point of 8. In the end, the scale ranges between  $-102$  and  $92$ , with 0 representing neutrality. While COPDAB records conflictual events with negative scores and cooperative events with positive ones, we inverted this scale to more intuitively align with our threat perception data.

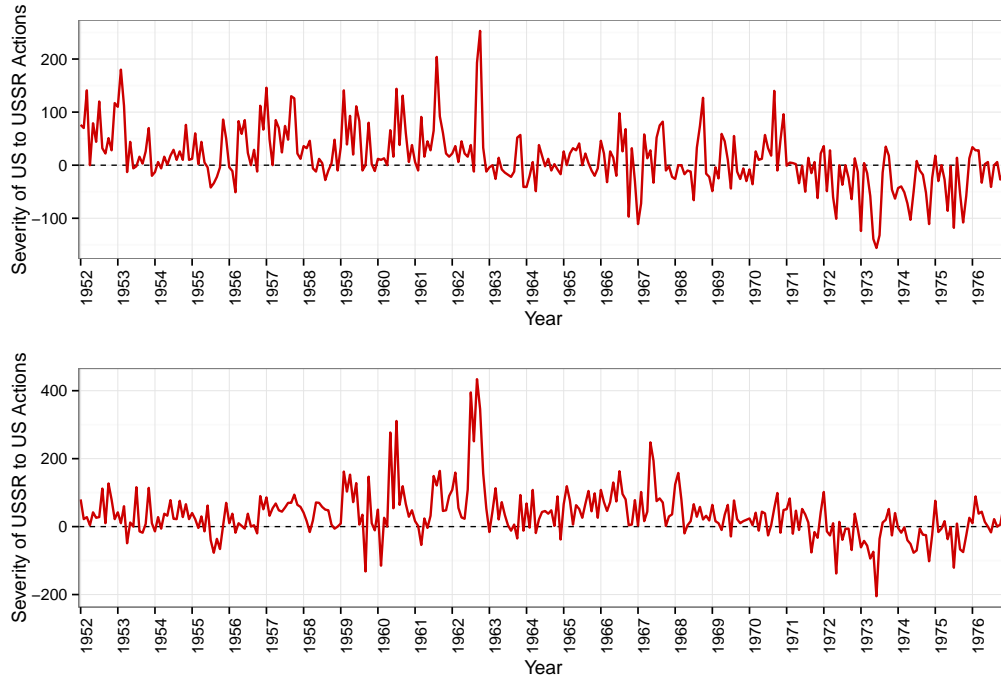
Of about 431,000 events in the dataset, 2,857 involve Soviet actions toward the United States, and 2,747 involve American actions against the Soviet Union. We aggregate these events’ scores to the

---

<sup>19</sup>At least two other event datasets are worth mention: the hand-coded World Event/Interaction Survey (WEIS) dataset (McClellan 1999); and the machine-coded Global Database of Events, Language, and Tone (GDELT). The former covers 1966 to 1978, which misses more than half the years in our threat data; the latter covers 1979 to present, precisely beginning where our threat data ends.

<sup>20</sup>Several other alternative typologies exist, such as the Goldstein index (Goldstein 1992); the Conflict and Mediation Event Observations, or CAMEO, system (Schrodtt 2012); the World Event/Interaction Survey, or WEIS, coding scale (used by the Kansas Event Data System, or KEDS; McClellan 1999); and the Integrated Data for Events Analysis, or IDEA, framework (Bond et al. 2003). The Goldstein, CAMEO, and IDEA typologies are all extensions or refinements of WEIS, which itself is highly correlated with COPDAB.

*weekly* level and match them with a weekly version of our predicted data. See Figure 3 (which displays monthly totals to reduce noisiness). This permits us to examine the relationship between *perceptions* (through our new data) and actual *events* (through the COPDAB data).



**Figure 3:** Actions between the United States and Soviet Union as measured in COPDAB. Top figure is  $US \rightarrow USSR$ , and bottom figure is  $USSR \rightarrow US$ . The  $y$ -axis increases with conflictual tendencies. Monthly totals are used here to reduce noisiness of the plot.

#### 4.1 Model Specification

Augmented Dickey-Fuller and Phillips-Perron tests indicate that the data satisfy the stationary assumption, allowing for use of a simple vector auto-regressive (VAR) model. Depending on the test or measure used (including AIC and BIC), lags between 2 and 7 appeared to be appropriate. For the sake of parsimony, and given that the results barely differ within this range of lags, we opt to use three weeks. This length of time is an intuitive period of time during which particular events may be anticipated or reacted upon. See Appendix 5 for more details.

In words, our main dependent variable is the degree of hostile acts from the Soviet Union against the United States at some week  $t$ , as determined from COPDAB data. The measure is directional and is modeled to be a function of three sets of predictors: lagged measures of US threat perception of the USSR (obtained through our new data,  $TP$ ), lagged measures of hostile events from the USSR to the US ( $CD_{USSR \rightarrow US}$ ), and lagged measures of hostile events from the US to the USSR ( $CD_{US \rightarrow USSR}$ ). Three weeks of lags are included for each measure. We are thus assuming that recent perceived threats and actions in both directions are related to actions taken at some time  $t$ ,

so some feedback may be possible.

In all analyses, we utilize a normalized version of the weekly threat measure,  $\frac{\# \text{ Threats in Week}}{\text{Memos in Week}}$ . This is to ensure that results are not simply driven by larger numbers of documents in any particular time period, particularly since the FRUS is a large but curated collection.

The VAR model is specified through three sub-models involving threat perception (TP) and COPDAB (CD) data, which are estimated separately.<sup>21</sup>

The main models are thus specified using the generated threat perception (TP) and COPDAB (CD) data:

$$\begin{aligned}
 CD_{USSR \rightarrow US,t} &= \alpha_0 + \sum_{i=1}^3 \beta_i TP_{t-i} + \sum_{j=1}^3 \gamma_j CD_{USSR \rightarrow US,t-j} + \sum_{k=1}^3 \eta_k CD_{US \rightarrow USSR,t-k} + w_1 \\
 CD_{US \rightarrow USSR,t} &= \bar{\alpha}_0 + \sum_{i=1}^3 \bar{\beta}_i TP_{t-i} + \sum_{j=1}^3 \bar{\gamma}_j CD_{USSR \rightarrow US,t-j} + \sum_{k=1}^3 \bar{\eta}_k CD_{US \rightarrow USSR,t-k} + w_2 \\
 TP_t &= \tilde{\alpha}_0 + \sum_{i=1}^3 \tilde{\beta}_i TP_{t-i} + \sum_{j=1}^3 \tilde{\gamma}_j CD_{USSR \rightarrow US,t-j} + \sum_{k=1}^3 \tilde{\eta}_k CD_{US \rightarrow USSR,t-k} + w_3
 \end{aligned}$$

The first of the three models, with Soviet actions against the US as the dependent variable, is our primary focus. The second model explores determinants of American hostility toward the USSR. The third, which treats threat perception as the outcome variable, is a rough manner in which to adjudicate whether events can predict subsequent changes in threat perception—that is, the existence of a reverse effect.

In sum, these three models seek to forecast or predict future events (or in the third equation, threat perception) based on past interactions and perceptions. While we emphasize that this time-series analysis cannot determine causality, the ability to forecast future behavior is of substantial importance in its own right.

## 4.2 Results

The main results of the vector auto-regression analyses are presented in Table 2 and are quite striking.

Model 2 is our main result, which indicates that perceived threat of Soviet intent *precedes* or *predicts* actually threatening Soviet behavior. This may seem intuitive or even unsurprising: an effective bureaucrat, particularly focused on the Soviet Union during the Cold War, seeks to anticipate Soviet

<sup>21</sup>Structural VAR (SVAR) models estimate these equations simultaneously. However, the SVAR results are substantively similar for our data.

**Table 2:** Results of Vector Auto-Regressive (VAR) Models

	Dependent Variables				
	(1)	(2)	(3)	(4)	(5)
	$TP_{USSR,t}$	$CD_{USSR \rightarrow US,t}$	$CD_{US \rightarrow USSR,t}$	$CD_{USSR \rightarrow US,t}$	$CD_{US \rightarrow USSR,t}$
Threat Perception					
$t - 1$	0.16*** (0.03)	10.65** (4.27)	0.81 (3.72)	11.42** (5.11)	4.32 (4.43)
$t - 2$	0.12*** (0.03)	10.17** (4.29)	4.47 (3.73)	5.29 (5.19)	-4.29 (4.48)
$t - 3$	0.11*** (0.03)	7.22 (4.26)	4.39 (3.71)	1.08 (5.13)	2.18 (4.45)
COPDAB					
$USSR \rightarrow US, t-1$	0.00 (0.00)	0.15*** (0.03)	0.07*** (0.02)	0.16*** (0.03)	0.08*** (0.02)
$USSR \rightarrow US, t-2$	0.00 (0.00)	0.13*** (0.03)	0.03 (0.02)	0.14*** (0.03)	0.03 (0.02)
$USSR \rightarrow US, t-3$	0.00 (0.00)	0.08*** (0.03)	-0.03 (0.02)	0.08*** (0.03)	-0.02 (0.02)
$US \rightarrow USSR, t-1$	0.00* (0.00)	0.02 (0.03)	0.10*** (0.03)	0.02 (0.03)	0.10*** (0.03)
$US \rightarrow USSR, t-2$	0.00 (0.00)	0.04 (0.03)	0.11*** (0.03)	0.04 (0.03)	0.11*** (0.03)
$US \rightarrow USSR, t-3$	0.00 (0.01)	0.02 (0.03)	0.11*** (0.03)	0.03 (0.03)	0.11*** (0.03)
Constant	0.08*** (0.01)	0.22 (0.94)	-0.79 (0.89)	-2.40 (2.60)	-0.37*** (2.26)
$N$	1407	1407	1407	1407	1407

Standard errors in parentheses.

Significance at \* $p < .10$ ; \*\* $p < .05$ ; \*\*\* $p < .01$  levels.

intentions. Nonetheless, this dynamic has never been tested in any rigorous manner.<sup>22</sup> These results likewise challenge the more general notion that foreign policymaking elite are largely reactive and unaware of future possibilities. Model 1 shows that the reverse effect does not appear to exist: Using threat perception,  $TP_t$ , as the dependent variable only indicated strong autocorrelation in the lagged  $TP$  measures.

As we would intuitively expect, Model 3 indicates that American threat perception does not appear to precede threatening American behavior against the Soviet Union. However, very recent Soviet actions against the United States ( $CD_{USSR \rightarrow US, t-1}$ ) do appear to systematically come before US hostility. This finding suggests a dynamic of reciprocity, in which the United States mildly responds to Soviet provocation.<sup>23</sup>

<sup>22</sup>This finding also contradicts Wolfe's (1979) study of Soviet threat, which uses historical research to assert that "[t]he real issue is not whether the Soviets become more aggressive, but whether the U.S. decides to view them as aggressive (31; his emphasis).

<sup>23</sup>We note that Zinnes (1968), who did a somewhat similar albeit less technical study on World War I, did not find evidence of reciprocity (see her Hypothesis 4).

The bootstrapped mean threat perception measure in Models 2 and 3 is based on a dichotomous classification. Models 4 and 5 mirror Models 2 and 3 respectively, but use a bootstrapped mean threat perceptions measure based on the individual estimated probabilities of each document in a week; see Figure 1. This roughly represents a continuous version of the threat perception variable. Table 2 indicates that the results are largely the same.

We highlight that these effects are short and do not create an infinite loop between threat perception and events, which would complicate our inferences.<sup>24</sup> The three coefficient estimates for  $CD_{USSR \rightarrow US}$  in Model 1 indicate that past Soviet actions against the United States do not influence subsequent perceptions of threat. Similarly, the three estimates for  $CD_{US \rightarrow USSR}$  in Model 3 show that past American actions do not affect later Soviet actions.

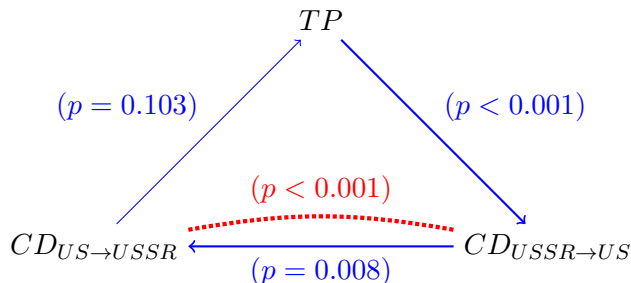
We can investigate this temporal aspect further through tests of Granger and instant causality.<sup>25</sup> Granger causality best applies to simple models where the outcome is modeled using only one other variable. When two or more are included, Granger causality between two time series may be confounded or biased by other time series in the model. We therefore test for partial Granger causality, which takes these relationships into account. See Guo et al. (2008). Instant causality exists if some correlation exists between two time series in the same period. This is determined by regressing the residuals from the time series (the unexplained variation in the dependent variable being regressed only on lagged covariates) on the contemporary covariates.

The results of these tests are summarized visually in Figure 4. A strong Granger arrow points from threat perception to Soviet hostility, which mirrors the main finding from Table 2. A similarly strong Granger relationship exists from Soviet hostility to US hostility, again reflecting the previous findings. But interestingly, we find a very strong degree of instant interaction between American and Soviet hostility, which indicates very swift reciprocity between the two entities. Responses to one side’s aggressive behavior are almost immediate, but quite disproportionate: American responses appear far less hostile than the Soviet provocation. An arguably non-significant Granger arrow connects American hostility to American threat perception. This might indicate some form of belief updating on the part of American elites. A more comprehensive and dynamic presentation of these results through impulse response functions is available in Appendix 5.

Figure 4 suggests a particular order to the three phenomena examined. Threat perception is followed by Soviet aggression, which is followed by American aggression in response. While Figure 4 hints at a weakly circular dynamic, the low degree of contemporaneous correlation of residuals in these models (as well as our previous discussion about Models 1 and 3) strongly indicates that threat perception is the initial trigger. Even if this does not explicitly prove causality, the predictive

<sup>24</sup>Impulse response functions in Appendix 5 also attest to the brevity of this dynamic linking threat perception and hostile events.

<sup>25</sup>Note that these concepts’ names have historical roots that are unrelated to the modern causal inference framework. Both only refer to predictive causality—the ability for a change in some  $X$  to provide information about future values of  $Y$ . Granger causality exists in a VAR model if past/lagged values of a covariate  $X$  appear to result in subsequent changes in the outcome  $Y$ .



**Figure 4:** Schematic to illustrate causal effects indicated by time series analysis. Blue solid lines represent Granger causality; red dotted line represents instant causality.

power of our model indicates a temporally cogent story: During the Cold War, the Soviet Union tended to be the instigator. Elites in the United States presaged these actions to some degree. However, instead of taking action based only on this perception, they would respond only once those provocations took place. Movements in American threat perception thus indicate predilection of Soviet hostilities, and not self-fulfilling prophecies where US elites take actions in anticipation of Soviet misconduct.

Threat perception, which has not been effectively measured or incorporated in past studies, is thus a key variable to understanding and better predicting aggressive behavior. Indeed, technically speaking, the correlation of residuals between  $CD_{US \rightarrow USSR}$  and  $CD_{USSR \rightarrow US}$  is 0.49 in Goldstein and Freeman (1990, 171). Our model, which includes threat perception, brings this number down to 0.35—a substantial improvement. We also note that our weekly level threat perception data represents a concerted effort to address Freeman’s (1991) proviso on high-level aggregation. This proves to be critical in revealing useful relationships: When the same analysis is run on monthly-level data, some of our findings about Granger causality disappear. Disaggregation to the weekly level illuminates real directionality of reciprocity in a manner that more highly aggregated measures struggle to identify.

## 5 Conclusion

The sheer importance of threats and perceptions to IR theory is rivaled by the field’s inability to evaluate theories on either concept in a truly rigorous way. Lack of systematic data has forced scholars to rely on case studies, formal models, and very coarse quantitative data to support their claims. While each of these holds its merits, none of them can claim to be as standardized and generalizable as would be ideal. Modern computational methods like text analysis and statistical learning are providing the tools to overcome these chronic limitations. These techniques are relatively recent to political science and have been even slower to infiltrate international relations—the field that arguably is in greatest need of more fine-grained data.

Our paper has sought to demonstrate this potential by scraping the *Foreign Relations of the United*

*States 1945-1980* (FRUS) collection and showing that that classification models—particularly balanced random forests—can identify useful and plausible moments in which elite U.S. policymakers perceived a sense of threat from the Soviet Union during the Cold War.

This data creation process is a novel contribution in its own right. However, our time-series analysis using the Conflict and Peace Data Bank also speaks to a substantial past literature on reciprocity between superpowers. Threat perception is a key variable to understanding aggressive behavior but has never been measured or incorporated into a quantitative study, particularly to this level of exactitude.

The finding that threat perception precedes truly threatening behavior has substantial empirical implications on IR, where threats have usually been operationalized using explicit threat-making and major events. Threat perception does not necessarily involve either, and would precede them in any case. In contrast to our data, which is based on private communications, measuring threats via publicly observed events is flawed, as it cannot capture discreet concerns or events that ultimately never come to pass. Studies that attempt to capture threats using explicit diplomatic overtures or visibly tense moments therefore may not be revealing what they claim.

These two parts, when considered together, yield a model that forecasts behavior better than any extant models that either use longer units of time or past events without accounting for elite perceptions. This note focused on the Soviet Union as the main opponent due to its historical significance, availability of declassified data, and connection to past reciprocity scholarship. Nonetheless, Russia's recent and renewed assertiveness, as well as current China-Japan and China-United States tensions, emphasize the utmost importance of using novel techniques and data to better understand and anticipate international interactions in the twenty-first century.

The initial successes of this paper provide compelling evidence that text analysis and machine learning techniques can be powerful and approachable tools for the study of international relations and foreign policy. The FRUS, in its digitized form online, is a rich and easily accessible source of data on American diplomacy between 1945 and 1980. However, the full FRUS extends back to 1861 in book form, and several archives across the country hold hundreds of linear feet in documents from numerous departments in the federal government. We hope to obtain and utilize these complete collections in future work. Many countries around the world also host enormous archives of government documents. These large collections of raw data, combined with the methods outlined in this paper, could be instrumental to creating novel time-series data that would allow us to develop and test an improved model to forecast inter-state events, while also finally permitting systematic tests of foundational theories in IR.

For example, one could additionally track threat perception as captured by the public media through printed periodicals. This would open an avenue to studying theories of elite cues (Zaller 1991; Watts et al. 1999; Gabel and Scheve 2007) and media indexing (Bennett 1990; Cook 1994), and the general interplay between these three potential pillars of foreign policy-making. Both convergence and



divergence, and understanding the conditions in which one or either occurs, would be an innovative accomplishment.

Threat perception data can also help empirically test formal models of diplomacy that focus on explicit threats. Many scholars have written on the rationale for public versus private diplomacy, and the potential differences each channel has on the ability to signal, coerce, and strike deals. A quantified measure of threat perception, matched with a source like the Military Compellent Threat dataset, could investigate the effectiveness and credibility of threats against the United States, or test competing versions of the crisis bargaining model.

Using computational techniques that have gained traction in other disciplines, IR scholars now have a tremendous opportunity to extend the field's knowledge in both novel and historically-grounded territory. Not to do so would be an equally tremendous loss. Our paper is an initial contribution and piece of evidence for this endeavor.

## 6 References

- Azar, Edward E. 1980. "The Conflict and Peace Data Bank (COPDAB) Project." *Journal of Conflict Resolution* 24(1): 143-152.
- Bennett, W. Lance. 1990. "Toward a Theory of Press-State Relations in the United States." *Journal of Communication* 40 (2): 103-127.
- Blechman, Barry M., Stephen S. Kaplan. 1978. *Force without War: U.S. Armed Forces as Political Instrument*. Washington: Brookings Institution.
- Bond, Doug, Joe Bond, Churl Oh, J. Craig Jenkins, and Charles L. Taylor. 2003. "Integrated Data for Events Analysis (IDEA): An Event Typology for Automated Events Data Development." *Journal of Peace Research* 40(6): 733-745.
- Brandt, Patrick T., Michael Colaresi, and John R. Freeman. 2006. "The Dynamics of Reciprocity, Accountability, and Credibility." *Journal of Conflict Resolution* 52 (3): 343-374.
- Brecher, Michael and Jonathan Wilkenfield. 2000. *A Study of Crisis*. Ann Arbor: University of Michigan Press.
- Breiman, Leo. 2001a. "Statistical Modeling: The Two Cultures." *Statistical Science* 16 (3): 199-231.
- Breiman, Leo. 2001b. "Random Forests." *Machine Learning* 45 (1): 5-32.
- Bueno de Mesquita, Bruce. 1975. "Measuring Systemic Polarity." *Journal of Conflict Resolution* 19 (2): 187-216.

- Buzan, Barry. 1983. *People, States, and Fear: The National Security Problem in International Relations*. Chapel Hill, NC: University of North Carolina Press.
- Chawla, Nitesh V. 2005. "Data Mining for Imbalanced Datasets: An Overview." In *Data Mining and Knowledge Discovery Handbook*, eds. Oded Maimon and Lior Rokach. New York: Springer.
- Chen, Chao, Andy Liaw, and Leo Breiman. 2004. "Using Random Forest to Learn Imbalanced Data." Technical Report 666, University of California, Berkeley.
- Cook, Timothy E. 1994. "Domesticating a Crisis: Washington Newsbeats and Network News after the Iraqi Invasion of Kuwait." In *Taken by Storm: The Media, Public Opinion, and U.S. Foreign Policy in the Gulf War*, eds. W. Lance Bennett and David L. Paletz. Chicago: University of Chicago Press.
- Dafoe, Allan, Jonathan Renshon, and Paul Huth. 2014. "Reputation and Status as Motives for War." *Annual Review of Political Science* 17: 371-393.
- Downes, Alexander B. and Todd S. Sechser. 2012. "The Illusion of Democratic Credibility." *International Organization* 66 (3): 457-489.
- Fearon, James D. 1994. "Domestic Political Audience Costs and the Escalation of International Disputes." *American Political Science Review* 88 (3): 577-592.
- Fordham, Benjamin. 1998. "The Politics of Threat Perception and the Use of Force: A Political Economy Model of U.S. Uses of Force, 1949-1994." *International Studies Quarterly* 42 (3): 567-590.
- Freeman, John R. 1989. "Systematic Sampling, Temporal Aggregation, and the Study of Political Relationships." *Political Analysis* 1: 61-98.
- Gabel, Matthew and Kenneth Scheve. 2007. "Estimating the Effect of Elite Communications on Public Opinion Using Instrumental Variables." *American Journal of Political Science* 51 (4): 1013-1028.
- Gartzke, Erik. 2006. *The Affinity of Nations Index, 1946-2002*. Dataset.
- Goldstein, Joshua S. and John R. Freeman. 1990. *Three-Way Street: Strategic Reciprocity in World Politics*. Chicago: University of Chicago Press.
- Goldstein, Joshua S. 1991. "Reciprocity in Superpower Relations: An Empirical Analysis." *International Studies Quarterly* 35 (2): 195-209.
- Goldstein, Joshua S. 1992. "A Conflict-Cooperation Scale for WEIS Events Data." *Journal of Conflict Resolution* 36(2): 369-385.
- Goldstein, Joshua S., Jon C. Pevehouse, Deborah J. Gerner, and Shibley Telhami. 2001. "Reciprocity, Triangularity, and Cooperation in the Middle East, 1979-97." *Journal of Conflict Resolution* 45(5): 594-620.

- Guo, Shuixia, Anil K. Seth, Keith M. Kendrick, Cong Zhou, and Jianfeng Feng. 2008. "Partial Granger causality—Eliminating exogenous inputs and latent variables." *Journal of Neuroscience Methods* 172: 79-93.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.
- Holsti, Ole R., Robert C. North, and Richard A. Brody. 1968. "Perception and Action in the 1914 Crisis." From *Quantitative International Politics: Insights and Evidence*, ed. J. David Singer. New York: The Free Press.
- Jervis, Robert. 1985. "Perceiving and Coping with Threat." In *Psychology and Deterrence*, eds. R. Jervis, R.N. Lebow, and J. Stein. Baltimore: Johns Hopkins University Press.
- Jervis, Robert. 1976. *Perception and Misperception in International Relations*. Princeton: Princeton University Press.
- Jo, Dong-Joon and Erik Gartzke. 2007. "Determinants of Nuclear Weapons Proliferation." *Journal of Conflict Resolution* 51 (1): 167-194.
- Kendall, M. G. 1938. "A New Measure of Rank Correlation." *Biometrika* 30 (1/2): 81-93.
- Koprinska, Irena, Josiah Poon, James Clark, and Jason Chan. 2007. "Learning to classify e-mail." *Information Sciences* 177 (10): 2,167-2,187.
- Kuhn, Max. 2008. "Building Predictive Models in R Using the `caret` Package." *Journal of Statistical Software* 28 (5): 1-26.
- Kurizaki, Shuhei. 2007. "Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy." *American Political Science Review* 101 (3): 543-558.
- Leventoğlu, Bahar and Ahmer Tarar. 2005. "Pre-negotiation Public Commitment in Domestic and International Bargaining." *American Political Science Review* 99 (3): 419-433.
- McClellan, Charles. 1999. *World Event/Interaction Survey (WEIS) Project, 1966-1978*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- McGinnis, Michael D. and John T. Williams. 1989. "Change and Stability in Superpower Rivalry." *American Political Science Review* 83 (4): 1101-1123.
- Meernik, James. 1994. "Presidential Decision Making and the Political Use of Military Force." *International Studies Quarterly* 38 (1): 121-138.
- Press, Daryl G. 2005. *Calculating Credibility: How Leaders Assess Military Threats*. Ithaca, NY: Cornell University Press.
- Ramsay, Kristopher W. 2011. "Cheap Talk Diplomacy, Voluntary Negotiations, and Variable Bargaining Power." *International Studies Quarterly* 55: 1003-1023.

- Rousseau, David L. and Rocio Garcia-Retamero. 2007. "Identity, Power, and Threat Perception: A Cross-National Experimental Study." *Journal of Conflict Resolution* 51 (5): 744-771.
- Sartori, Anne E. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 56 (1): 126-149.
- Sartori, Anne E. 2005. *Deterrence by Diplomacy*. Princeton: Princeton University Press.
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.
- Schelling, Thomas C. 1966. *Arms and Influence*. New Haven, CT: Yale University Press.
- Schmueli, Galit. 2010. "To explain or to predict?" *Statistical Science* 25 (3): 289-310.
- Schrodt, Philip A. 1995. "Event Data in Foreign Policy Analysis." In *Foreign Policy Analysis: Continuity and Change in Its Second Generation*, eds. Laura Neack, Patrick J. Haney, and Jeanne A.K. Hey. New York: Prentice Hall.
- Schrodt, Philip A. 2012. *CAMEO: Conflict and Mediation Event Observations Event and Actor Codebook*. <http://data.gdeltproject.org/documentation/CAMEO.Manual.1.1b3.pdf>.
- Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy*. New York: Cambridge University Press.
- Sechser, Todd S. 2011. "Militarized Compellent Threats, 1918-2011." *Conflict Management and Peace Science* 28 (4): 377-401.
- Shellman, Stephen M. 2004. "Time Series Intervals and Statistical Inference: The Effects of Temporal Aggregation on Event Data Analysis." *Political Analysis* 12(1): 97-104.
- 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): 115-144.
- Singer, J. David. 1958. "Threat-perception and the armament-tension dilemma." *Journal of Conflict* 2 (1): 90-105.
- Stein, Janice G. 2013. "Threat Perception in International Relations." In *Oxford Handbook of Political Psychology, Second Edition*, eds. Leonie Huddy, David O. Sears, and Jack S. Levy. Oxford: Oxford University Press.
- Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin, and Achim Zeileis. 2008. "Conditional variable importance for random forests." *BMC Bioinformatics* 9: 307-317.
- Thomas, G. Dale. 2014. "Minimizing the Effects of Temporal Aggregation on Event Data Analysis." *International Interactions* 40: 837-852.
- Tomz, Michael. 2007. "Audience Costs in International Relations: An Experimental Approach." *International Organization* 61 (4), 821-840.

- Trager, Robert F. 2010. "Diplomatic Calculus in Anarchy." *American Political Science Review* 104 (2): 347-368.
- Trager, Robert F. 2012. *Diplomacy in Anarchy: The Construction and Consequences of Perceptions of Intentions*. Book manuscript; currently titled *Diplomacy: How Adversaries Communicate and Signal Intention*, under contract with Cambridge University Press.
- Walt, Stephen M. 1987. *The Origins of Alliances*. Ithaca, NY: Cornell University Press.
- Watts, Mark D., David Domke, Dhavan V. Shah, and David P. Fan. 1999. "Elite Cues and Media Bias in Presidential Campaigns: Explaining Public Perceptions of a Liberal Press." *Communication Research* 26 (2): 144-175.
- Weeks, Jessica L. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62 (1): 35-64.
- Weiss, Gary M. and Foster Provost. 2003. "Learning When Training Data are Costly: The Effect of Class Distribution on Tree Induction." *Journal of Artificial Intelligence Research* 19: 315-354.
- Williams, John T. and Michael D. McGinnis. 1988. "Sophisticated Reaction in the U.S.-Soviet Arms Race: Evidence of Rational Expectations." *American Journal of Political Science* 32 (4): 968-995.
- Wolfe, Alan. 1979. *The Rise and Fall of the 'Soviet Threat'*. Washington, D.C.: Institute for Policy Studies.
- Yarhi-Milo, Keren. 2013. "In the Eye of the Beholder: How Leaders and Intelligence Communities Assess the Intentions of Adversaries." *International Security* 38 (1): 7-51.
- Zaller, John. "Elite Leadership of Mass Opinion." In *Taken by Storm: The Media, Public Opinion, and U.S. Foreign Policy in the Gulf War*, eds. W. Lance Bennett and David L. Paletz. Chicago: University of Chicago Press.
- Zinnes, Dina A. 1968. "The Expression and Perception of Hostility in Prewar Crisis: 1914." From *Quantitative International Politics: Insights and Evidence*, ed. J. David Singer. New York: The Free Press.