



Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research

SEAN P. O'BRIEN

Defense Advanced Research Projects Agency

Military planners and other decision makers require advanced early warning of impending crises so they can devise effective mitigation plans, mobilize resources, and coordinate responses with their foreign counterparts. Over the last 40 years, the US government has invested generously in several attempts to build crisis forecasting systems that were analytically defensible and capable of processing and making sense of vast amounts of information in real or near real time. This article describes the most recent attempt by the US military to develop an Integrated Crisis Early Warning System (ICEWS). Although ICEWS relies heavily on social science theories, data, and methods, our experiences thus far reveal some strengths and limitations of contemporary quantitative approaches to addressing social science questions with real world implications. The article concludes with a sketch of a new paradigmatic approach—a Computational Social Science Experimentation Proving Ground—that could not only improve crisis early warning and response, but also revolutionize how social science knowledge is developed, evaluated, and applied more broadly.

For more than 40 years, the US government has invested in research to develop crisis early warning systems that could inform decisions on how to allocate resources to mitigate them. Some of these efforts are reported to have had real world impacts.¹ The output of Andriole and Young's (1977) DARPA-funded crisis early warning system is reported to have been published in President Reagan's daily briefing book.² Bueno de Mesquita's expected utility model (called Policon), which models the dynamics in local, national, and international contexts (see Bueno de Mesquita 1981; Bueno de Mesquita, Newman, and Rabushka 1985; Bueno de Mesquita and Stockman 1994), is reported by Feder (1995, 2002)³ to have been applied by the CIA to accurately forecast thousands of decisions made by foreign leaders, including:

¹The opinions and interpretations expressed herein are mine alone and are not necessarily shared by the US Department of Defense or anyone associated with any official agency of the US government.

²This according to Stephen J. Andriole's webpage <http://www11.homepage.villanova.edu/stephen.andriole/> (April 21, 2009).

³Feder reports that a declassified CIA study concluded that Bueno de Mesquita's Policon model was correct in over 90% of the real world applications for which the CIA used it. Furthermore, in every case in which the forecasts generated by Policon differed from the forecasts of the subject matter experts who provided the input data, it was the Policon forecasts that proved to be correct. Bueno de Mesquita (2009) recently acknowledged that he himself is not exactly sure how to interpret this accuracy claim since most of the reported assessments he has seen were not explicit about how accuracy was measured.

- What policy is Egypt likely to adopt toward Israel?
- How fully will France participate in the Strategic Defense Initiative?
- What is the Philippines likely to do about US bases?
- What policy will Beijing adopt toward Taiwan's role in the Asian Development Bank?

A contemporary version of Policon (called Senturion) was reportedly used by the Department of Defense (See Abdollahian, Baranick, Efird, and Kugler 2006) to accurately forecast developments in the aftermath of Operation Iraqi Freedom including, among others:

- The lack of consensus around Iraqi government institutions
- The deteriorating disposition of Iraqis toward US forces
- Al Sadr's active opposition toward US interests
- How removing Saddam Hussein would cause more vigorous opposition from Shia parties

The rationale behind US interest and investment in crisis early warning capabilities is straightforward: the 2005 *US National Defense Strategy* (Department of Defense 2005), for instance, recognized that it is easier to influence global events in their earliest stages, before they become more threatening and less manageable. Furthermore, in an era of increasing budgetary pressures, military commanders have incentives to develop rigorous methodologies to assist them in proactively managing and justifying explicit tradeoffs in the allocation of resources designed to help stabilize the countries in their areas of responsibility around the world.

To be effective, a crisis early warning system should ideally meet several criteria:

1. *Accuracy and precision.* Forecasts should have documented accuracy and precision in terms of the nature of the expected event, the location of its occurrence, and its timing. In general, there is a tradeoff among the factors specifying the timing, location, and precision with which the event is identified and measured. All else being equal, it becomes much more difficult to achieve accuracy and precision targets as the unit of analysis is reduced to a specific event occurring at a specific time and place (that is, point prediction). Thus, an interpretation of the "goodness" with which a system generates accurate and precise forecasts must factor in the length of the forecast horizon, its geo-specificity, and the specific or general nature of the phenomenon that is expected to occur (for example, a bombing of a particular government institution vs. anti-government violence).
2. *Generalizable crisis antecedents.* A scientific approach to crisis early warning demands a systematic search for those conditions, events, and circumstances that consistently precede or occur in association with specific kinds of crises. Ideally, these antecedents or crisis drivers are identifiable in any particular crisis forecast, and their relationships and causal dynamics are clearly elucidated.
3. *A robust decision support system* is needed to evaluate the range of different configurations of resources that could be brought to bear to mitigate the crisis, through their estimated effects on the crisis antecedents. Forecasts of impending crises alone are insufficient; decision makers require informed insights into how the options at their disposal might mitigate, or even exacerbate, the crisis.
4. Forecasts should support the needs of *multiple constituencies*, from resource planners who have a long-term planning horizon, to intelligence

personnel who continuously monitor events and keep senior decision makers apprised of any dramatic shifts that might occur in highly fluid situations.

This article summarizes the objectives and emerging results of the most recent attempt by the US military to develop a crisis early warning system that is intended to meet the criteria enumerated above. Space restrictions prohibit a detailed description and defense of the methods used to generate the results and findings described herein. I will leave it to the university-based scholars⁴ and industry partners involved in the project to provide that accounting in future publications. My more limited aim is to describe the objectives of the Integrated Crisis Early Warning System (ICEWS) program, its major accomplishments to date, and the lessons learned about both the strengths and limitations of contemporary quantitative approaches to crisis early warning. I close out the article with a sketch of a new paradigmatic approach for synthesizing, formalizing, and testing the full range of social science theories for anticipating and responding to adverse societal conflagrations.

The Integrated Crisis Early Warning System

Program Goals

The objective of the ICEWS program is to develop a comprehensive, integrated, automated, generalizable, and validated system to monitor, assess, and forecast national, sub-national, and international crises in a way that supports decisions on how to allocate resources to mitigate them. Specifically, ICEWS is intended to provide military commanders with answers to three crucial questions:

1. Which countries in a commander's Area of Responsibility (AOR) are likely to become more or less unstable in the near-, mid-, and long-term?
2. What are the factors that are driving instability?
3. Given the array of national security resources across the entire Diplomatic, Information, Military, Economic (DIME) spectrum, what combinations of strategies, tactics, and resources are likely to have the greatest positive impact on mitigating the instability?

In Phase 1 of the ICEWS program (concluded in December 2008), several teams comprised of experts from industry and academia were provided funding to develop, test, and evaluate computational social science models for forecasting various forms of country-specific instability. The modeling capabilities were evaluated at the end of Phase 1 based on their ability to retrospectively "forecast" several classes of events often associated with (and consequences of) country instability. In Phase 2 (beginning March 2009), the ICEWS program will transition from retrospective forecasting to near-real time forecasting of a variety of Events of Interest (EoIs), which are defined below. An additional key objective of Phase 2 involves developing a capability to generate robust DIME strategies and resource packages that could be applied to mitigate particular configurations of factors driving the instability. Phase 3 will involve a live, in-theater test of the system's ability to generate robust solutions to fulfill Combatant Command stability objectives in both resource-constrained and unconstrained environments.

⁴They include Barry Silverman and Ian Lustick (University of Pennsylvania), Philip Schrodt (University of Kansas), Stephen Shellman (College of William and Mary), and Michael Ward (University of Washington).

Technical Approach

The overarching technical goal of the program is to automatically monitor, assess, and forecast the consequences of national and sub-national events and interactions that could affect US national security interests, and inform decisions on how to allocate DIME (diplomatic, information, military, and economic) resources to mitigate them. The tools and methodologies developed in ICEWS are designed to allow users to:

- Account for the complexity of interactions between governments and government institutions, the people they govern (or claim to govern), and nonstate actors such as al-Qaeda and other similar groups that are not tied to any specific geographic location.
- Identify the *generalizable* patterns in these interactions (that is, “early warning indicators”) that allow users to estimate with a high degree of accuracy the probability that an insurgency will develop, a civil war will occur, one or more countries will attack another with military force, or a military coup will be hatched to dispatch a current set of rulers, to name but a few examples.
- Determine how a country’s macro-structural conditions (political, social, demographic, economic) affect the way in which the country’s citizens interact with their government; determine how these conditions enable or constrain the way in which the country’s leadership interacts with its people, in addition to other governments.
- Identify whether there are certain characteristics of a government’s leaders that are particularly telling about their propensity to defuse or exacerbate potentially volatile situations.
- Leverage the latest information processing technologies that can capture and process vast quantities of data from digitized news media, web sites, blogs, and other sources of information that reflect the dynamic and rapidly changing character and intensity of interactions between people and governments.

Program Scope

The goal of ICEWS is to cover the broadest possible spectrum of events encompassing instability and political violence. Toward that end, each ICEWS performer was tested on its ability to forecast the following discrete EoIs:⁵

- *Domestic Political Crisis*—Significant opposition to the government, but not to the level of rebellion or insurgency (for example, power struggle between two political factions involving disruptive strikes or violent clashes between supporters)
- *Rebellion*—Organized opposition where the objective is to seek autonomy or independence
- *Insurgency*—Organized opposition where the objective is to overthrow the central government
- *Ethnic/Religious Violence*—Violence between ethnic or religious groups that is not specifically directed against the government
- *International Crisis*—Conflict between two or more states or elevated tensions between two or more states that could lead to conflict

⁵We also collected data on major instances of government repression, economic collapses, regime changes, and civil wars. But given our geographic (Asia-Pacific region) and temporal (1998–2006) scope, there were too few instances of these EoIs so they were excluded from further analyses.

Historical data on these events were collected on a quarterly basis from a variety of open sources.⁶ The data were collected for 28 major countries in the US Pacific Command (USPACOM) AOR.⁷ This AOR ranges from Mongolia and northern China to the north, western China and India to the west and the Pacific Ocean spanning beyond Hawaii to the east and Australia and New Zealand to the south. The USPACOM AOR was chosen as an initial focus area because it includes a mix of very stable, semi-stable, and highly unstable countries, providing us with elements of both the most similar and most different research designs (Przeworski and Teune 1970). Because one of the objectives of ICEWS is fielding a *generalizable* capability to forecast progress toward or away from stability, a viable solution involves the ability to forecast country instability using a consistent set of factors applicable to all countries in the PACOM AOR. A portion of the historical data (1998–2004) was provided to the ICEWS performers to use as training in model development; the remaining data (2005–2006) were withheld for testing the modeling solutions according to the following performance metrics:

- Accuracy = $\frac{\# \text{ of correct predictions}}{\# \text{ of predictions made}}$
- Recall = $\frac{\# \text{ of correctly predicted conflicts}}{\# \text{ of conflicts that occurred}}$
- Precision = $\frac{\# \text{ of correctly predicted conflicts}}{\# \text{ of conflicts predicted to occur}}$

The split-sample testing protocol, along with the decision rules governing probability thresholds, followed the procedures described in O'Brien (2002). In addition, to impose some consistency on the overall evaluation, each performing team was evaluated based on its ability to forecast the instability index for each country (none/low, moderate, or high intensity level of instability). This typology, depicted in Figure 1, defines an aggregate index of instability. This index was used to establish a baseline level of stability against which progress toward stability goals in the AOR could be assessed. The index definitions as well as reference data were derived from reports of the Conflict Barometer, published annually by the Heidelberg Institute of International Conflict Research.⁸

⁶Heidelberg Institute for International Conflict Research <<http://www.hiik.de/start/index.html>> (accessed May 12, 2009); Political Instability Task Force <<http://globalpolicy.gmu.edu/pitf/>> (accessed May 12, 2009); Armed Conflict Database, <<http://www.iiss.org/publications/armed-conflict-database/>> (May 12, 2009); Reuters data provided by Virtual Research Associates (VRA) <<http://www.vranet.com/index.html>> (accessed May 12, 2009); MIPT Terrorism Knowledge Base <<http://www.tkb.org/Home2.jsp>> (May 12, 2009); Global Security Data <<http://www.globalsecurity.org/>> (accessed May 12, 2009); BBC, International Herald Tribune, CNN news archives, and other news websites as needed.

⁷The data were collected by Strategic Analysis, Inc. under US government contract. The effort was led Drs. Philippe Loustaunau and Evelyn Dahm. The top-level EoIs were extracted from the textual descriptions in the PITF, HIIK, and ACD databases. The Reuters event data provided by Doug Bond at VRA were used to verify the EoIs and identify missing EoIs. The remaining sources were used to verify beginning and end dates to the extent possible. The ICEWS teams were provided at the beginning of Phase 1 with the training data and EoI definitions, but not the list of sources from which the data were extracted. Doing so might have called into question the validity of the ultimate out of sample test and evaluation. However, the teams were given 90 days to review and comment on the consistency and accuracy of the training data. The review identified some miscodings (that is, rebellions coded as insurgencies or vice versa), which were corrected in the test data and the final version of the training data provided to the ICEWS teams.

⁸See <http://www.hiik.de/en/index_e.htm> (accessed June 8, 2009). The original index contained two additional categories—latent and manifest conflicts—that were collapsed into the no-conflict category because their intensities did not rise to a level that was deemed significant or appropriate given our objectives.

None/low intensity	1— No Conflict	No crisis has occurred
	2 – Crisis	A crisis is a tense situation in which at least one of the parties uses violent force in sporadic instances
Moderate Intensity	3 – Violent crisis	A conflict is a severe crisis if violent force is repeatedly used in an organized way
High intensity	4 – War	A war is a type of violent conflict in which violent force is used with a certain continuity in an organized and systematic way. The conflict parties exercise extensive measures, depending on the situation. The extend of destruction is massive and of long duration

FIG 1. Index of Instability, modified from the Heidelberg Institute of International Conflict Research
Notes. If the combined probability of conflict types 1 and 2 occurring in a given country-quarter is greater than 67%, then the expectation is that the country will experience no or a low intensity level of instability. If the combined probability of conflict types 2 and 3 occurring in a given country-quarter is greater than 67%, then the expectation is that the country will experience a moderate intensity level of instability. If the combined probability of conflict types 3 and 4 occurring is greater than 67%, then the expectation is that the country will experience a high intensity level of instability.

Modeling Approaches

Scholars typically approach crisis early warning from one of several different perspectives. The different perspectives can be distinguished by the types of factors they consider, and their corresponding forecast horizons. Macro-structural forecasting models examine the relationship between country instability and broad trends in political, social, economic, and demographic factors. Together, these factors describe an environment that makes a country more or less susceptible to various forms of instability, by enabling or constraining the options available to its political leaders in their ability to respond to a variety of internal and external challenges. Because these macro-structural indicators change slowly, and often along predictable trend lines, the macro-structural forecast horizon is typically 2–5 years (Esty, Gladstone, Robert Gurr, Surko, and Unger 1995; Esty, Gladstone, Robert Gurr, Harff, Levy, Dabelko, Surko, and Unger 1998; O'Brien 2002).

A second approach involves searching for patterns in how the character, intensity, and sequence of interactions between or within states, portend future conflict events. Schrodtt (Schrodtt, Simpson, and Gerner 2001) and others (Bond, Jenkins, and Schock 1997; Shellman 2004) have developed and matured techniques to automatically parse and convert digital news reports into structured indices that reflect the character and intensity of interactions between key leaders, organizations, and countries—who is doing *what* to *whom*, *when*, *where* and *how* around the world—based in large measure on schemas first developed by McClelland (McClelland and Hoggard 1969). These behavioral event data have been leveraged in several successful early warning projects (Schrodtt 1997, 1998; Schrodtt and Gerner 1997; Pevehouse and Goldstein 1999). Because of the highly dynamic nature of these data, and the timeliness with which they can now be maintained, the forecast window is often weeks or months ahead.

Still others have highlighted the importance of decision makers themselves, and how their personalities, leadership styles, and operational codes influence how they are likely to respond to internal and external pressures that could culminate into serious conflicts. Within this third perspective, Hermann (1999) has developed automated content analysis tools to extract, in near real time from spontaneous speech transcripts, indicators measuring a range of theoretically relevant leadership trait characteristics (see also Walker 2000; Lazarevska and Sholl

2005). The presence, absence, intensity and configuration of these characteristics, provide a profile of a leader, along with a set of theoretical expectations for how the leader will approach decisions, relate to other leaders, or behave in certain circumstances.

The goal of ICEWS was to bring together the best modeling approaches from each of these perspectives to evaluate the extent to which their integration could allow for generating forecasts that were more accurate, precise, and actionable than any one of the modeling approaches alone could produce. As discussed below, that hypothesis was largely confirmed. The most successful team was led by Lockheed Martin-Advanced Technology Laboratories (LM-ATL), assisted by several prominent scholars and industry partners. The team successfully integrated and applied six different conflict modeling systems, including the following:

- Agent-based models, grounded in theory and parameterized with data about leaders and societies, are ideally suited for evaluating the causal dynamics associated with multiple potential futures, including the actual future a country experienced. ICEWS used Barry Silverman's *Factionsim* (Silverman, Bharathy, Nye, and Smith 2008)⁹ and Ian Lustick's *Political Science-Identify (PS-I)* computational modeling platforms (Lustick, Miodownik, and Eidelson 2004). *Factionsim* uses surveys of Subject Matter Experts (SMEs) to develop highly detailed profiles of archetypical leaders and followers for government and non-governmental groups within each country. PS-I models were created with agents representing population elements of various ethnic/political identities organized geographically and in authority structures designed to mirror the society being studied. Data on these factors were also elicited from SMEs with extensive knowledge of each country under examination.
- Phil Schrodt and Steve Shellman (Shellman 2008; Shellman, Hatfield, and Mills 2010) developed separate logistic regression models that used macro-structural and event data factors commonly analyzed in the academic literature, including regime type, GDP per capita, and indices reflecting the degree of cooperation and hostility between government and civil society actors, among many others. Shellman evaluated a similar set of factors using a Bayesian statistics model.
- Michael Ward built geo-spatial network models that used structural factors, event counts, and various types of spatial networks—trade ties, people flows, and “social similarity” profiles—that embody potential EOI co-dependencies between proximate countries (Ward and Gleditsch 2002; Hoff and Ward 2004).
- A final model was developed by aggregating the forecasts from the abovementioned models using Bayesian techniques.¹⁰ This involved creating a Bayesian network that combined evidence (that is, forecasts) from all the models with prior probability estimates for each of the EoIs to produce posterior probabilities for the EoIs. The network consisted of three types of nodes: EoI nodes; EoI estimate nodes, one for each model's forecast of an EoI; context nodes to infer priors for the EoI nodes. A Bayesian network consists of a structure representing the conditional relationships among the nodes and parameters encoding the strength of those relationships (Pearl 1988). Structure learning based upon mutual information scores was used to discover the relationships

⁹See Silverman's website http://works.bepress.com/barry_silverman/ (accessed May 5, 2009) for additional papers describing *Factionsim*.

¹⁰The Bayesian Aggregator model was developed by Dr. Suzanne Mahoney from Innovative Decisions, Inc.

between the context nodes and the EoI nodes (Cheng, Bell, and Liu 1998). EoI estimate nodes were assumed to be conditionally independent given the EoI. For each of these, a conditional probability table representing the probability of a model's forecast given the ground truth for the EoI was learned. Discretization learning partitioned training data forecasts into bins that effectively recalibrated the forecasts (Yang and Webb 2002). Then standard parameter learning was used to fill in the parameters for both the context and EoI estimate nodes (Neapolitan 2003). EoI posteriors given country-specific context and EoI estimates were inferred using a common inference algorithm (Lauritzen and Spiegelhalter 1988). The idea here is to ascribe more or less confidence to any individual model forecast, according to the countries and EoIs for which that model has demonstrated high performance.

Input Data

In addition to the SME data used by the agent-based models, the ICEWS performers used input data from a variety of sources. Notably, they collected 6.5 million news stories about countries in the Pacific Command (PACOM) AOR for the period 1998–2006. This resulted in a dataset about two orders of magnitude greater than any other with which we are aware. These stories comprise 253 million lines of text and came from over 75 international sources (AP, UPI, and BBC Monitor) as well as regional sources (*India Today*, *Jakarta Post*, *Pakistan Newswire*, and *Saigon Times*). News stories were complemented by country data information, including data from the Economist Intelligence Unit, Freedom House, International Monetary Fund (IMF), World Bank, Political Instability Task Force (PITF), and the Correlates of War project (COW). Using an actor dictionary (with over 8,000 entries) and a verb-phrase dictionary (with over 15,000 entries), news stories were coded using Schrod's TABARI¹¹ event data coding system in four major categories—verbal cooperation/conflict, material cooperation/conflict—comprising 130 variables to measure and monitor the character and intensity of a broad range of political activities. The TABARI system automatically parses and converts these news reports into structured indices that describe *who* is doing *what* to *whom*, *when*, *where* and *how* around the world.

Results

The LM-ATL ICEWS team was allowed to train its models using the historical data for 1998–2004. The out of sample performance was assessed using the data withheld for the period 2005–2006. The team was required to generate eight quarterly forecasts, using data no more current than the quarter immediately preceding the quarterly forecast horizon. The results from the aggregated forecasts are shown in Figure 2.

The minimum performance metrics were set at 80% accuracy and recall, and 70% precision. These figures are a rough average of the results reported in O'Brien (2002), which served as the ICEWS program's benchmark study. The ICEWS team surpassed this benchmark in forecasting the Heidelberg index of instability and three of the discrete EoIs: rebellion, insurgency, and ethnic/religious violence. The disappointing results for the domestic political crisis EoI may be driven in part by the fact that these crises are less intense, less violent forms of both rebellions and ethnic/religious violence. In other words, rebellions

¹¹See <http://raven.cc.ukans.edu/~keds/index.html> (accessed May 5, 2009) for information on the KEDS and TABARI projects.

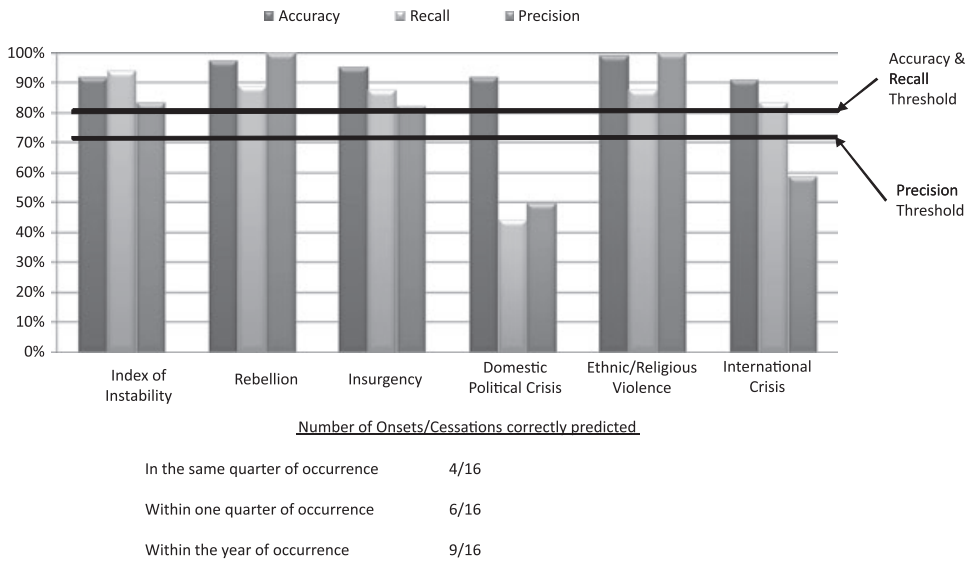


FIG 2. Retrospective, Out of Sample Performance Metrics, 2005–2006 (ICEWS Aggregate Model)

and ethnic/religious violence often begin as domestic political crises, and we have yet to identify the factors that would allow us to detect these events in their earlier stages. The poor performance on international crises may at least in part be attributed to our country-quarter research design.

Figure 2 also reveals a need to improve our ability to accurately forecast both the onset and cessation of EoIs. There were 16 new EoI onsets and cessations in the 2005–2006 test set, roughly equally distributed across the six EoIs. Of these 16, we correctly forecast only 25% to occur or end in the exact quarter in which they actually occurred or ended. If we use an annualized forecast horizon, as is standard in most academic research, that figure more than doubles to 56% (9/16). These relatively poor performance metrics may be due in part to the many restrictions we placed on the ICEWS modelers in our effort to represent as closely as possible the operational environment in which ICEWS is designed to ultimately be applied. For instance, as in O'Brien (2002), we used a high probability threshold for generating a forecast (67%), in contrast to the 50% cutoff most commonly used in academic studies. We did so because we believe that a forecast that is intended to be taken seriously is one that demands a level of confidence greater than that of a coin toss. We also prohibited the use of dependent variable lags as well as variables measuring some aspect of the country's conflict history—standard methodological practices and measures in most studies—in a conscious effort to avoid inflating performance metrics, a point on which I elaborate below.

For purposes of comparison, Figure 3 depicts the results of taking a completely naïve approach to forecasting the EoIs. It is based on projecting what occurred in each country in the fourth quarter of 2004 over the eight subsequent quarters in the 2005–2006 test set. Because the ICEWS team was required to complete the test and evaluation in a single iteration, and because the team was not allowed to use lags of the dependent variables, the results of the naïve model reported in Figure 3 are directly comparable to the results reported in Figure 2.

Although, by definition, the naïve model will not correctly forecast any onsets or cessations, it performs remarkably well in indicating the presence of some

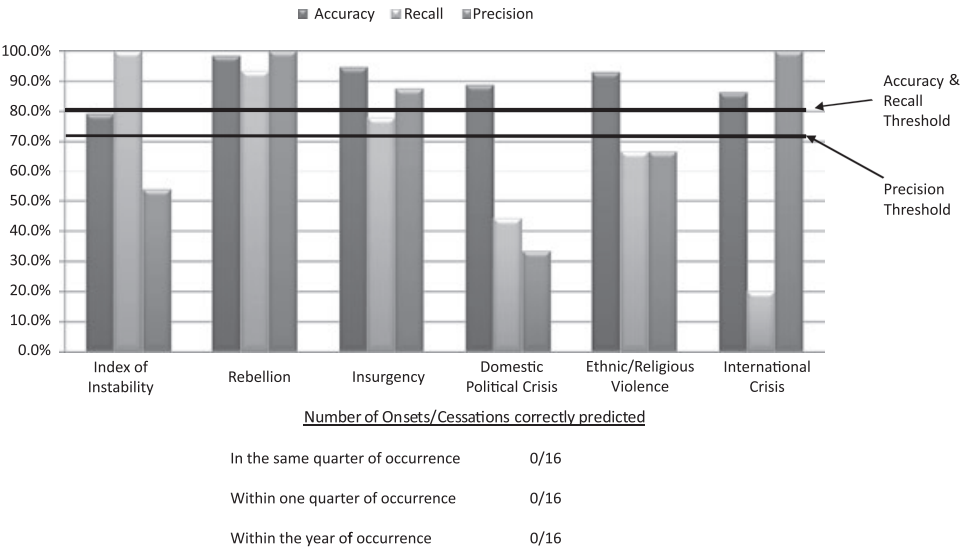


FIG 3. Retrospective, Out of Sample Performance Metrics, 2005–2006 (Naïve Model)

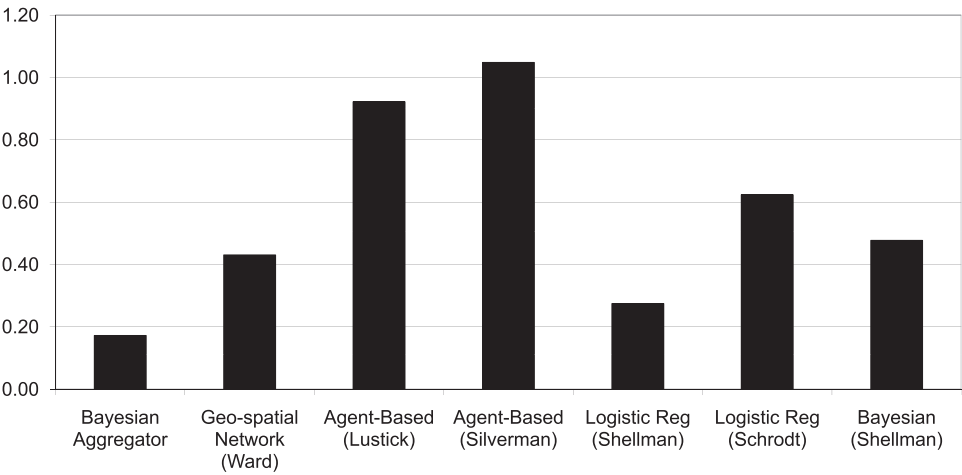


FIG 4. Brier Scores Comparing Model Performance on Rebellion EoI

Note. Models with lower Brier scores forecast with greater accuracy and precision than models with higher Brier scores.

EoIs. For instance, it surpasses the minimum benchmarks for both the rebellion and insurgency EoIs, and performs well at forecasting ethnic/religious violence. By contrast, the model performs unacceptably poor on the Heidelberg index of instability, domestic political crisis, and international crisis EoIs, the recall score on the latter EOI barely reaching 20%. In any case, the results in Figure 3 illustrate the significant impact that dependent variable lags or measures of conflict history can have on model performance metrics. Given the out of sample forecasting design used in the ICEWS program, an accurate assessment of the contribution of more substantive, policy-relevant antecedents can only be completed in the absence of these types of variables.

Figure 4 displays relative performance measures for how each model performs on forecasting the rebellion EoI. The height of each bar is the average

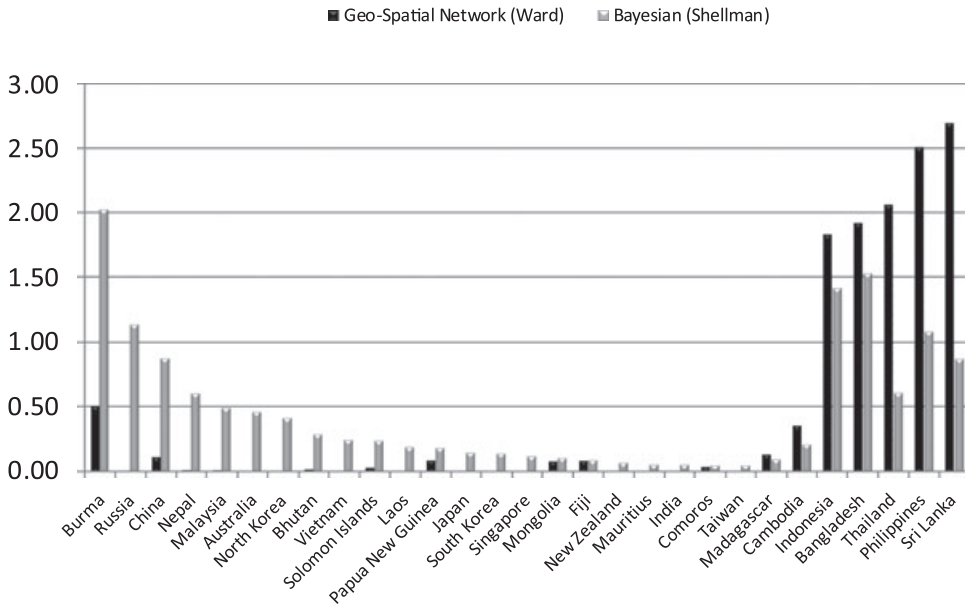


FIG 5. Brier Score Comparison between Ward and Shellman Models (Rebellion EoI)

squared deviation between the predicted probabilities generated by each model and the “ground truth” represented by the test data (Brier 1950). This difference is called the Brier score and it offers a useful way to assess the performance of models relative to one another. Models with lower Brier scores forecast with greater accuracy and precision than models with higher Brier scores. Figure 4 provides confirmatory evidence that an integrated crisis early warning model performs better than any individual approach considered here.

Some models generate forecasts that are more accurate for some EoIs and countries compared to other models. Here again, the Brier score is a useful way to conduct such an assessment. For instance, Figure 5 compares Ward’s geo-spatial model with Shellman’s Bayesian model with respect to their performance on forecasting the rebellion EoI. The height of the Brier score bars indicate that Ward’s model is performing better on countries toward the left of the figure, whereas the Bayesian model out performs the geo-spatial model for countries on the right side of the chart. The five countries located farthest to the right (Indonesia, Bangladesh, Thailand, Philippines, and Sri Lanka) were also among those most active in experiencing rebellions. A nice feature of the multi-model approach used in ICEWS is that in instances in which one model fails to correctly forecast an EoI for a particular country, it is often the case that a different model will correctly make the forecast. Again, the prior performance evaluations for each modeling approach are used by the aggregator model to weight each model’s forecast according to the countries and EoIs for which it best performs.

We are aware of no similar forecasting project that has generated quarterly-level forecasts for as many EoIs and at equivalent levels of accuracy and precision, than what we demonstrated in Phase 1 of the ICEWS program. Nevertheless, though retrospective forecasting provides us with impressions of how our forecasting models are likely to perform in the future, a true performance assessment will have to await sufficient operational tests and evaluations, which are forthcoming in subsequent phases of the program.

Lessons Learned

Integrated Crisis Early Warning System is fundamentally concerned with identifying those perhaps seemingly benign, policy relevant factors that, when combined with other factors, systematically precede crises in a probabilistic way. By policy-relevant factors, I mean factors that could conceivably be changed through government policy or intervention, regardless of whether the political will or requisite amount of resources are immediately present in any given instance. We early on discovered that we could come close to achieving our benchmark performance metrics using naïve models, which included lagged values of the EoI dependent variable, and a small number of policy-irrelevant correlates like size of population, presence or absence of mountainous terrain, and the like. Though such a naïve model may retrospectively achieve acceptable levels of overall performance, it is useless for real world applications for two reasons.

First, models that rely on dependent variable lags, as seen above, provide only an illusion of high performance or goodness of fit. A naïve model containing only lags of the dependent variable may score well on indicating the presence of some EoIs, but will miss every new onset and cessation of conflict, literally by definition. The illusory good performance metrics also operate as a disincentive to continue the search for more insightful, actionable crisis antecedents.

Second, the inclusion in forecasting models of static, policy irrelevant variables like mountainous terrain and border contiguity seems to add little value independent of the contribution they provide to model performance metrics such as R^2 or the overall level of accuracy as defined and used in this examination. The theoretical arguments justifying the explanatory power of these statistically significant but substantively less meaningful variables are often simplistic and vague. Furthermore, short of changing the entire nation-state system setup by the Treaty of Westphalia in 1648, no country or group of countries can act upon the relationship between contiguous borders and inter-state disputes by, say, relocating effected countries as if they were chess pieces on a board. Mountainous terrain seems no more susceptible to reasonable mitigation strategies, regardless of its correlation to countries that experience rebellions. Therefore, to consciously avoid exaggerating forecast model performance metrics, the performers were prohibited from using in their models dependent variable lags or variables that mattered little beyond their contribution to forecast performance metrics.

A second lesson learned is that out of sample forecasting is a much better way to assess model performance than the more common approach of using statistical significance tests. Ward, Greenhill, and Bakke (Forthcoming), who recently evaluated the predictive ability of the Fearon and Laitin (2003) and Collier and Hoeffler (2004) models of civil war, corroborate this observation. Their findings deliver a serious blow to the predominant way in which most conflict models are evaluated using statistical significance tests. For Ward et al. show that at a reasonable probability cut off (50%) these models predict few if any civil war cases with which they are concerned.¹² Surprisingly, the Fearon and Laitin model does not even appear to generate a probability greater than 30%. Given the number and length of civil wars in Africa, Asia, and elsewhere, the fact that these frequently cited models do not include enough information to be more than 30% confident a civil war might occur is cause for serious reflection upon the predominant paradigm informing quantitative conflict studies. For if it is the case, as Ward et al. seem to show, that we cannot correctly predict over 90% of the cases with which our model is concerned, then we have little basis to assert our *understanding* of a phenomenon, never mind our ability to *explain* it. Rather than relying

¹²Indeed, at the 50% cutoff the Fearon and Laitin model correctly predicts 0 of 107 civil war onsets, whereas the Collier and Hoeffler model correctly predicts 3 of 46 onsets.

on statistical significance tests to evaluate variable and model performance, a better approach might involve evaluating theoretically relevant variables according to the contribution they make to distinguishing conflict cases from non-conflict cases in out of sample forecasts, using the techniques suggested by Ward et al. King and Zeng (2001) make essentially the same case and note that predicting with high degrees of accuracy and precision is a sign that one has identified causal structure.

A final lesson learned pertains to the utility of taking a multi-method approach to crisis forecasting, that extends beyond its contribution to forecasting accuracy and precision. In particular, the agent-based model has significant potential that has yet to be fully tapped by the political science and international relations academic communities. Agent-based models generate multiple potential futures that could emerge from dynamical perturbations of any particular configuration of current societal conditions and organizational profiles. These futures allow for the examination of the causal dynamics that could generate any particular future, including the one that is ultimately realized. What is more, agent-based models—and simulations more generally—may provide a suitable, or even superior alternative to the predominant approach of testing competing theoretical claims using Large-N statistical research designs.

Many of the most interesting, policy-relevant theoretical questions are also the most complex, nonlinear, and highly context-dependent. They demand consideration of hundreds of massively interacting variables that are difficult to measure systematically and at a level of granularity consistent with the theory. In such cases it is at best impractical and at worst impossible to apply standard regression techniques within the context of a Large N study, short of invoking unreasonable, oversimplifying assumptions. This may in part account for contradictory findings in the literature relative to the validity of alternative theoretical claims.¹³

Lustick et al. (2004) provide a good case in point. The authors lament that neither Small-N nor Large-N studies have reached consensus on whether power sharing prevents or encourages secessionism in multicultural states. The authors identify three separate positions on the issue: federalism encourages nationalism; federalism reduces nationalism; federalism has no bearing on the strength of nationalism. To evaluate these outstanding theoretical claims, Lustick et al. used the agent-based PS-I system to develop Beita, a virtual multiethnic state. Beita corresponds directly with no particular country but contains features of many nation-states. It was specifically designed to possess those features of multiethnic societies that are held to be crucial by the relevant theories and so bears resemblance to those states that may be predisposed to secessionism. Through their ability to control some aspects of the experimental environment while varying others, Lustick et al. are able to perturb the system in ways consistent with the alternative competing theoretical claims. After generating multiple histories, the results are analyzed and assessed relative to the expectations of the different theories.

The results of their exercise allow for interpretations that are much more highly sophisticated and nuanced than what can be obtained from traditional Large-N regression analyses, including the following:

¹³For instance, does commercial interdependence increase (Gasirowski 1986; Barbieri 1996, 2002) or decrease (Oneal and Russett 1997; Gartzke 1998; Russett, Oneal, and Davis 1998) the likelihood of bilateral conflict? Does a country's political instability have a positive effect on (Woodward and Rolfe 1993), negative effect on (Schneider and Frey 1985), or bear no relationship (Fatehi-Sedeh and Safizadeh 1989; Olibe and Crumbley 1997) to the amount of Foreign Direct Investment (FDI) it receives? Are ethnic civil wars virtually impossible to resolve through power-sharing settlements (Kaufmann 1996) or are they no less amenable to power-sharing than other civil wars (Mason and Fett 1996)? Finally, for a review of the inconsistent findings relative to the conditions under which truth commissions emerge, and the influence they have on transitional societies, see Brahm (2009).

1. Excluding potential challengers could drive them into rebellion but it also reduces their salience in the political system.
2. Excluding potential challengers can neutralize the threat of rebellion, but only at very high levels of exclusion.
3. Including potential challengers can enhance regime stability but only if the level of inclusion is high and associated with increased salience in their political system.
4. Including potential challengers does not disrupt the regime, but it does weaken the political position of the incumbents in the political system.

A New Paradigm: Computational Social Science Experimental Proving Ground

Lustick et al. (2004) provide inspiration for a new paradigmatic approach to developing, testing, and ultimately applying social science knowledge to complex, highly nuanced issues that are important, both intellectually and from a policy perspective. I close this article with a brief sketch of what such an approach might resemble.

Lustick et al. remind us that the social science literature, political science in particular, is beset with competing theoretical claims, even on some of the most fundamental questions. For instance, does government repression (i) increase, (ii) decrease, or (iii) bear no relationship to insurgent violence? One can point to cases where heavy-handed government repression was followed by dramatic decreases in insurgent violence (Fujimori's Peru, Chechnya during Putin administration) as well as cases in which repression was followed by increases in insurgent activity (Northern Ireland, Palestine, Iraq). But neither Small-N nor Large-N studies have resolved the issue to the point where we could confidently anticipate the consequences of government repression in any particular instance. Whether repression is associated with an increase or decrease in insurgent activity is likely to depend on other factors and theoretical expectations from psychology, history, anthropology, religious studies, and economics. Yet, most social science theories are examined in intra-disciplinary isolation from other theories upon which their resolution depends. So what can we do to harvest the theoretical knowledge we have accumulated thus far, identify and expose gaps in our knowledge, and provide the incentives and tools for scholars to reorient their research agendas toward filling these gaps with alternative theoretical specifications? The answer may lie in the development of a computational social science experimental proving ground.

Whereas Lustick et al. built a virtualization of a country that corresponded with no country in particular, it is possible to imagine the development of simulations and test beds on a global scale that empirically represent the nuances of different societies, at various levels of abstraction. Scholars at Virginia Tech have already developed a 100 million agent simulation that includes synthetic versions of many American citizens, and plan to expand to 300 million agents this year (Upson 2008). Each synthetic agent has as many as 163 variables describing age, ethnicity, socio-economic status, gender, and various attitudinal factors. The simulation is used to assess how different types of pandemics could spread across the United States under different scenarios.

Having synthetic software agents that are empirical reflections of the people and organizations under examination, though an important first step, tell us nothing about how those agents are likely to react to various events under different circumstances. The comprehensive rules or laws that govern human behavioral responses to various social stimuli should be theoretically informed and empirically derived. In any case, they have yet to be discovered. However, spread throughout the social science literature are hundreds of theories (that is,

candidate rules of human behavior) that provide a set of expectations for how diverse groups and individuals, endowed the specific cognitive, demographic, and cultural characteristics, develop goals, preferences and standards of behavior; form, alter, and act upon beliefs; join or leave radical/violent organizations; and respond to events around them. Some of these theories have been tested for correspondence to the real world, but most have not because the tools of the social scientist often do not lend themselves to rigorous empirical evaluations of alternative, competing theoretical claims.

It is conceivable that we could synthesize, formalize, and semantically integrate all known social science theories—even those subject to contested claims—and instantiate them within the context of a large simulation experimental platform. This synthesized theoretical framework would need to be run across rich test beds, consisting of multiple societal representations on a scale similar to the one developed by Virginia Tech, in the United States. We would want to see, for instance, whether this integrated set of theories could account for or replicate the dynamics associated with the dissolution of Yugoslavia, or the evolution of the insurgency in post-2003 Iraq, among many other cases. This would allow us to prove or disprove alternative theories and identify the boundary conditions under which any particular theory or set of theories applies. We would also need to identify gaps, where new theories would need to be specified and tested to account for discrepancies between simulated and real world behavior.

One could imagine competitions within and between various disciplines and universities to develop and test within the experimental proving ground alternative theoretical specifications to address some of the outstanding social science “grand challenges.” These grand challenges might consist of observed but unexplainable patterns of behavior or other outstanding questions subject to competing theoretical claims. Foundations, governmental and non-governmental organizations could nominate their own grand challenge questions, which could be categorized and maintained on a project web site. Financial and prestige awards could be provided to those who discover solutions to various social science grand challenges, and demonstrate the validity of those solutions within the context of the community-owned experimental proving ground. By leveraging the entire social science community within the context of a single experimental proving ground, social science knowledge might rapidly accumulate, leading to significant breakthroughs that have eluded the field for so long. Such a bold undertaking conjures images of the Manhattan Project, but perhaps that is what the discipline needs at this time to begin maturing as a science. But only if several criteria are met is such a vision likely to be feasible and garner the necessary support of the social science research community.

First, a major investment would be required upfront to jump start the effort. This initial effort would focus on developing the simulation architecture along with a first cut, “best of breed” synthesis of some of the major theories from across the different social science disciplines. The synthesized theoretical framework would need to be formalized and tested, generating the first peer-reviewed reports of initial findings. Second, the experimental proving ground would have to be open-source, transparent, comprehensively documented, and “owned” by no one but the community of interest. Of course, an institutional entity would have to assume responsibility for maintaining the code base and test bed. One possibility is to rotate the responsibility among interested universities and laboratories. To finance the maintenance of the code base and test bed data, the responsible institution could charge a nominal fee to provide the community with access to the experimental proving ground.

Third, an Oversight Committee, consisting of professionals popularly elected from among the user community would establish standards, oversee test bed development, vet and publish grand challenge nominations, and adjudicate

competitions. The Committee might be elected to staggered terms of 3–4 years, for instance. Fourth, user-friendly tools would need to be developed to assist scholars in formalizing and connecting their theoretical specification to the overarching, validated theoretical framework. The TurboTax[®] software, which guides non-technical users through the morass of a highly complicated US tax code, is a model worth emulating. Finally, the social science academic community would need incentives to participate in such an endeavor. To the extent that such an experimentation platform could garner credibility within the discipline, social science journals, particularly the leading journals, might consider a successful proving ground test of a new theory or approach to social science questions as a necessary prerequisite for publication. The financial and prestige awards mentioned earlier would also serve as incentives.

This concept of a computational social science proving ground is entirely consistent with the movement to more closely integrate empirical and formal modeling, as exemplified by the Empirical Implications of Theoretical Models (EITM) Workshop sponsored by the Political Science Program of the National Science Foundation in 2001. The subsequent report¹⁴ bemoaned the schism that has developed between those who engage in formal modeling, which emphasizes rigorous mathematics and computational simulation, and those who focus on empirical modeling involving data analysis using statistical tools. The over-specialization of political scientists in one approach at the expense of the other has resulted in research that has formal clarity but little or no empirical tests, or in research that takes an empirical approach with no formal clarity. Bridging the gap between formalism and empiricism requires identifying or parsing out causal linkages among many factors, and the boundary conditions under which any particular relationship applies.

This is precisely the intent behind the notion of a computational social science proving ground, which is presented here primarily to stimulate discussion. Whatever new paradigm eventually emerges with consensus support, it must take seriously the need to bridge this gap between formalism and empiricism which serves as a primary impediment to the accumulation of an integrated set of social science theoretical knowledge. Only then will we be able to better understand the political and social world, generate accurate predictions about its dynamics, and support better, more informed policy decisions that can have profound and often unanticipated consequences.

References

- ABDOLLAHIAN, MARK, MICHAEL BARANICK, BRIAN EFIRD, AND JACEK KUGLER. (2006) Senturion: A Predictive Political Simulation Model. Center for Technology and National Security Policy National Defense University. <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA454175&Location=U2&doc=GetTRDoc.pdf> (Accessed May 9, 2009).
- ANDRIOLE, STEPHEN. J., AND ROBERT A. YOUNG. (1977) Toward the Development of an Integrated Crisis Warning System. *International Studies Quarterly* 21: 107–150.
- BARBIERI, KATHERINE. (1996) Economic Interdependence: A Path to Peace or Source of Interstate Conflict? *Journal of Peace Research* 33: 29–49.
- BARBIERI, KATHERINE. (2002) *The Liberal Illusion: Does Trade Promote Peace?* Ann Arbor: University of Michigan Press.
- BOND, DOUG, CRAIG JENKINS, AND KURT SCHOCK. (1997) Mapping Mass Political Conflict and Civil Society: Issues and Prospects for the Automated Development of Event Data. *Journal of Conflict Resolution* 41: 553–579.
- BRAHM, ERIC. (2009) What is a Truth Commission and Why Does it Matter? *Peace and Conflict Review* 3: 1–14.

¹⁴See <http://www.nsf.gov/sbe/ses/polisci/reports/pdf/eitmreport.pdf> (accessed June 8, 2009).

- BRIER, GLENN W. (1950) Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review* 78: 1.
- BUENO DE MESQUITA, BRUCE. (1981) *The War Trap*. New Haven, CT: Yale University Press.
- BUENO DE MESQUITA, BRUCE. (2009) A New Model for Predicting Policy Choices: Preliminary Tests. Paper Prepared for the 50th Meeting of the International Studies Association New York, February 15–18, 2009. http://www.allacademic.com/meta/p312200_index.html (Accessed June 8, 2009).
- BUENO DE MESQUITA, BRUCE, AND FRANS N. STOCKMAN, Eds. (1994) *European Community Decision Making*. New Haven, CT: Yale University Press.
- BUENO DE MESQUITA, BRUCE, DAVID NEWMAN, AND ALVIN RABUSHKA. (1985) *Forecasting Political Events: The Future of Hong Kong*. New Haven, CT: Yale University Press.
- CHENG, JIE, DAVID BELL, AND WEIRU LIU. (1998) Learning Bayesian Networks from Data: An Efficient Approach Based on Information Theory. <http://www.cs.ualberta.ca/~jcheng/bnpc.htm> (Accessed June 3, 2009).
- COLLIER, PAUL, AND ANKE HOEFFLER. (2004) Greed and Grievance in Civil War. *Oxford Economic Papers* 56: 563–595.
- DEPARTMENT OF DEFENSE. (2005) The National Defense Strategy of the United States of America. <http://www.defenselink.mil/news/Mar2005/d20050318nds1.pdf> (Accessed May 4, 2009).
- ESTY, DANIEL C., JACK GLADSTONE, TED ROBERT GURR, PAMELA SURKO, AND ALAN N. UNGER. (1995) Working Papers: State Failure Task Force Report. McLean, VA: Science Applications International Corporation. http://www.allacademic.com/meta/p312200_index.html (Accessed June 8, 2009).
- ESTY, DANIEL C., JACK GLADSTONE, TED ROBERT GURR, BARBARA HARFF, MARC LEVY, GEOFFREY D. DABELKO, PAMELA T. SURKO, AND ALAN N. UNGER (1998) *The State Failure Task Force Report: Phase II Findings*. McLean, VA: Science Applications International Corporation. <http://globalpolicy.gmu.edu/pitf/SFTF%20Phase%20II%20Report.pdf> (Accessed June 8, 2009).
- FATEHI-SEDEH, KAMAL, AND HOSSEIN M. SAFIZADEH. (1989) The Association Between Political Instability and Flow of Foreign Direct Investment. *Management International Review* 29: 4–13.
- FEARON, JAMES D., AND DAVID D. LAITIN. (2003) Ethnicity, Insurgency, and Civil War. *American Political Science Review* 97: 75–90.
- FEDER, STANLEY. (1995) Faction and Policon: New Ways to Analyze Politics. In *Inside CIA's Private World: Declassified Articles from the Agency's Internal Journal, 1955–1992*, edited by H. Bradford Westerfield. New Haven: Yale University Press.
- FEDER, STANLEY. (2002) Forecasting for Policy Making in the Post-Cold War Period. *Annual Review of Political Science* 5: 111–125.
- GARTZKE, ERIK. (1998) Kant We All Just Get Along? Opportunity, Willingness, and the Origins of the Democratic Peace. *American Journal of Political Science* 42: 1–27.
- GASIOROWSKI, MARK. (1986) Economic Interdependence and International Conflict: Some Cross-National Evidence. *International Studies Quarterly* 30: 23–38.
- HERMANN, MARGARET G. (1999) *Assessing Leadership Style: A Trait Analysis*. Social Science Automation, Inc., Hilliard, OH. <http://www.socialscienceautomation.com/docs/Lta.pdf> (Accessed May 4, 2009).
- HOFF, PETER D., AND MICHAEL D. WARD. (2004) Modeling Dependencies in International Relations Networks. *Political Analysis* 12: 160–175.
- KAUFMANN, CHAIM. (1996) Possible and Impossible Solutions to Ethnic Civil Wars. *International Security* 20: 136–175.
- KING, GARY, AND LANGCHE ZENG. (2001) Improving Forecasts of State Failure. *World Politics* 53: 623–658.
- LAURITZEN, STEFFEN L., AND DAVID J. SPIEGELHALTER. (1988) Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems. *Journal of the Royal Statistical Society, Series B* 50: 157–224.
- LAZAREVSKA, ELENA, AND JAYNE M. SHOLL. (2005) *The Distinctive Language of Terrorism*. Social Science Automation, Inc., Hilliard, OH. <http://christophe.lalanne.free.fr/Papers/Lazarevska-2005The%20distinctive%20lang.pdf> (Accessed May 4, 2009).
- LUSTICK, IAN S., DAN MIODOWNIK, AND ROY J. EIDELSON. (2004) Secessionism in Multicultural States: Does Sharing Power Prevent or Encourage It? *American Political Science Review* 98: 209–230.
- MASON, T. DAVID, AND PATRICK J. FETT. (1996) How Civil Wars End: A Rational Choice Approach. *Journal of Conflict Resolution* 40: 546–568.
- MCCLELLAND, CHARLES A., AND GARY HOGGARD. (1969) Conflict Patterns in the Interactions Among Nations. In *International Politics and Foreign Policy*, edited by James N. Rosenau. (Revised edition) New York: The Free Press.

- NEAPOLITAN, RICHARD E. (2003) *Learning Bayesian Networks*. Upper Saddle River, NJ: Prentice Hall.
- O'BRIEN, SEAN P. (2002) Anticipating the Good, the Bad, and the Ugly: An Early Warning Approach to Conflict and Instability Analysis. *Journal of Conflict Resolution* 46: 791–811.
- OLIBE, KINGSLEY O., AND D. LARRY CRUMBLY. (1997) Determinants of US Private Foreign Direct Investments in OPEC Nations: From Public and Non-Public Policy Perspectives. *Journal of Public Budgeting, Accounting and Financial Management* 9: 331–355.
- ONEAL, JOHN R., AND BRUCE M. RUSSETT. (1997) The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985. *International Studies Quarterly* 41: 267–294.
- PEARL, JUDEA. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.
- PEVEHOUSE, JON C., AND JOSHUA S. GOLDSTEIN. (1999) Serbian Compliance or Defiance in Kosovo? Statistical Analysis and Real-Time Predictions *Journal of Conflict Resolution* 43: 538–546.
- PRZEWORSKI, ADAM, AND HENRY TEUNE. (1970) *The Logic of Comparative Social Inquiry*. New York, London, Toronto, Sidney: Wiley-Interscience.
- RUSSETT, BRUCE, JOHN ONEAL, AND DAVID R. DAVIS. (1998) The Third Leg of the Kantian Tripod for Peace: International Organizations and Militarized Disputes, 1950–1985. *International Organization* 52: 441–467.
- SCHNEIDER, FRIEDRICH, AND BRUNO S. FREY. (1985) Economic and Political Determinants of Foreign Direct Investment. *World Development* 13: 161–175.
- SCHRODT, PHILIP A. (1997) *Early Warning of Conflict in Southern Lebanon Using Hidden Markov Models*. Paper prepared for presentation at the annual meeting of the American Political Science Association, Washington, DC. <http://web.ku.edu/~keds/papers.dir/schro97b.pdf> (Accessed June 8, 2009).
- SCHRODT, PHILIP A. (1998) Pattern Recognition of International Crises Using Hidden Markov Models. In *Non-Linear Models and Methods in Political Science*, edited by Diana Richards. Ann Arbor: University of Michigan Press.
- SCHRODT, PHILIP A., AND DEBORAH J. GERNER. (1997) Empirical Indicators of Crisis Phase in the Middle East, 1979–1995. *Journal of Conflict Resolution* 41: 529–552.
- SCHRODT, PHILIP A., ERIN M. SIMPSON, AND DEBORAH J. GERNER. (2001) Monitoring Conflict Using Automated Coding of Newswire Sources: A Comparison of Five Geographical Regions. Paper presented at the PRIO/Uppsala University/DECRC High-Level Scientific Conference on Identifying Wars: Systematic Conflict Research and Its Utility in Conflict Resolution and Prevention, Uppsala, Sweden, June 8–9 2001. <http://web.ku.edu/~keds/papers.dir/schro97b.pdf> (Accessed June 8, 2009).
- SHELLMAN, STEPHEN M. (2004) Measuring the Intensity of Intranational Political Interactions Event Data: Two Interval-Like Scales. *International Interactions* 30: 109–141.
- SHELLMAN, STEPHEN M. (2008) Machine Coding Non-state Actors' Behavior in Intrastate Conflict. *Political Analysis* 16: 464–477.
- SHELLMAN, STEPHEN M., CLARE HATFIELD, AND MAGGIE MILLS. (2010) Disaggregating Actors in Intrastate Conflict. *Journal of Peace Research* 47: 1.
- SILVERMAN, BARRY G., GNANA BHARATHY, BENJAMIN NYE, AND TONY SMITH. (2008) Modeling Factions for Effects Based Operations, Part II: Behavioral Game Theory. *Computational & Mathematical Organization Theory* 14: 120–155.
- UPSON, SANDRA. (2008) Virginia Tech Is Building an Artificial America in a Supercomputer, December 8. *IEEE Spectrum* <http://www.spectrum.ieee.org/dec08/7051> (Accessed May 6, 2009).
- WALKER, STEPHEN G. (2000) *Forecasting the Political Behavior of Leaders with the Verbs in Context System of Operational Code Analysis*. Hillard, OH: Social Science Automation.
- WARD, MICHAEL D., AND KRISTIAN SKREDE GLEDITSCH. (2002) Location, Location, Location: An MCMC Approach to Modeling the Spatial Context of War and Peace. *Political Analysis* 10: 244–260.
- WARD, MICHAEL D., BRIAN B. GREENHILL, AND KRISTIN BAKKE. (Forthcoming) The Perils of Policy by P-Value: Predicting Civil Conflicts. *Journal of Peace Research* 47: 5.
- WOODWARD, DOUGLAS, AND ROBERT ROLFE. (1993) The Location of Export-Oriented Foreign Direct Investment in the Caribbean Basin. *Journal of International Business Studies* 24: 121–144.
- YANG, YING, AND GEOFFREY I. WEBB. (2002) A Comparative Study of Discretization Methods for Naïve-Bayes Classifiers. Proceedings of PKAW 2002, The 2002 Pacific Rim Knowledge Acquisition Workshop, Tokyo, Japan. <http://www.csse.monash.edu/~webb/Files/YangWebb02a.pdf> (Accessed June 3, 2009).