Precedents, Progress and Prospects in Political Event Data *

Philip A. Schrodt

Department of Political Science

Pennsylvania State University

University Park, PA 16802

schrodt@psu.edu

Version 1.0 : March 20, 2012

*This project was funded in part by National Science Foundation grant SES-1004414. The paper has benefitted from extended discussions and experimentation within the ICEWS team and the KEDS research group at the University of Kansas; I would note in particular contributions from Matthias Heilke, Baris Kesgin, Jennifer Lautenschlager, Hans Leonard, Vladimir Petrov, Steve Purpura, Andrew Shilliday, Steve Shellman, Brandon Stewart, and David Van Brackle, as well as input from Will Lowe and Peter Nardulli.

Abstract

The past decade has seen a renaissance in the development of political event data sets. This has been due to at least the three sets of factors. First, there have been technological changes that have reduced the cost of producing event data, including the availability of information on the Web, the development of specialized systems for automated coding, and the development of machine-assisted systems which reduce the cost of human coding. Second, event data have become much more elaborate than the original state-centric data sets such as WEIS and COPDAB, with a far greater emphasis on sub-state and non-state actors, and in some data sets, the incorporation of geospatial information. Finally, there have been major institutional investments, such as support for a number of Uppsala and PRIO data sets, the DARPA ICEWS Asian and global data sets, and various political violence data sets from the U.S. government. This paper will first review the major new contributions, with a focus on those represented in this special issue, discuss some of the open problems in the existing data and finally discuss prospects for future development, including the enhanced use of open-source natural language processing tools, standardizing the coding taxonomies, and prospects for near-real-time coding systems.

1 Introduction

Political event data have had a long presence in the quantitative study of international politics, dating back to the early efforts of Edward Azar's COPDAB [Azar, 1980] and Charles McClelland's WEIS [McClelland, 1976] as well as a variety of more specialized efforts such as Leng's BCOW [Leng, 1987]. By the late 1980s, the NSF-funded *Data Development in International Relations* project [Merritt et al., 1993] had identified event data as the second most common form of data—behind the various Correlates of War data sets—used in quantitative studies [McGowan et al., 1988]. The 1990s saw the development of two practical automated event data coding systems, the NSF-funded KEDS [Gerner et al., 1994, Schrodt and Gerner, 1994] and the proprietary VRA-Reader (`http://vranet.com`; King and Lowe 2004) and in the 2000s, the development of two new political event coding taxonomies—CAMEO [Gerner et al., 2009] and IDEA [Bond et al., 2003]—designed for implementation in automated coding systems.

While these efforts had built a substantial foundation for event data, the overall investment in the technology remained relatively small. This situation changed dramatically after 2000 with two major initiatives. First, there was a substantial and sustained investment beginning around 2003 in a variety of inter-linked event data sets—some following the older COW norms of coding extended conflict episodes, others focusing on specific acts of violence—by the Peace Research Institute Oslo and the Department of Peace and Conflict Research at Uppsala University.

Second, in 2007 the U.S. Defense Research Projects Agency (DARPA) initiated the Integrated Conflict Early Warning System (ICEWS; O'Brien 2010) which invested very substantial resources in event data development, particularly automated coding. The key difference between the ICEWS event data coding efforts and those of earlier NSF-funded efforts was the scale. As O'Brien—the ICEWS project director—notes,

> ...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 [area of responsibility] 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*).

As the extended scope and variety of the data sets discussed in this volume demonstrate, this field is now very active and the number of data sets is increasing rapidly

In the discussion which follows, I will use the following terminology to distinguish between the types of information being coded. An "event" is a discrete incident that can be located at a single time (usually precise to a day) and set of actors, usually a dyad of a source and target, although policy statements and other announcements may not have a target actor other than "the world." "Events" are distinct from "structural data" such as GDP or Polity scores (`http://www.systemicpeace.org/polity/polity4.htm`). "Episodic" data are those coding the characteristics of an extended set of events such as a war or a crisis: the Correlates of War project (COW; `http://www.correlatesofwar.org/`) is the archetype; International Crisis Behavior (`http://www.cidcm.umd.edu/icb/`) would be a more recent example. "Composite" events are those which occur in a relatively short period of time and limited geographical space—for example a terrorist attack—and multiple characteristics of the incident are coded. Finally, "atomic" events are basic units of political interaction—date, source, target, event—found in classic event data sets such as WEIS and COPDAB, and in contemporary coding schemes such as IDEA and CAMEO. As with any typology, not all of the data sets fit clearly into a single category, but most will.

2 Contemporary Data Sets

Table 1 is an extensive, but by no means comprehensive, list of event data sets in production after 2000. The inclusion criterion was largely those data sets that are either used in the articles in this volume, or which I'm referring to elsewhere in this essay. Several of these projects—notably MID, UCDP, SIGACTS, KEDS and KOSVED—involve a variety of inter-related data sets, and in the case of MID and UCDP, are a mix of episodic and composite event data.[1]

Space does not allow a detailed discussion of each of these—in most cases, it is best just to go to their web sites for the details, as some of the coding protocols run for hundreds of pages—but for the purposes of the discussion to follow, a few additional points on some of the projects:

- VRA: Despite the citation to King and Lowe, this data, and the underlying IDEA event taxonomy, were the product of the commercial firm VRA (`http://vranet.com`). VRA worked on one of the sub-projects of the first Phase of ICEWS, which would require data for Asia through 2006, though it is unclear whether their global data set is being maintained.

- ICEWS: The ICEWS Asia data has now been used in several conference presentations, as well as O'Brien [2010], though it is still unclear when and whether it will be released to the public. During the summer of 2011, the prime ICEWS contractor, Lockheed's Advanced Technology Laboratory (ATL) produced a global data set, its status following the end of ICEWS in December-2011 is also unclear. If the global data is released

[1] I have not included a number protest event data sets that are primarily coming out of sociology: a review of these can be found in Earl et al. [2004].

Table 1: Selected event data projects

Dataset	Focus	Geographical	Years	Geo-located?	Reference
VAR	general	global	1990-2004	no	[King and Lowe, 2006]
KEDS	general	primarily Middle East	1979-present	no	[Schrodt and Gerner, 2010]
ICEWS	general	Asia; global	1998-2010	no	[O'Brien, 2010]
SPEED	general	global	1946-present	city	[Nardulli, 2011]
ACLED	conflict	primarily Africa	1997-2010	yes	[Raleigh et al., 2010]
MID3 Incidents	conflict	global	1993-2001	no	[COW, 2007]
ACD	conflict	global	1998-2010	no	[Harbom and Wallensteen, 2010]
UCDP-GED	conflict	global	1989-2009	yes	[Melander and Sundberg, 2011]
EDACS	violence	failed states	1990-2009	yes	[Chojnacki et al., 2012]
SCAD	protest	Africa	1990-2010	yes	[Salehyan et al., 2012]
GTD	terrorism	global	1970-2010	yes	[START, 2012]
WITS	terrorism	global	2004-2010	city	[NCTC, 2011]
PITFWAD	one-sided violence	global	1995-present	yes	[PITF, 2011]
KOSVED	one-sided violence	selected states	varies by case	yes	[Schneider et al., 2012]
NIRI	violence	Northern Ireland	1968-1998	yes	[Sullivan et al., 2012]
WARICC	water-related conflict	Africa Middle East	1997-2009	yes	[Bernauer et al., 2012]
Urban Violence	urban disorder	Africa , Asia	1960-2009	yes	[Urdal and Hoelscher, 2012]
SIGACTS	violence	Afghanistan Iraq	2004-2010	yes	[HSRP, 2010, Linke et al., 2012]

and a suitable splicing with the VRA dataset could be done (not a trivial task given the difference in sources and coding methods), this would provide global coverage for the entire post-Cold-War period, and splicing it with WEIS would produce a data set back to 1966.

- MID currently goes only to 2001 but an update to 2010, presumably including the incident data, is expected within the year.

- SIGACTS is U.S. military parlance for "significant [enemy] actions" and comprises a series of data sets for the U.S. wars in Iraq and Afghanistan. Versions of these have made their way onto several web sites—for example `http://publicintelligence.net/nato-restricted-afghanistan-security-reports-and-sigacts-maps-august-october-2010/` `http://cryptome.org/0002/sigacts/centcom-sigacts.htm; both accessed 19 March 2012`—and into academic presentations (see Linke et al. [2012]), though at present no "standard" version of these appears to exist. Additional unauthorized versions these data may also be distilled from various WikiLeaks files.

Several things are apparent from this table. The first is the sheer volume and variety of the data, and again, this is not a comprehensive list. This is all the more remarkable given that during the period from about 1980 to 1995, there were essentially no new event data sets developed other than the largely experimental work related to the KEDS project [Schrodt, 2006b]. Furthermore, most of these are relatively recent; in fact several have not been fully released, though this is likely to have changed by the time this volume is in print.

A major reason for the proliferation of the new data is evident from the temporal coverage, which usually begins in 1990, the point at which news sources become readily available in

the major data services such as Lexis-Nexis (LN) and Factiva. Data sets which have used a mix of electronic and conventional sources such as NIRI [Urdal and Hoelscher, 2012] go back further, and SPEED has obtained digitized versions of the *New York Times* and *Wall Street Journal*, the CIA *Foreign Broadcast Information Service* and the BBC *Summary of World Broadcasts* back to 1946, an extraordinarily rich resource. But the low-hanging fruit is clearly 1990 to the present, an interval now more than two decades in length and increasing with each passing year.

In contrast to the pre-2000 data sets, which were generally global, the contemporary data involve a mix of global and regional coverage. This is in part due to the more focused character of a number of the data sets—if one is studying substate conflict, it makes more sense to code countries where such behavior can be found than, say, New Zealand and Japan—but also to the problems of downloading and coding the quantity of text required for a global data set.

With the exception of the atomic data sets—VRA, KEDS, ICEWS and SPEED—most of the new data sets are focused on specific categories of behavior. But they provide considerably greater information about these events, for example characteristics of the victims and perpetrators of violent events, the number of victims, and location. These collections are often driven by specific theoretical questions—for example the circumstances under which government repression will increase or decrease dissent, the impact of climate and resources on conflict, and the contagion of violence—rather than the "one size fits all" approach of general coding schemes such as WEIS, COPDAB, IDEA and CAMEO.

This detail, combined with the very fine temporal detail of event data—a huge contrast to the generally annual level of structural data sets—means these new data sets are coming

close to a form of "thick description" usually associated with qualitative data.[2] This is also leading to "quantitative case studies" where a single case (or small number of cases) is studied in detail over time—for example the NIRI focus on Northern Ireland [Sullivan et al., 2012], the KEDS focus on the Levant, and the examination of the Bosnia case by [Schneider et al., 2012]. This is a substantial contrast to the annualized and global country-level studies which predominated in research prior to 2000.

The scale of the projects varies widely. Most have some institutional support, usually a university research center and/or national research funding such as the U.S. National Science Foundation and its European counterparts, but are generally the product of a small number of researchers working for a small number of years. Some, however—notably MID, ICEWS, SPEED, UCDP-PRIO, WITS, GTD and SIGACTS—involve funding in the range of a million or more dollars, large interdisciplinary and multi-institution teams, and have been sustained for a number of years, with the expectation of relatively frequent future updates. They are "big science" with substantial investments of public funding, again a complete contrast to the pre-2000 experience.

Geospatial coding is now found in about half of the data sets, a complete change from the pre-2000 data, where the most finely-grained spatial coding was the nation-state. In a number of instances, standardized coordinates are provided, to the extent that these can be determined by the news reports—Chojnacki et al. [2012] discuss just how difficult this can be—and as such the datasets can easily interface with contemporary geospatial visualization and analysis tools, in particular *GoogleEarth*. This continues a general trend towards greater

[2]Complex coding was also found in the pre-2000 BCOW data set, but probably due to its complexity, BCOW was almost never used in published analyses except by Leng [1993a,b].

use of geospatial analysis of conflict behavior [Buhaug and Gates, 2002, Buhaug and Lujala, 2005, Buhaug and Rød, 2006, Cederman and Gleditsch, 2009, Franzese and Hays, 2008] and it is likely to be of increasing importance due to climate-related studies [Urdal, 2008, Gleditsch, 2012, Bernauer et al., 2012].

While only three of the projects doing atomic event coding—VRA, KEDS and ICEWS—are using fully automated coding, several others—SPEED, EDACS, MID[3]—are employing sophisticated natural language processing (NLP) tools for partial data field extraction and particularly for source text filtering. This latter task—automatically going through a large set of texts generated by a general set of Boolean search terms and identifying only those stories likely to require coding—can result in a dramatic reduction, 90% or higher, in the set of stories coders need to evaluate. This partially alleviates the "drinking from a firehose" problem inherent in the very large text databases now available. The SPEED project has gone well beyond this and, in collaboration with computer scientists, has developed several suites of tools that automate much of the coding process, while using human coders for the final coding decisions (see `http://www.clinecenter.illinois.edu/research/speed-components.html`)

3 Problems

3.1 Sources

The impact of the diversity of sources now available remains unclear. Chojnacki et al. [2012] for example, document a very worrisome divergence between EDACS and ACLED in areas where the two should show similar trends. No consensus has emerged on several key issues,

[3]This is used in the current MID4 work: see Schrodt et al. [2008], Landis et al. [2011]

including

- The use of summaries such as the BBC *Summary of World Broadcasts* compared to primary news feeds such as Reuters and Agence France Press (AFP);

- The value added by additional sources: some studies suggest that as few as two general sources may be sufficient, whereas other projects make extensive use of non-news sources such as NGO and legal reports. The number of sources varies widely: VRA and KEDS use a single source; SCAD uses two; SPEED and EADAC four sources; ACLED and ICEWS fifty or more, and the news aggregators provide around 4,000 sources;

- The circumstances under which local sources are needed. These are increasingly available either directly from the Web or through news aggegators such as GoogleNews (GN; `http://news.google.com/`) and the European Media Monitor (EMM; `http://emm.newsbrief.eu/`). The answer may vary by region: for example international sources may be sufficient in the Middle East and the China-Taiwan dyad but may provide very important additional information in Africa and South Asia. This also raises the issue of the need for non-English sources.

3.2 Duplicate Stories

Downloads from aggregators such as LN, Factiva, GN and EMM contain a very large number of stories that are either literally or effectively duplicates. These generally come from five sources

- Exact duplicates, where a local source simply reprints the contents of an international newswire story. This is why newswires exist, so it happens a lot;
- Multiple reports of the same event—for example a suicide bombing—as it develops;
- Stories repeated to correct minor errors such as incorrect dates or spelling;
- Lead sentences that are repeated in general news summaries issued during the day;
- Multiple independent reports of the event from different news sources:

Duplicate detection, particularly across multiple sources in very large files, is a very difficult problem: The term-of-art used in the NLP literature is "near duplicate detection."

When used simply for prediction mode, for example in the ICEWS project, duplicates are not necessarily a bad thing, since they generally will amplify politically-relevant signals. If reporters or editors think that something is important, it is more likely to be repeated, both within sources and across sources, than something that is mundane.

However, when trying to measure trends in "ground-truth" behavior against a baseline over a long period time, duplicates are a serious problem, both across sources and within sources. Cross-source duplication has probably changed considerably over the past fifteen years due to local sources putting increasing amounts of material on the Web, and more generally with the globalization of the news economy, so that events in once-obscure places are now available in machine-readable form. In-source duplication can change due both to changes in the resources available to an organization—Reuters went through something of an organizational near-death experience during the period 1998-2002 [Mooney and Simpson, 2003] and the frequency of its reporting dropped dramatically during that time—and editorial

policies on updating, corrections and the production of summaries.

3.3 Irrelevant Stories

Irrelevant stories have been the bane of the automated processing of event data source texts from the beginning. For example, the search strategy for our now-30-year "Levant" data set simply looks for stories containing the names or synonyms of the six actors we are tracking: Egypt, Israel, Jordan, Lebanon, the Palestinians, and Syria. However, our early downloads covered the peak of the career of basketball player Michael Jordan and we ended up with quite a number of basketball stories. These are relatively harmless and easily discarded in automated coding, but they do involve needless downloading. This is further complicated by the unpredictability of the LN search engine: one of the downloads for MID4, for example, included a 1,829-line listing of every commercial cruise for 2004, with destinations, prices, and amenities, presumably because somewhere in that text was a term contained in our search string for acts and threats of political violence.

While irrelevant stories are largely an issue for automated coding, their sheer magnitude is a problem for human coding as well, since each story must processed and read (or at least skimmed) by coders before being rejected. Both MID4 and SPEED, while using human coding, devoted substantial efforts—generally successful—in identifying irrelevant stories. In the case of MID4, this reduced the irrelevant stories by over 90% with no false negatives. The SPEED "BIN" system (`http://www.clinecenter.illinois.edu/research/publications/SPEED-BIN.pdf` has only a 1%-3% false negative rate.

Stories that are easily rejected by human coders who can put the story into context can

be much more problematic for automated coding systems which interpret stories literally and on a sentence-by-sentence basis. The most important are chronologies and retrospectives, which describe political events that occurred in the sometimes distant past, yet with a contemporary dateline. For example various World War II commemorations typically receive extensive coverage and could be miscoded as conflict behavior between the US, Germany and Japan.

Another longstanding problem are international sports competitions that use military metaphors. World Cup reports, for example, always use the simple national names—Netherlands versus Spain—and not infrequently use terms such as "battle," "fought," "stand-off" and the like. These can usually by solved by discard phrases involving every imaginable form of competition, sporting and others. But even this will fail when the sports context is implicit, such as a [hypothetical] report on 11 July 2010 that might begin, with little concern that it will be misinterpreted, "Fans eagerly await tonight's battle between the Netherlands and Spain." Such stories can be as much as a third of the texts in areas where little seems to be happening except sports—in the ICEWS downloads, Australia.

3.4 Coding Error

As I noted in Schrodt [1994], coding error is only one potential source of error between "events on the ground" and the evaluation of a statistical model. News reports are only a tiny, tiny fraction of all of the events that occur daily, and they are non-randomly selected by reporters and editors; event taxonomies such as WEIS, CAMEO and IDEA are very generic and bin together events that may not always belong together; and statistical models

invariably contain specification error, coefficient estimate standard errors, and the intrinsic randomness found in any open complex system. In this chain of events, the impact of coding error, while still relevant, is not necessarily dominant.

Multiple independent tests, most recently King and Lowe [2004] have shown that machine coding is comparable in accuracy to human coding. But the human coding accuracy in some of those tests is quite low: King and Lowe's coder accuracy on the individual VRA codes alone (Table 2, pg 631)—not the complete record with source and target identification, another major potential source of error—is in the range 25% (!) to 50% for the detailed codes and 55% - 70% for the cue categories.

The King and Lowe results are not anomalous: Mikhaylov et al. [2012] find similar results in cross-checking the human coding of the Comparative Manifestos Project, a topic classification task quite similar to event data coding. Despite CMP's claim of 80% reliability, their experiments find the intercoder reliability averages about half that figure, with ranges quite similar to those found in King and Lowe; Ruggeri et al. [2011] report an agreement rate of 16% to 73% in another human event coding exercise. Those levels are almost certainly well within the range of existing machine coding technology, at least for atomic event coding.

An extensive recent body of psychological work—see Baumeister and Tierney [2011] for a popular treatment—indicates that the sustained decision-making required for human coding presents almost a perfect storm for inducing fatigue, inattention, and a tendency to use heuristic shortcuts.These physiological costs are far more deeply rooted than previously assumed, and can only be reduced, not eliminated, by improved coding protocols, training, coder selection and supervision. The human brain was simply never intended for the tasks we impose on coders.

3.5 Statistical Methods

Event data, as a categorical time series, are difficult to fit into the standard repertoire of statistical methods found in political science. There has unquestionably been progress made in this regard—the articles in this volume show considerable breadth of technique, particularly in comparison to pre-2000 works that generally used simple OLS regression and time-series, often with scaled data, and the original work in the 1970s which often used only crosstabulation [Azar et al., 1972, Daly and Andriole, 1980, Cimbala, 1987]. However, quite an assortment of open issues—perhaps better thought of as opportunities—remain.

- Event data analysis requires a great number of choices in aggregation: temporal, event types, definitions of actor dyads, and spatial in the case of geolocated data. Each of these decisions can potentially affect the results [Shellman, 2004].

- Events are probably best thought of as sequences rather than conventional time series, and we are just beginning to see temporal sequence methods applied [Sullivan et al., 2012]: a variety of these are available from the fields of genetic and linguistic analysis;

- in predictive models, it is quite likely that machine learning methods will be useful, since these are usually more robust than conventional statistical approaches in situations with high dimensional and irregularly distributed data. These have not seen extensive experimentation

- The issue of assessing the degree of accuracy in a probabilistic forecast is quite complex, and there is no single answer to this issue, though considerable guidance can be found in the areas of economic and meteorological forecasting [Brandt et al., 2011]

- Event data is perfectly suited for network analysis, a topic that has gotten a great deal of attention in quantitative political science recently (see [Barnett, 2011])

- The issue of splicing data sets—across time, across different data sets with different coverage (e.g. WEIS, VRA, ICEWS) across different sources, and across different regions—remains an open issue [Reuveny and Kang, 1996]. In principle these data sets are all measuring the same underlying phenomena and should be commensurate with appropriate scaling, but in practice this has proven problematic. The rapid changes in the international news media are further complicating this, particularly in very long time frames such as SPEED's 1946-2010.

3.6 Error detection and correction

Several projects—ACLED, PITFWAD, EDACS—are now trying to provide some indication of the reliability of the report at a qualitative level. In theory, at least, this could be incorporated into statistical estimation methods though to date I am not aware of any analyses that have used this information.

Some ICEWS experiments have shown that even very simple filters can eliminate egregious errors such coding USA/Japanese conflict events based on Pearl Harbor travel and movie reviews or anniversaries of the bombings of Hiroshima and Nagasaki. However, far more sophisticated filtering methods are available, and many of these are of relatively recent vintage due to the computing power required. A multi-category support vector machine (SVM), for example, could be applied to the full text of a story to determine whether the story is likely to have produced events of the type coded, based on previously verified correct

codings.

More generally, we need a theory and statistical adjustment for how events get into the text record in the first place. This needs to account for at least the following

- media fatigue—conflict episodes are more likely to be covered at their beginning than later and this interest probably decays exponentially [Gerner and Schrodt, 1998]
- media competition—when do multiple sources specialize and when do they exhibit herd behavior?
- systematic variation in the level of coverage across countries and regions, including the effects of government suppression of the press [Davenport and Ball, 2002].
- the accuracy of fatality counts [Eck and Hultman, 2007, Spagat et al., 2010, Chojnacki et al., 2012]
- the fact that high intensity events such as violence and routine events such as scheduled meetings are more likely to be reported than unexpected events such as a threat issued during an unplanned campaign stop
- incorrect classification is likely to be systematic rather than random, and the estimated misclassification matrices could be used to adjust the aggregate counts

4 Prospects

4.1 Fully-automated coding

TABARI and the VRA-Reader were developed around 2000, and both were based on the experience of KEDS , the first practical automated coder. As a result of the ICEWS project, several additional proprietary coders have either been developed or are under development: these include Lockheed ATL's JABARI family and Strategic Analysis Enterprise's XENOPHON, `http://strategicanalysisenterprises.com/services.php` as well as the inclusion of event coders on top of existing software, for example the "Behavior and Events from News" (BEN) system built into Social Science Automation's *Profiler Plus* and efforts by BBN and IBM to use their existing NLP "event triple" software for event coding.

While the projects coding composite events still use human coding, for real-time coding of atomic events, there is simply no alternative to automated methods. Sustained human coding projects, once one takes in the issues of training, retraining, replacement, cross-coding, re-coding due to effects of coding drift and/or slacker-coders and so forth, usually ends up coding about six events per hour. Individual coders, particularly working for short periods of time, and definitely if they are a principal investigator trying to assess coding speed, can reliably code much faster than this. But for the *overall* labor requirements—that is, the total time invested in the enterprise divided by the resulting useable events—the 6 events per hour is a pretty good rule of thumb and—like the labor requirements of a string quartet—has changed little over time. Machine-assisted coding software *may* be able to increase this rate—SPEED is hoping to reduce the time required for coding its composite events to 3 minutes—but the sheer quantity of the available text presents a daunting challenge.

SPEED reports that its intake is currently 100,000 stories per day. Even assuming that 80% of these are duplicates or irrelevant—roughly the proportion in the original ICEWS downloads—and could be removed completely efficiently, the remaining coding would require a permanent team of around 100 to 400 *coders*—depending on the effectiveness of the machine-assisted coding software—a labor requirement that would probably need to be multiplied by at least 1.5 to account for management, training, quality control and turnover. This is simply beyond the capacity of existing academic political science enterprises. TABARI codes about 5,000 sentences per second, and any automated coder can be trivially scaled to any speed needed by dividing the texts across multiple processors in now widely-available cluster computers and thus presents no such problems.

The challenge is extending the success of automated coding to projects which are coding composite events, which at the present time can still only be done using human coding. This may be possible to develop specific "data field extraction" methods, for example locating reports of the number of individuals killed or the amount of aid promised. An NLP literature exists on this and several such methods are used in SPEED and EDACS. One would then define composite events such as "civil war" by using patterns of the atomic events and the extracted fields; Hudson et al. [2008] demonstrates a tool for doing this with CAMEO data. Such automation would have the additional benefit of making the differences between definitions used by various project unambiguous (or at least comparable) and allow the composite events to be easily constructed out of existing data sets rather than starting every new project from the beginning. MID, in moving from the original episodic definitions to coding composite incidents as well, would be an example of this approach, albeit with human coding.

4.2 Open Source NLP[4]

Earlier coding software was largely self-contained. Now, however, it makes far more sense to leave the NLP software development to the computational linguists whenever possible, and focus only on those remaining tasks specifically required for coding events. This would be consistent with the broader incorporation of automated text processing into political science [Monroe and Schrodt, 2008] in contexts such as the analysis of legislative debate and party platforms.

Major open-source NLP software sites include

- Open-NLP `http://opennlp.apache.org/`
- GATE; `http://gate.ac.uk/`
- University of Illinois Cognitive Computation Group:
  `http://cogcomp.cs.illinois.edu/page/software`
- Stanford NLP Group: `http://nlp.stanford.edu/software/index.shtml`

LingPipe's "Competition" page (`http://alias-i.com/lingpipe/web/competition.html`) lists—as of March 2012—no fewer than 23 academic/open-source NLP projects, and 122 commercial projects. This is quite different than the situation in statistical software, where at present there are only four major systems in wide use (SPSS, SAS, Stata and R), and perhaps a dozen or some additional specialized systems.

NLP tasks relevant to event coding include the following:

[4]This section in particular has benefitted from a variety of discussions with David Van Brackle, Will Lowe, Steve Purpura and Steve Shellman

- Full-parsing. An assortment of full-parsers are available, and the *TreeBank* parse format (`http://www.cis.upenn.edu/~treebank/`) appears to be a fairly stable and standard output format, so a researcher could use the parser of his or her choice so long as these could produce *TreeBank*-formatted output.

- Disambiguation by parts-of-speech markup. One of the major tasks of the TABARI dictionaries is noun-verb disambiguation: this issue accounts for much of their size and complexity. Parts-of-speech (POS) marking eliminates this problem. Stemming—most frequently the Porter stemming algorithm for English—further simplifies and generalizes coding dictionaries.

- Named Entity Recognition and Resolution (NER). These are systems for the recognition of names within text, and also resolving equivalent names, e.g. "President Obama," "President Barak Obama," "United States President Barak Hussein Obama" and so forth. Some of these methods are very sophisticated—for example using conditional random fields and hidden Markov models—though it is not entirely clear these are needed for NER when dealing with political actors found in news reports, who have fairly regularized names and titles.

- Synonym and relational dictionaries. The *WordNet* lexical database provides a nearly comprehensive list of synonyms ("synsets"), hyponyms and hypernyms for the English language; this could be used to replace specific instances of nouns and verbs with general classes. Various projects have also assembled extensive lists of specialized words such as currencies, occupations, first names, titles and so forth: See for example the files in the `LbjNerTagger1.11.release/Data/KnownLists` folder that can be downloaded as

part of the Illinois NER system at `http://cogcomp.cs.illinois.edu/page/download_view/4`.

- Regular expressions. Given the ubiquity of regular expressions in the contemporary computing environment—regular expressions have been called "the calculus of string processing"—it would be very useful to allow these to be used as the pattern-specification language, rather than the ad hoc syntax used in most of the existing systems.

- NGA GEOnet Names Server (GNS). The National Geospatial-Intelligence Agency maintains a continuously-updatyed database of approximately 5.5-million geographical place names (`http://earth-info.nga.mil/gns/html/`). This could be incorporated into a program for identifying the location of events described in a story which would improve disambiguation of actors and agents, as well as providing geo-located data for those events that have an unambiguous location.

- Machine translation. These systems are gradually improving—and are the subject of massive investment by a variety of firms—and it is quite likely that these translations, however imperfect and frustrating they may seem to a fluent speaker, may be quite adequate for the relative limited needs of event coding. Machine translation is likely to develop most rapidly in areas where there is a large population of elites who communicate in a language other than English—Chinese, Arabic, Spanish and French are prime candidates—and this could compensate for coverage biases in the English-language international news services.

The use of these tools would align automated event coding and machine-assisted coding with the research output of the larger NLP community, and as their tools improved, we

could incorporate those improvements into our work immediately. Two systems have already moved in this direction. Lockheed modied their JABARI coder—originally just a proprietary Java version of the open-source, C/C++ TABARI —to produce JABARI-NLP, which uses open-source NLP software for considerable pre-processing of the texts. This resulted in substantial improvement in coding accuracy, particularly by insuring that the event target identified by the coder was actually syntactically part of the verb phrase generating the event. SPEED has also worked extensively with the computer science community at Illinois to integrate state-of-the-art NLP into their machine-assisted coding.

More generally, off-loading most of the NLP tasks to sophisticated pre-processing programs means that programs for automated coding or machine-assisted coding can be much shorter and in more easily maintained forms using a robust scripting language—probably, given its wide application in NLP tasks, `Python`—rather than in a fully compiled language such as `C/C++` or `Java`. Scripting languages take a performance hit but, due to their internal memory management, are usually considerably shorter and easier to maintain. In addition, since almost all high-volume coding will now be done in parallel environments, speed is less critical than when coding was done on single processors.

5 Near-Real-Time Coding

The widespread availability of real-time news reports on the Web and through alternative media such as RSS feeds opens the possibility of near-real-time coding of atomic events. Implementation simply involves linking together the appropriate web-scraping scripts, formatters, duplicate detection programs, an automated coder, a database such as `mySQL`, and

a web-based interface.[5] We implemented an experimental system in 2008-2009, and our 18-month experiment has found at least three characteristics of the data that should be taken into account in the design of any future systems.

- While *in principle* one could get real-time coding—news monitoring services used in support of automated financial trading systems routinely do this—there is little reason to do so for existing event data applications, which generally do not work on data that is less finely grained than a day. Furthermore, the news feeds received during the course of a day are considerably messier—for example with minor corrections and duplications—than those available at the end of a day. Consequently, after initial experiments we updated the data only once a day rather than as soon as the reports became available.

- These are definitely not "build and forget" systems due to the changing organization of the news web sites. We found Reuters went through three or four major reorganizations of their web site during the period we were coding data, and in one instance was off-line for close to a week. The changes in code resulting from these reorganizations were relatively minor, primarily dealing with the locations of files rather than the file formats, but it necessitated periodic—and unexpected—maintenance.

- A semi-automated NER system would be needed to track new actors entering and changing their status in the system, for example a politician losing office or a militant leader getting killed. However, with machine-assisted NER tools, this is within the capabilities of a relatively small project.

[5]UCDP provides a sophisticated model for this for its various composite event sets

5.1 Standardized codes

A final area with considerable promise would be standardizing as many coding taxonomies as possible. This was an area where the quantitative international relations community initially made significant contributions, with the original standardization of COW codes [Russett et al., 1968]. Geospatial coding is largely standardized in event data sets—and hence immediately compatible with other tools—because standards from other fields existed at the time geocoding was added. But in other areas the field has lagged.

Of immediate concern is the persistence of COW codes despite the availability—and widespread application—of an international system of nation-state codes, ISO 3166, which overlaps the COW codes just enough to be confusing. Granted, there is still a legacy issue, but it is difficult to imagine why any new system should not use the ISO standard.

The atomic event coding taxonomies WEIS, IDEA and CAMEO are fairly close to convergence but still not quite there.[6] IDEA and CAMEO were both designed as WEIS replacements but both have problems. IDEA introduced a number of new categories for politically relevant events such as natural disasters but retained backward compatibility with WEIS and thus some of the event category still require distinctions that are difficult to differentiate in machine (or, likely, human) coding. CAMEO combined ambiguous WEIS categories and greatly expanded the level of detail on acts of violence, but contains excessive detail on events related to mediation, the original focus of the taxonomy. SPEED, meanwhile, has introduced some additional refinements not found in either IDEA or CAMEO, particularly in the realm of protest and other forms of direct action.

[6]COPDAB's very simplified ordinal scale has not survived into contemporary use, nor has the highly complex BCOW scale.

The very large scale efforts of ICEWS and SPEED could provide a basis for a new atomic event taxonomy that could survive for some time—WEIS, after all, was in use for close to four decades, despite being intended as a first draft McClelland [1983]. That standard might also extend the definitions of some of the atomic events to specify the coding of auxiliary information when it is available, much as several data sets are already doing with location. For example, acts of violence would have a number of victims; demonstrations would have a size, location and topics. To the extent that these definitions can be standardized, one could then focus on the development of data field extraction modules to automatically code them.

A similar situation, though less stable, is found for the substate and transnational actor taxonomies: the coding of general actors such as chief executives, police, military, rebel groups, IGOs and NGOs and so forth. Again, IDEA and CAMEO are not that far apart, though CAMEO has been evolving (for example it recently introduced a new 1,500 entry classification of world religions and a 650 entry set of ethnic group codes) and still has a number of idiosyncratic elements linked to its origins in regional, rather than global, coding. ICEWS, while using the CAMEO *event* taxonomy with few changes, developed a more extended substate and transnational scheme for its global coding. If the full details of this are released, it might form a basis for a standardized system.

The coding of ethnic groups also remains problematic. One of the most dramatic illustrations of this issue is a comparison of the Joshua Project (`http://www.joshuaproject.net/`) list of 16,598 ethnic groups and the widely-used `Minorities at Risk` data set which tracks only 283 groups. While these efforts are not strictly comparable—the Joshua Project list undoubtedly includes many groups that are not politically relevant and repeats groups across countries—the fact that it is almost two orders of magnitude larger than MAR suggests that

conventional political science coding is missing some detail. Similarly, it would be useful to have standardized codes from persistent international militarized groups such as al-Qaeda and Hizbullah.

6 Conclusion

In a history of the first fifteen years of the KEDS/TABARI project [Schrodt, 2006a], the final section—titled "Mama don't let your babies grow up to be event data analysts"—lamented the low visibility of event data analysis in the political science literature despite major advances in automated coding and the acceptance of that data in all of the major refereed political science journals.

Due to the UCDP, PRIO and ICEWS efforts, the situation now very different. As indicated in Table 1, there is a very active and expanding set of event data sets, and theoretically-motivated papers applying these data, usually with methods far more sophisticated than those seen prior to 2000, are common at conferences, and gradually beginning to find their way into mainstream journals.

Nonetheless, there still appear to be some barriers in the wider acceptance of the method, largely due to unrealistic expectations about the accuracy of coding—human or machine—and an insufficient appreciation of the ability of statistical models to deal with noise. In this final section, I will address some of those concerns.

The situation we have with event data is, I suspect, somewhat similar to that faced by George Gallup and other early scientific pollsters in the development of statistical public opinion assessment prior during the period 1930-1970. Two features are of particular note.

First, many experts were convinced that they could accurately assess opinion using unsystematic methods. Notably the infamous political hacks, the pols hanging around the bars near the statehouse chewing tobacco and drinking whiskey and happy to tell any reporter what "their people" really thought. In the not-infrequent instances when these politicians were rigging the election outcomes they probably were worth listening to, but not as an indicator of true public opinion. But could never be persuaded otherwise—"the pollsters don't know what they are doing" continues to be the refrain of every candidate, anywhere in the world, who is behind in the polls. Likewise, systematic evidence that humans—including experts in critical fields such as medicine and finance—vastly over-estimate their predictive accuracy [Tetlock, 2005, Kahneman, 2011] has been of little avail.

Second, people simply could not grasp the implications of random sampling. This is, in fact, quite amazing: a properly done sample of 2,000 people will usually predict the outcome of a national election of a million voters within a percentage point or so. Extraordinary, really. But then so are iPhones. George Gallup had a quick come-back when confronted with the skeptical "How can your poll possibly be accurate: no pollster has ever contacted me!" Gallup's retort: "The next time you go to the doctor for a blood test, don't settle on a little sample, ask him to test all 5 quarts."

We're in a similar situation, though as yet without the clever response. Despite the demonstrated utility of models based on event data, whether in explanatory or predictive settings, people will not believe that a statistical model generated with methods that are 100% transparent and replicable can do better than their anything-but-transparent-and-replicable intuition. Meehl [1954] pointed out this problem sixty years ago—and six decades of work has yet to provide any refutation of that work—but we still live with these perceptions.

The expert, of course, will usually seek to demonstrate the flaws of the data by pointing to an incorrectly coded sentence or neglected incident. Any event data system, human coded or machine coded, will have plenty of those. But this ignores the fact that the *total* amount of information in the data is vastly greater than that which can be processed by an individual, and while intuitive analysis may be better in a specific selected case (and certainly for a specific selected news report), the *combination* of systematic statistical analysis and large amounts of data has better performance. The individual whiskey-swigging pol might do better than a small random sample in their particular ward, or even their city, but what can they say about a city on the other side of the continent? A subject-matter-expert may perform better on their area of expertise in a particular time frame—though Tetlock's work suggests otherwise—but the event-based ICEWS forecasting models predicted five indicators for 29 countries at a monthly granularity for almost fifteen years, and could readily scale this to cover the entire world.

Efforts should, of course, be made to improve the quality of event coding, and I have outlined a number of fairly clear areas where research could be done, or has already been done and simply needs to be applied. Survey research has demonstrated such incremental improvements: voter-turnout models are continually refined, and researchers deal with technologically-driven issues such as the displacement of easily sampled land-lines by cell phones. As I have indicated in this article, we have made tremendous progress in the past decade, particularly for a field that was essentially dead in 1990. Both technical and coordination issues remain, but solutions should be possible for most of them, and looking towards the future, the combination of the Web and NLP techniques show the possibility of detailed, up-to-date, and relatively inexpensive data sets that can be used to address some of the core

issues in our field. The future today looks very bright.

References

Edward E. Azar. The conflict and peace data bank (COPDAB) project. *Journal of Conflict Resolution*, 24:143–152, 1980.

Edward E. Azar, Richard A. Brody, and Charles A. McClelland, editors. *International Events Interaction Analysis: Some Research Considerations*. Sage Publications, Beverly Hills, 1972.

G. Barnett. *Encyclopedia of Social Networks*. Sage, Sage eReference (Online service), 2011.

Roy F. Baumeister and John Tierney. *Willpower: Rediscovering the Greatest Human Strength*. Penguin, New York, 2011.

Thomas Bernauer, Tobias Böhmelt, Halvard Buhaug, Nils Petter Gleditsch, Theresa Tribaldos, Eivind Berg Weibust, and Gerdis Wischnath. Water-related intrastate conflict and cooperation (WARICC): A new event dataset. *International Interactions*, 38(5):this volume, 2012.

Doug Bond, Joe Bond, Churl Oh, J. Craig Jenkins, and Charles L. Taylor. Integrated data for events analysis (IDEA): An event typology for automated events data development. *Journal of Peace Research*, 40(6):733–745, 2003.

Patrick Brandt, John Freeman, and Philip A. Schrodt. "racing horses: Constructing and evaluating forecasts in political science". Twenty-Eighth Political Methodology Summer Conference, Princeton University (https://webspace.princeton.edu/users/dwintjen/Racing2011.

Halvard Buhaug and Scott Gates. The geography of civil war. *Journal of Peace Research*, 39(4):417–433, 2002.

Halvard Buhaug and Päivi Lujala. Accounting for scale: Measuring geography in quantitative studies of civil war. *Political Geography*, 24(4):399–418, 2005.

Halvard Buhaug and Jan Ketil Rød. Local determinants of African civil wars, 1970-2001. *Political Geography*, 25(6):315–335, 2006.

Lars-Erik Cederman and Kristian Skrede Gleditsch. Introduction to special issue of "disaggregating civil war". *Journal of Conflict Resolution*, 24(4):590–617, 2009.

Sven Chojnacki, Christian Ickler, Michael Spies, and John Wiesel. Event data on armed conflict and security: New perspectives, old challenges, and some solutions. *International Interactions*, 38(5):this volume, 2012.

Stephen Cimbala. *Artificial Intelligence and National Security*. Lexington Books, Lexington, MA, 1987.

Correlates of War Project COW. Militarized interstate disputes (v3.10). `http://www.correlatesofwar.org/COW2%20Data/MIDs/MID310.html`, 2007.

Judith Ayres Daly and Stephen J. Andriole. The use of events/interaction research by the intelligence community. *Policy Sciences*, 12:215–236, 1980.

Christian Davenport and Patrick Ball. Views to a kill: Exploring the implications of source selection in the case of guatemalan state terror, 1977-1995. *Journal of Conflict Resolution*, 46(3):427–450, 2002.

Jennifer Earl, Andrew Martin, John D. McCarthy, and Sarah A. Soule. The use of newspaper data on the study of collective action. *Annual Review of Sociology*, 30:65–80, 2004.

Kristine Eck and Lisa Hultman. One-sided violence against civilians in war: Insights from new fatality data. *Journal of Peace Research*, 44(2):233–246, 2007.

Robert J. Franzese and Jude C. Hays. Interdependence in comparative politics: Substance, theory, empirics, substance. *Comparative Political Studies*, 41(4/5):742–80, 2008.

Deborah J. Gerner and Philip A. Schrodt. The effects of media coverage on crisis assessment and early warning in the middle east. In Susanne Schmeidl and Howard Adelman, editors, *Early Warning and Early Response*. Columbia University Press-Columbia International Affairs Online, 1998.

Deborah J. Gerner, Philip A. Schrodt, Ronald A. Francisco, and Judith L. Weddle. The machine coding of events from regional and international sources. *International Studies Quarterly*, 38:91–119, 1994.

Deborah J. Gerner, Philip A. Schrodt, and Ömür Yilmaz. *Conflict and Mediation Event Observations (CAMEO) Codebook*. `http://eventdata.psu.edu/data.dir/cameo.html`, 2009.

Nils Petter Gleditsch. Whither the weather? climate change and conflict: Introduction to special issue. *Journal of Peace Research*, 49(1):3–9, 2012.

Lotta Harbom and Peter Wallensteen. Armed conflict 1946-2009. *Journal of Peace Research*, 47(4):501–509, 2010.

Human Security Report Project HSRP. Afghanistan conflict monitor. http://www.conflictmonitors.org/countries/afghanistan/facts-and-figures/violent-incide 2010.

Valerie M. Hudson, Philip A. Schrodt, and Ray D. Whitmer. Discrete sequence rule models as a social science methodology: An exploratory analysis of foreign policy rule enactment within Palestinian-Israeli event data. *Foreign Policy Analysis*, 4(2):105–126, 2008.

Daniel Kahneman. *Thinking Fast and Slow*. Farrar, Straus and Giroux, New York, 2011.

Gary King and Will Lowe. An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. *International Organization*, 57(3):617–642, 2004.

Gary King and Will Lowe. 10 million international dyadic events. http://gking.harvard.edu/data?dvn_subpage=/faces/study/StudyPage.xhtml?globalId=hdl:19 2006.

Stephen Landis, Vito D'Orazio, Glenn Palmer, and Philip A. Schrodt. Separating the wheat from the chaff: Application of two-step support vector machines to MID 4 text classification. Midwest Political Science Association, Chicago, 2011.

Russell J Leng. *Behavioral Correlates of War, 1816-1975. (ICPSR 8606)*. Inter-University Consortium for Political and Social Research, Ann Arbor, 1987.

Russell J. Leng. Reciprocating influence strategies in interstate crisis bargaining. *Journal of Conflict Resolution*, 37:3–41, 1993a.

Russell J. Leng. *Interstate Crisis Behavior, 1816-1980.* Cambridge University Press, New York, 1993b.

Andrew M. Linke, Frank D. W. Witmer, and John O'Loughlin. Space-time Granger analysis of the war in Iraq: A study of coalition and insurgent action-reaction. *International Interactions*, 38(5):forthcoming, 2012.

Charles A. McClelland. *World Event/Interaction Survey Codebook (ICPSR 5211).* Inter-University Consortium for Political and Social Research, Ann Arbor, 1976.

Charles A. McClelland. Let the user beware. *International Studies Quarterly*, 27(2):169–177, 1983.

Patrick McGowan, Harvey Starr, Gretchen Hower, Richard L. Merrit, and Dina A. Zinnes. International data as a national resource. *International Interactions*, 14:101–113, 1988.

Paul Meehl. *Clinical and Statistical Prediction: A Theoretical Analysis and a Review of the Evidence.* University of Minnesota Press, Minneapolis, 1954.

Erik Melander and Ralph Sundberg. Climate change, environmental stress, and violent conflict: Tests introducing the UCDP georeferenced event dataset. Presented at the International Studies Association Meetings, Montréal, 2011.

Richard L. Merritt, Robert G. Muncaster, and Dina A. Zinnes, editors. *International Event Data Developments: DDIR Phase II.* University of Michigan Press, Ann Arbor, 1993.

Slava Mikhaylov, Michael Laver, and Kenneth Benoit. Coder reliability and misclassification in the human coding of party manifestos. *Political Analysis*, 20(1):78–91, 2012.

Burt Monroe and Philip A. Schrodt. Editors' introduction: The statistical analysis of political text. *Political Analysis*, 16(4), 2008.

Brian Mooney and Barry Simpson. *Breaking News: How the Wheels Came off at Reuters*. Capstone, Mankato, MN, 2003.

Peter Nardulli. The social, political and economic event database project (SPEED). `http://www.clinecenter.illinois.edu/research/speed.html`, 2011.

National Countterrorism Center NCTC. Worldwide incidents tracking system. `http://www.nctc.gov/wits/witsnextgen.html`, 2011.

Sean P. O'Brien. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review*, 12(1):87–104, 2010.

PITF. Political instability task force worldwide atrocities dataset. `http://eventdata.psu.edu/data.dir/atrocities.html`, 2011.

Clionadh Raleigh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. Introducing ACLED: An armed conflict location and event dataset. *Journal of Peace Research*, 47(5):651–660, 2010.

Rafael Reuveny and Heejoon Kang. International conflict and cooperation: Splicing COPDAB and WEIS series. *International Studies Quarterly*, 40(2):281–305, 1996.

Andrea Ruggeri, Theodora-Ismene Gizelis, and Han Dorussen. Events data as bismarck's sausages? intercoder reliability, coders' selection, and data quality. *International Interactions*, 37(1):340–361, 2011.

Bruce M. Russett, J. David Singer, and Melvin Small. National political units in the twentieth century: A standardized list. *American Political Science Review*, 62(3):932–951, 1968.

Idean Salehyan, Cullen S. Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, Jennifer Williams, Kaiba White, and Sarah Williams. Social conflict in Africa: A new database. *International Interactions*, 38(5):this volume, 2012.

Gerald Schneider, Margit Bussmann, and Constantin Ruhe. The dynamics of mass killings: Testing time- series models of one-sided violence in in the Bosnian civil war. *International Interactions*, 38(4):this volume, 2012.

Philip A. Schrodt. Statistical characteristics of events data. *International Interactions*, 20 (1-2):35–53, 1994.

Philip A. Schrodt. Beyond the linear frequentist orthodoxy. *Political Analysis*, 14(3):335–339, 2006a.

Philip A. Schrodt. Twenty years of the Kansas event data system project. *The Political Methodologist*, 14(1):2–8, 2006b.

Philip A. Schrodt and Deborah J. Gerner. Validity assessment of a machine-coded event data set for the Middle East, 1982-1992. *American Journal of Political Science*, 38:825–854, 1994.

Philip A. Schrodt and Deborah J. Gerner. Kansas event data system. `http://eventdata.psu.edu`, 2010.

Philip A. Schrodt, Glenn Palmer, and Mehmet Emre Hatipoglu. Automated detection of reports of militarized interstate disputes using the SVM document classification algorithm. Paper presented at American Political Science Association, 2008.

Stephen Shellman. Time series intervals and statistical inference: The effects of temporal aggregation on event data analysis. *Security Studies*, 12(1):97–104, 2004.

Michael Spagat, Andrew Mack, Tara Cooper, and Joackim Kreutz. Estimating war deaths: An arena of contestation. *Journal of Peace Research*, 53:934–950, 2010.

START. Global terrorism database. National Consortium for the Study of Terrorism and Responses to Terrorism, 2012. URL `http://www.start.umd.edu/gtd`.

Christopher Sullivan, Cyanne E. Loyle, and Christian Davenport. The coercive weight of the past: Temporal dependence in the conflict-repression nexus. *International Interactions*, 38(5):this volume, 2012.

Philip E. Tetlock. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press, Princeton, NJ, 2005.

Henrik Urdal. Population, resources and violent conflict: A sub-national study of India 1956–2002. *Journal of Conflict Resolution*, 52(4):590–617, 2008.

Henrik Urdal and Kristian Hoelscher. Explaining urban social disorder and violence: An empirical study of event data from Asian and Sub-Saharan African cities. *International Interactions*, 38(5):this volume, 2012.