Analyzing the dynamics of international mediation processes in the Middle East and Balkans

Philip A. Schrodt, Deborah J. Gerner, Rajaa Abu-Jabr, Ömür Yilmaz and Erin M. Simpson

Department of Political Science
University of Kansas
Lawrence, KS 66045 USA
email contact: `schrodt@ku.edu`
project web site: `http://www.ku.edu/~keds`
phone: +1.785.864.9024
fax: +1.785.864.5700

Paper prepared for delivery at the 2001 Annual Meeting of the American Political Science Association, Hilton San Francisco and Towers, 30 August - 2 September 2001.

This research was supported in part by the U.S. National Science Foundation grant SES-0096086 and by the General Research Fund of the University of Kansas. An earlier version of this paper was presented at the Eighteenth Annual Political Methodology Summer Conference, Emory University, July 19-21, 2001 We would like to thank John Freeman and Steve Chan for useful comments on this earlier version. The data and computer programs used for this research are available at the KEDS web site, http://www.ku.edu/~keds.

Abstract

This paper presents initial results from a project that will formally test a number of hypotheses embedded in the theoretical and qualitative literatures on mediation, using automated coding of event data from news-wire sources. Third-party mediation is one of the most common international responses to political conflict. Studies show that mediation was attempted in over half of the conflicts in the post-WWII period; it is likely that the use of mediation has increased following the end of the Cold War. Surprisingly, there have been few systematic studies on mediation. Those that do exist have generally focused on relatively static contextual factors such as the conflict's attributes and the prior relationship between the mediator and protagonists rather than on dynamic factors—both contextual and process—that may contribute to the success or failure of mediation activities. In contrast, the extensive qualitative literature provides numerous hypotheses about dynamic aspects of mediation.

The initial part of the paper focuses on two issues of design. First, we discuss the advantages of generating data using fully automated methods, which increases the transparency and replicability of the research. This transparency is extended to the development of complex variables that cannot be captured as single events: these are defined as patterns of the underlying event data.

Second, we justify the "statistical case study" approach that focuses on a small number of cases limited in geographical and temporal scope. While the risk of this approach is that one will find patterns of behavior that apply only in those circumstances, we point out that the more conventional large-N time-series cross-sectional studies also carry inferential risks.

The statistical tests reported in this paper look at three different issues using data on the Israel-Lebanon and Israel-Palestinian conflicts in the Levant (1979-1999), and the Serbia-Croatia and Serbia-Bosnia conflicts in the Balkans (1991-1999). First, cross-correlation is used to examine the effects of mediation on the level of violence over time; we show that these differ substantially depending on who is mediating. Second, we test the "sticks-or-carrots" hypothesis on whether mediation is more effective in reducing violence if accompanied by cooperative or conflictual behavior by the mediator. The results for these cases generally indicate that reduction in violence is associated with mediation combined with conflictual action directed to the stronger antagonist and cooperative action directed to the weaker antagonist. Finally, we estimate Cox proportional hazard models to assess the factors that influence (1) whether mediation is accepted by the parties in a conflict, (2) whether formal agreements are reached, and (3) whether the agreements actually reduce the level of conflict. The results from these models are more ambiguous, and in a number of instances appear unreliable due to small sample sizes.

Future work in the project involves development of a new event coding scheme specifically designed for the study of international mediation. The current draft of this system is included in an appendix. We also intend to expand our list of cases to include mediated conflicts in West Africa.

1. Introduction

This paper discusses the initial analysis from a project that is examining the dynamics of third-party international mediation using statistical time-series analyses of political event data. Event data—nominal or ordinal codes recording the interactions between international actors as reported in the open press—provide a rich set of indicators about the results of mediation, the political circumstances of the mediation (for example, prior military success or failure by the protagonists), and the various strategies employed by the mediating parties.

The quantitative study of international mediation dates back to the 1960s. The initial work was done by Haas (1967, 1986), who focused specifically on the efforts of international organizations to control conflict through mediation and other active measures such as collective security. This work was later extended by Nye (1968) and Butterworth (Haas, Butterworth, & Nye, 1972; Butterworth & Scranton 1980); the Butterworth also included mediation efforts by individual nation-states and by organizations not set up for collective security. Sherman (1987, 1994; Sherman & Neack 1993; Alker & Sherman 1982) further extended this work in the SHERFACS data set. The CASCON data set developed by Bloomfield and his associates (Bloomfield & Leiss 1969; Bloomfield & Moulton 1997) is another resource dating from this period; it shares many of the concepts of the Haas-Butterworth-Sherman effort, notably the coding of “crisis phase” and the categorization of mediator types. Unfortunately, very little statistical work employing contemporary methods has been done with these data collections—Dixon’s (1996) study using SHERFACS is one of the few exceptions—and they have largely been used for descriptive rather than inferential purposes.

During the 1990s, the most extensive quantitative analysis of mediation has been in the work of Bercovitch and his associates (Bercovitch, Anagnoson, & Wille 1991; Bercovitch & Wells 1993; Bercovitch 1996a, 1996b; Trappl et al. 1997;Wickbolt, Bercovitch & Piramuthu 1999; Bercovitch & Schneider 2000; Bercovitch & Houston 2000). Bercovitch has assembled a data set on mediation efforts for 295 conflicts from 1945 to 1995, and used state-of-the-art statistical methods to test a variety of hypotheses about mediation. This research has also demonstrates clearly that there are testable hypotheses in the qualitative literature and identifies many of the key mediation characteristics of theoretical interest.

The objective of our research will be to shift from the generally *structural* focus of the Haas-Butterworth-Sherman, CASCON, and Bercovitch studies—which examine the characteristics of mediators and the conflicting parties—to an emphasis on the *dynamics* of the mediation process as reflected in news reports coded as international event data. In other words, we will be looking at the impact of variables that change over time. In the qualitative mediation literature, these are generally referred to as "process" variables, although we will also be looking at some dynamic variables that are usually put in the "contextual" category. For example, the relationship between the mediator and a disputant is generally considered a "contextual variable," but it can change at critical moments, as with the December, 1988 decision by the United States to deal directly with the Palestine Liberation Organization (Gerner & Wilbur 2000). We see this research as filling a gap in the literature between the macro-level variables emphasized in the existing quantitative studies and the micro-level advice to individual negotiators that is found in the "wisdom literature" (e.g. Fisher & Ury 1978, Fisher et al 1997) and the case studies.

In general, our dependent variable will be the success or failure of international mediation. However, as Kleiboer (1996) points out, this can be measured in a variety of different ways. We will look the following measures, among others:

❖ Do the disputants openly agree to mediation?

❖ Do the parties formally reach an agreement?

❖ Is the agreement successfully implemented, in the sense that violence is reduced?

These variables capture the main behaviors emphasized in the literature and can be readily coded using event data.

2. Research Design Issues

The general project will focus on the statistical analysis of events in three geographical regions in the recent past—the Middle East (1979-1999), the Balkans (1991-1999), and West Africa (1989-1999). Our analytical techniques will emphasize conventional inferential statistics. The temporally-limited case study emphasis is a departure from the global, multi-century approaches found in much of the quantitative research on international politics, and the use of statistical inference is a departure from our earlier work with computational and algorithmic models; this section will explain these design decisions.

2.1. Statistical Case Studies versus Time Series Cross-Sectional Approaches

Our approach of doing a time-series analysis of selected case studies differs significantly from the fundamental designs used in much of the statistical work in international relations. These either define a type of behavior and then look at all instances of that behavior across a large set of specified actors in a specified time period (e.g. the approach of COW, MID, ICB, the Butterworth-Haas-Sherman mediation study and the Bercovitch mediation studies)[1] or else define a set of behaviors and code them across a large set of specified actors in a specified time period (e.g. the approach of the Polity data set, as well as the WEIS, COPDAB and SherFacts event data sets.)

In recent years, the case study approach—long derided as “slow journalism”—has undergone a rehabilitation among researchers using formal methods. For example, Bueno de Mesquita is his 2001 International Studies Association presidential address put the case study on par with large-sample studies and formal theory as a crucial element in developing a scientific understanding of international behavior.

> One path to insight is the detailed analysis of individual events; the method that today we call the case study. This technique, often relying on archival research, proves to be a fertile foundation from which new and interesting ideas germinate, ideas that suggest hypotheses about regularities in the world that are worthy of being probed through close analysis of individual events and through careful repetition across many events. The close probing of case study analysis enhances the prospects of achieving verisimilitude as it brings the proposed explanation into close proximity with the known details of the situation. It does not, however, provide evidence that the specific details are germane to other, similar occurrences. (Bueno de Mesquita 2001,2)

Bueno de Mesquita’s approach echoes that of other recent works on political science methodology—notably King, Keohane & Verba (1994) and Van Evera (1997)—that have also placed the properly-designed case study firmly in the realm of scientific studies.

[1] A variant found in event data projects—for examples BCOW and CASCON—is to sample the cases rather than analyzing the entire population.

Our project differs from the usual interpretation of “case study” in that we primarily will be using statistical, rather than interpretative, methods. In this respect it is closest to the various studies of Goldstein (e.g. Goldstein and Freeman 1990, Goldstein and Pevehouse 1997; Goldstein et al forthcoming), though it also is similar to a number of other dyad-specific event data studies, for example Mooradian & Druckman 1999, Moore 1995, Somer & Scarritt 1998, Thomas 1999, and Ward & Rajmaira 1992. In making the choice to use a time series rather than a cross-section, we were motivated by three considerations.

The first is the issue of data quality. The Middle East and the Balkans are probably the most thoroughly reported conflicts in human history. As we note below, this does not mean that we have a “god’s eye view,” but subject to that caveat, the data don’t get any better than this. These data sets each contain tens of thousands of events, so aggregating to the monthly or even weekly level for time series analysis is feasible. As a control, we will also be analyzing conflicts in West Africa, where we know that journalistic coverage is very poor. We also have extensive field experience for the Middle East case, and we have access to individuals at the University of Kansas who have extensive field experience on the West Africa case.[2]

The second issue deals with the variance in the independent and dependent variables. Within each geographical region there are a number of distinct (though inter-related) conflicts, so while we are considering only three geographical foci, there are at least a dozen different conflict/mediation sequences within those areas. A wide variety of different approaches to mediation have been attempted, with a variety of different outcomes. The Middle East is arguably the most mediated conflict of the post-WWII period—quite possibly to the point of diminishing returns—as every U.S. Secretary of State since at least Kissinger has seemingly felt obligated to spend a disproportionate amount of time engaged in the region. The Balkans, in contrast, witnesses a period of ineffective mediation prior to 1995, followed by the thus-far successful Dayton Agreement that halted the violence (but did not necessarily resolve the conflict), followed by a renewed conflict and subsequent intervention in Kosovo-Macedonia. West Africa largely has seen regional rather than super-power mediation, again with mixed results.

---

[2] and inadvertently found ourselves in the middle of a Senegal-Mauritania border dispute last summer

Finally, we are focusing on a limited number of regions and a relatively limited period of time in order to control, at least in part, for the effects of cultural and historical context. This is probably the most controversial aspect of the case study approach, and requires some justification.

Studies that cover a long period of time and a large number of cases are, presumably, seeking to find very general "laws" (or at least correlations) that hold across all of those times and places. This approach has been part of the behavioralist agenda from its earliest days, and is based on the classical agenda of post-World War II realist writers such as Morgenthau and Kissinger, as well as the contemporary interpreters of the classical works of Thucydides, Sun-Tzu and others.

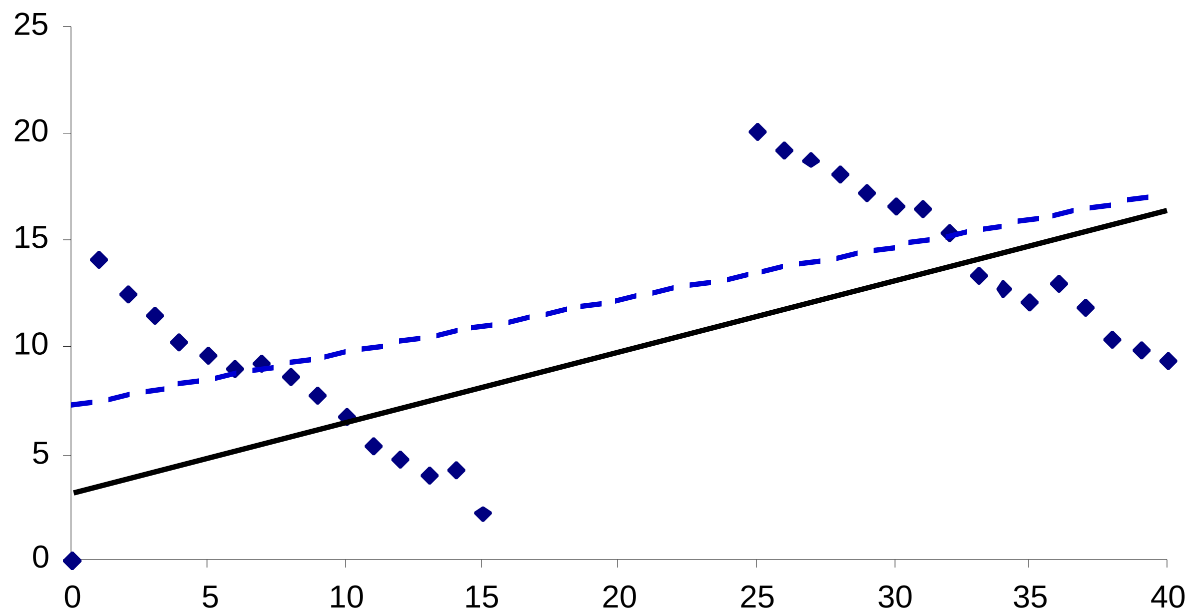
The most common argument against the large-scale approach is that the focus on law-like generalizations that hold across very large numbers of cases bypasses a variety of useful generalizations that apply only in more limited (but non-trivial) times or places. Conrad and Schlichte, summarizing the twenty years of experience of the *Arbeitsgemeinschaft Kriegsursachenforschung*, (Study Group for the Causes of War) at the University of Hamburg, note:

> The underlying idea within the mainstream of quantitative research on wars is to isolate ‘factors’ that contribute to the outbreak of war or make warfare more likely. … The reason for our lack of enthusiasm today is that such a universal modelling of factor relations and interactions does not take into account that differences in institutional settings and historical times are myriad and can only be included into models of high complexity and tremendous scope which are no longer of practicality. ... Although they sometimes hint towards an interesting relation between a specific ‘factor’ and the outbreak of war, plenty of other, disturbing ‘factors’ render the result unhelpful for practical purposes.
>
> … Quantitative research on wars assumes that there are generally no built-in differences between historical epochs, between different ‘logics of action’ (*Handlungslogiken*) of state leaders and other relevant personnel in divergent historical settings, or of different historical formations as such. Instead of adapting the method to the object of investigation, i.e. war at various points of time and manifold locations, the object is subdued to a methodology that might be clear and rigid, but does not follow the historical development (*Formwandel*) of the causes of political violence. (Conrad and Schlichte 2001; emphasis in original).

The number of statistical generalizations about international behavior that have held up in time-series cross-sectional studies is very small: The link between borders and wars is probably the only "law" that has near universal statistical support, and this is scarcely an generalization that should have required massive amounts of data collection and computer power to establish. The democratic peace hypothesis has generated a huge literature, but much of that consists of statistical studies refuting the hypothesis (or contradicting earlier studies); the statistical literature on the "diversionary" hypothesis is equally ambiguous. Most of the "laws" that were asserted in the classical literature to hold across multiple cultures and times—for example linkages between various alliance or power configurations and the likelihood of war—have failed to withstand statistical scrutiny. While this statistical brush-clearing has some utility, it primarily shows that the classical literature has been no better than the scientific in developing non-trivial generalizations. There may not be a pony in there.

It gets worse. Correlational studies across multiple, distinct sub-sample may produce results that are actually misleading. Consider for example the case in Figure 3.1. The true relationship within two of the subgroups is Y = -07.X + c except for a cluster of points—a third of the sample—at the origin.[3] Yet the slope of the regression line estimated on the complete set of data is positive and highly significant.

Figure 3.1. The risks of correlation across sub-populations



Solid line: N = 45; r = 0.72 (signif < 0.001); slope = 0.33 (t = 6.81, signif < 0.001)

Dashed line: N = 30; r = 0.42 (signif = 0.018); slope = 0.15 (t = 2.50, signif = 0.018)

While Figure 3.1 has been artificially constructed to illustrate our argument, the possibility of a pattern such as this is hardly implausible. Let the Y-axis be a measure of conflict and the X-axis some variable measuring mediation efforts. The group of points on the left are low-conflict cases with low levels of mediation, the group on the right are high-conflict cases with higher levels of mediation, and there are a cluster of cases with no conflict and no mediation. Estimation on the entire sample would indicate that the mediation is counter-productive.

This is not just an effect of the cluster of points at the origin. The dashed line shows the regression that results when these are eliminated; while less dramatic than the first case, the relationship is still significant at the 0.02 level and still shows a positive relationship between mediation and conflict. This artifact would be obvious in the two-dimensional case illustrated here but could very easily be lost in a more typical large-N study where many independent variables are used and where the high-dimension space is impossible to visualize.

Our point with Figure 3.1 is to indicate that in a world where sub-populations exist—and this is almost certainly a characteristic of the world we are studying—large-sample studies are not risk free. Our approach therefore has been to start by looking at some cases that we know very well, and which by virtue of limited time and limited geography are relatively homogenous. Not completely homogeneous—we recognize fully that there were critical differences between, say, the Lebanese civil war and the Palestinian *intifada*, or between Serbian attacks in Bosnia and Kosovo—but we would argue that these still have more in common than, say, the Chaco War and the Cuban Missile Crisis. As we find relationships that hold in some or all of these conflicts, we can then extend our analysis to others.

2.2. Statistical Methods versus Computational Pattern Recognition

Much of the prior work in the Kansas Event Data System (KEDS) project has involved the development (or adaptation) of computational methods for the analysis of event data. Generally, these methods have come out of the algorithmic pattern recognition literature—for example ID3 (Schrodt 1991a); genetic algorithms (Schrodt 1989); neural networks (Schrodt 1991b); cluster analysis (Schrodt & Gerner 1997; 2000); and hidden Markov models (Schrodt 1999, 2000).

---

[3] Equations generating Figure 3.1:
$x \quad \{1, 15\}$: $y = 16 - 0.75x + e$; $x \quad \{25, 40\}$: $y = 41 - 0.75x + e$; $e \sim$ Uniform(-1,1); 15 cases at (0,0)

We adopted this approach for several reasons: Pattern recognition was strongly supported by the theoretical literature on political decision-making, many of the pattern recognition algorithms could be employed without the arbitrary intermediate step of scaling the event data into interval-level measures, and with a few exceptions, most of the statistical methods used with event data prior to 1990 were very crude, often little more than contingency table analyses. Nonetheless, despite our rather extensive investment in algorithmic methods, we are currently inclined to abandon that approach and return to conventional statistics. This change in approach is motivated by four factors:

First, while there has been some additional use of computational methods to analyze political behavior—for example neural networks are used by King and Zeng (Beck, King & Zeng 2000), genetic algorithms by Sekhon and Mebane (1998) and classification methods by some artificial intelligence researchers (Wickbolt, Bercovitch & Piramuthu 1999; Kovar et al 2000)—computational pattern recognition is still not widely employed in the political science literature. Due to the required investment in specialized or custom-written software, these approaches are difficult to use without a substantial knowledge of computer programming.

Second, ten years of experimentation with algorithmic techniques failed to demonstrate dramatic advantages sufficient to offset the computational costs and computer programming involved. Event data are noisy and generated by processes that have a large stochastic component, and any estimates based on event data will necessarily have a substantial amount of error. Given that many computational methods require huge amounts of computer time to get oftentimes indeterminate results, the comparative advantage of those methods are not clear. (Computer time itself is inexpensive, but the time of the human analyst awaiting those results is not.)

Third, and probably most important for the purposes of this project, computational pattern recognition algorithms lack a clearly defined inferential mode. Because this project is evaluating hypotheses from a rich, if inconsistent, theoretical literature, inference is our primary concern. The level of sophistication in the time series techniques found in political analysis has increased dramatically in recent years (see, for example, King 1989; Beck & Katz 1995; Box-Steffensmeier & Jones 1997; Beck, Katz & Tucker 1998; Bennett 1997, 1999) . Consequently, while time series models do not fit perfectly to the theoretical explanations for the success and failure of mediation, the inferential power of these methods far outweighs the sacrifices one may need in terms of explanation.[4] Because these methods can be implemented with existing statistical packages such as *Stata* and *SAS*, we will be able to focus most of our efforts on analysis rather than software development.

We are still left with the challenge of figuring out how to analyze sequences. Most of the existing time-series methods were designed to study interval-level data reported at regular time intervals (for example, GDP, stock prices, or unemployment rates). An event sequence, in contrast, consists of nominal-level variables reported at irregular time intervals. Event data are further complicated by the fact that events occur between pairs of actors ("dyads") and, as we will note below, the sequencing of events within a single day is indeterminant. We do not regard these problems as insurmountable—for example there is already a sizeable time-series literature that converts event data sequences to interval-level data through scaled aggregations, and duration models such as the Cox proportional hazard model deal effectively with irregular reports—but the fit between the available data and the available methods remains less than perfect.

3. Transparency and Pattern Recognition

One of the objectives of our analysis will be to raise the level of transparency and replicability to the highest level possible. Machine coding already provides this with respect to generating the basic event data, and our objective is to extend that to other variables as well.

We originally became involved with machine coding because, after initial start-up costs, it is dramatically faster and less expensive than human coding. Once a researcher has established vocabulary lists of actors and verb phrases, the only significant expense involved in generating event data is the acquisition of machine-readable news reports. Furthermore, a coding system developed at one institution can be used by other researchers through the sharing of vocabulary lists and coding software.

In working with KEDS, we discovered an additional advantage to machine coding: It is free of non-reproducible coding biases and is therefore both reliable and transparent. Human coding is

---

[4] This emphasis on inferential statistics is also appropriate given the differences between our earlier focus on prediction (e.g. Schrodt & Gerner 2000; Schrodt 2000) and the explanatory nature of this project. Good predictive models do not necessarily involve good explanations; in fact when models with diffuse parameter

subject to systematic biases because of unconscious assumptions made by the coders. For example, Laurance (1990) notes that even expert coders in the military tended to over-estimate the military capability of China in the 1980s because they knew China to be a large Communist country. When event coding is done part-time by students, coder biases are even more unpredictable and difficult to control.

In contrast, with machine-coding the words describing an activity will receive the same code irrespective of the actors or time period involved. Any biases embedded in the machine coding system are preserved explicitly in an index of its vocabulary such as[5]

```
092             "ASK POLICY AID"
    $ *MUSTER SUPPORT
    *REACH OUT TO FORMER ALLIES
    *SEEK SUPPORT
    *SOUGHT SUPPORT
    *WANT CLOSER TIE
    + WILL *ASK INVESTIGATOR
    ANGLING FOR *HELP
    JOIN *SEEK MEDIATION
    LOBBIED
    SAID MUST *PERSUADE
```

Human coding produces no such record beyond the codebooks indicating the rules that the coders were supposed to be implementing.

By analogy, human coded events are similar to the summary of an open-ended interview, where the process by which the information was extracted from the respondent varies from interviewer to interviewer, and probably respondent to respondent, and the archival record does not retain all of the information. Automated coding, in contrast, is similar to a survey instrument with a fixed set of questions that are preserved along with the data. While the fixed-question format has limitations with respect to the information that can be obtained, and questions may be context dependent—for example, a 1960s question about fallout shelters or the problem of

---

structures are used (for example hidden Markov models, neural networks, or VAR), there may actually be a tradeoff between effective prediction and coherent explanation.

[5] The example give here is a partial index of the phrases used by in a KEDS/TABARI dictionary to code the WEIS 092 category. The “*” preceding a word indicate that it is the “verb” that is being coded in the phrase; in some cases this verb is actually being used as a noun in the phrase, as in “angling for help” and “said must persuade”. At the conclusion of a coding session, the TABARI system can also produce an annotated list of the dictionaries that shows how many times a phrase was actually used to generate an event; this allows an analyst to determine which phrases are actually being found in the texts.

Communist influence in labor unions would probably produce only puzzled looks from most respondents in 2001—the stimulus is known exactly and can be preserved with the data.

We believe that transparency is especially important in the study of mediation because of the potential problem of "hind-sight bias": knowing the outcome of a mediation effort can potentially affect how informed coders assign values to the independent variables. This is an unavoidable risk in human-coded data. But our emphasis on coding transparency is a substantial departure from the data-generation work in international politics that has emphasized the importance of human coders having an understanding of a situation's full historical context before categorizing a case. In the absence of machine coding, this made sense—the only thing worse than having a stupid machine assign codes is having a stupid human assigning codes, since humans (unlike machines) have biases and preconceptions. But now that automated coding is available as a data generating method, it makes sense to eliminate the human as an uncontrollable source of error.

This contention that we have eliminated the human factor from the *coding* of the texts has led some critics to assume that we are contending that we have created an "objective" view of the world. Nothing could be further from the truth: Because we have done extensive field work on our primary case, we are acutely aware that any source of reports—whether Reuters, *Agence France Presse*, *The New York Times*, CNN, FBIS, *al-Fajr* or *Ha'aretz*—is selective. We've been there, watched the sausage being made, and it ain't pretty. No news source, or combination of news sources, provides the "god's eye view" of events on the ground. Machine coding from a given set of *texts* merely eliminates the *additional* biases introduced by the coder.

Having eliminated irreproducible human factors at the stage of coding events from a given set of texts, it seems appropriate to also eliminate these "downstream" in the construction of more complex variables as *patterns* of events. This goes back to McClelland's (1970) original assumption that event data would break down complex political activities into a sequence of basic building blocks (e.g., comments, visits, grants, rewards, protests, demands, threats, and military engagements) from which more complex political activities were constructed..

Patterns are central to the entire issue of sequence analysis, and patterns more generally are at the core of analyzing categorical data. Clinick, commenting on the most well-developed contemporary pattern specification system, the "regular expressions" of awk, sed, perl and other Unix utilities (Wall, Christiansen & Orwant 2000), noted

> One of Perl's key features as a language is regular expressions; in fact, Perl has probably done more to evolve regular expressions than any other language. If you are not familiar with regular expressions, think of them as the ultimate string manipulation tool for serious string processing. Regular expressions are to strings what math is to numbers.
>
> (Andrew Clinick, Microsoft Program Manager, January 22, 1999. http://msdn.microsoft.com/workshop/languages/clinic/scripting012299.asp; accessed 18 December 2000)

Given that a string of text is nothing more than a categorical sequence, in principle regular expressions could be to sequences what math is to numbers, and a pattern could be specified using a regular expression.

Unfortunately, we can't quite do this with sequences of event data because of two factors involving calendar time. First, the precision of the time measure in machine-coded event data is—at best—accurate to about a day or two, and the sequencing of events within that period is indeterminate. In other words, if events A, B, C occur on 5 Jan 96, they could appear in the sequence as either A-B-C, A-C-B, B-C-A, B-A-C, C-A-B, or C-B-A, and there is no substantive difference between these.[6] Second, the passage of time itself may be substantively important—a sequence of events for a dyad might experience two consecutive uses of force, but the substantive interpretation of this will probably be different depending on whether those events were separated by a day or by a year.[7]

The absence of a compact notation for event sequence patterns does not, however, mean that these cannot be specified unambiguously. In our project, we have implemented these as a series of relatively simple C programs that operate on the original stream of event data, detect various

---

[6] While the accuracy of sequencing is probably slightly worse in machine-coded data than in the highest-quality human-coded data, we believe that getting calendar precision finer than a day is virtually impossible with wire service data. There are three major problems. First, events such as meetings and military clashes occur over an *interval* of time, and one would need to decide whether the event was coded when it began, ended or somewhere between. Second, information on the time that an event occurred is frequently missing from news reports, and tracking it down would be very time consuming (and in many cases, impossible). Third, the system would need to adjust for time zones—this is technically possible but requires locating the geographical location of an event to greater precision that we are currently doing and in some cases (such as announcements) this information will not be reported.

The dates in the KEDS project data sets use the date of the *report* of the event, which is unambiguous and usually—but not always—occurs within 24 hours of the actual event. The existing system does not deal with temporal modifiers such as "yesterday," "last week," or "tomorrow", and we are hoping to add this capability in the near future.

[7] A partial solution to this problem is to pad the sequence with "non-events" whenever no interaction occurs: the work we have done with hidden Markov models (Schrodt 1999, 2000) makes extensive use of this technique.

patterns, and then produce a new file that can be analyzed by a statistics package. The pattern-recognition routines within these programs are only a few lines in length, and the programs themselves preserve the “coding rules” by which the variables were generated.

Event data provide an extremely rich set of potential variables for the analysis of mediation activities. Most of the information considered theoretically relevant to the mediation “process” can be coded from event data (as long as the information is reported in some news source), as can quite a few of the “contextual” variables. These include information on the chronology of the conflict, changes in the relations between potential mediators and the protagonists, the initiation and cessation of formal negotiations, and the level of violence between the disputants. Figures 3.1, 3.2, and 3.3 show some examples of how a complex behavior can be derived from a specific pattern of events (as well as showing the general patterns of scaled conflict in the cases we will be analyzing). The “mediation pattern ” was defined as cooperative behavior (WEIS cue categories 01 through 10) between a designated mediator (for example, the USA, UN or EU) and both parties in the dyad that occurring within a period of seven days. While this is not a sufficient condition for mediation—a representative of a state might visit multiple parties to a conflict without trying to mediate—it is probably a necessary condition (any mediation will involve such cooperation, at least within the limitation of the news reports)

Figures 3.1 and 3.2 show both the Goldstein-scaled (Goldstein 1992) monthly aggregations of events of Israel to the Palestinians and Israel to Lebanon, as well as the frequency of “mediation events.” This measure of mediation activity tracks the historical record fairly well. The Israel-Palestinian dyad receives mediation efforts almost continuously except during the 1983-1988 period, with conspicuous spikes corresponding to events such as the 1982 Lebanon invasion, US resumption of formal negotiations with the PLO in 1988, and various agreements in the Oslo process. In contrast, mediation in Lebanon tends—necessarily but not sufficiently—to coincide with periods of violence.

Figure 3.1. Israel-Palestinian Cooperation and mediation

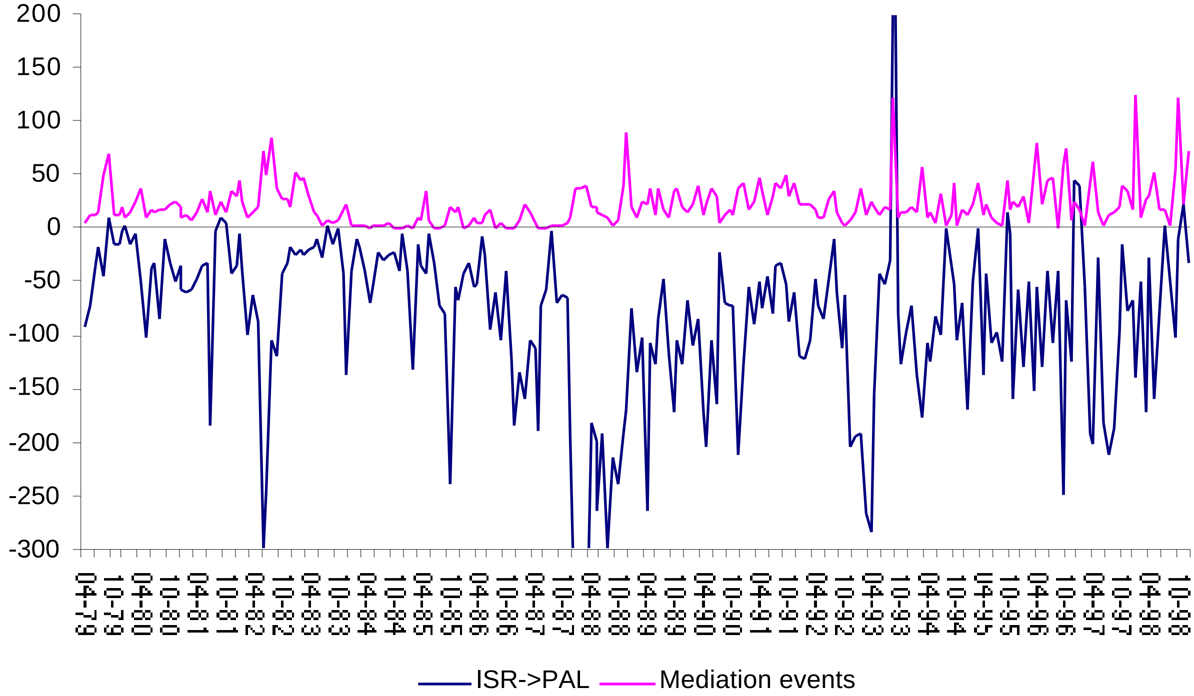


Figure 3.3. Israel-Lebanon Cooperation and mediation

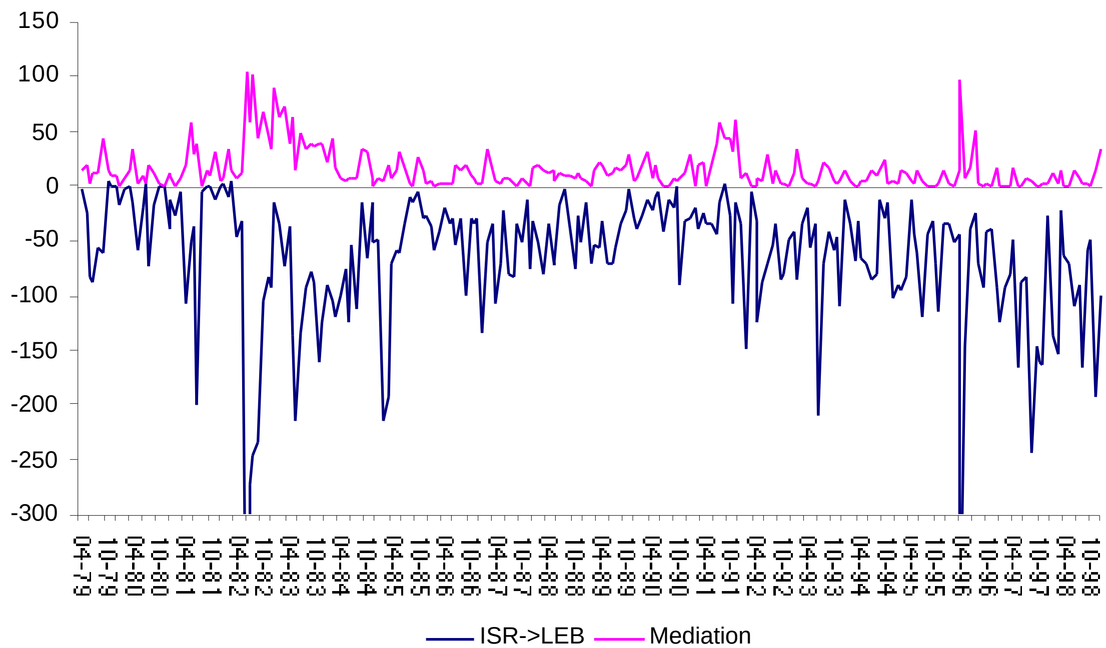
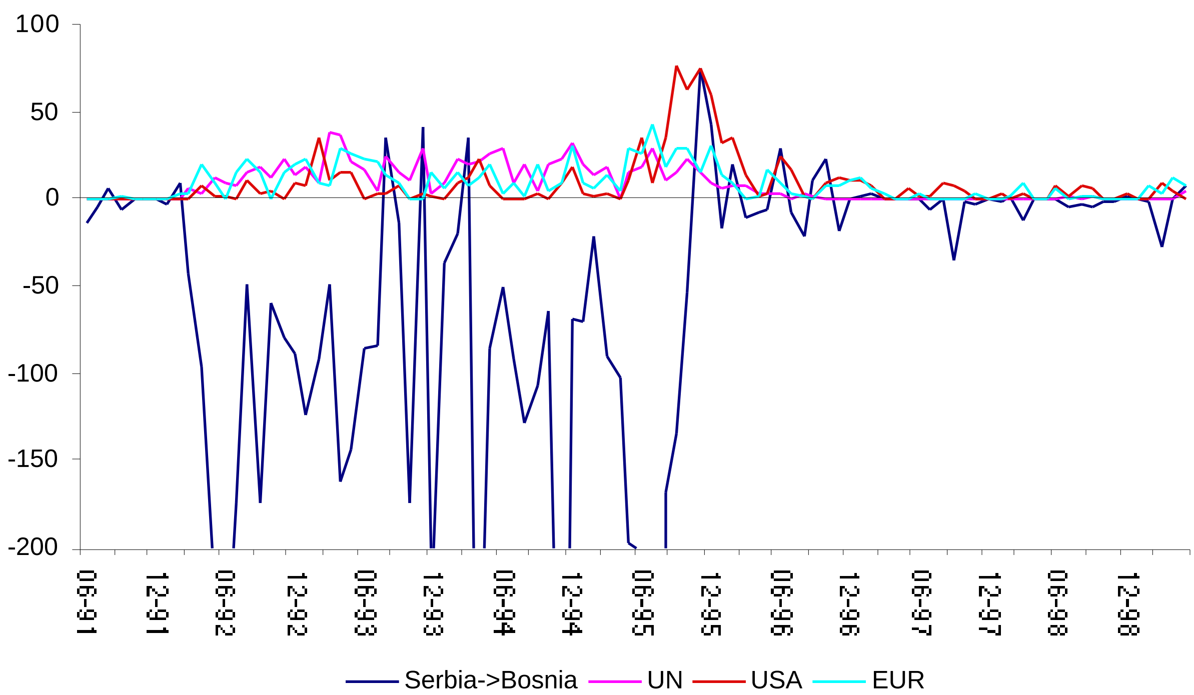


Figure 3.3 shows Balkans mediation and the Goldstein-scaled net cooperation from Serbia to Bosnia; in this graph Bosnia Serbs are included in the "Serbian" activity, although in the original event data they are assigned a distinct code. The major periods of conflict are evident, as is the period of the Dayton agreement. The dyadic summary—correctly—does *not* show conflict during the periods when the major source of regional military activity was between Serbia and Croatia in 1991 or Serbia and Kosovo in 1998.

The lines labeled "UN", "USA" and "EUR" count the number of mediation events that involved the United Nations, United States and major European states (plus the EU) respectively. As we will show below, cross-correlation tests show substantial differences between the effects of the three mediating groups: UN efforts were associated with subsequent increased levels of conflict; United States efforts were associated with decreased levels; and there was no discernible change following European efforts.

Figure 3.4. Serbia-Bosnia Goldstein values and mediation



4. Analysis

4.1. Data

The data used in this study were coded into the WEIS scheme (McClelland 1976; also see Appendix II) using the Kansas Event Data System (KEDS), a computer program that creates event data from machine-readable text.[8] KEDS is a pattern-matching system that uses a computational method called "sparse parsing." Instead of trying to decipher a sentence fully, KEDS determines only the parts required for event coding—for instance, political actors, compound nouns and compound verb phrases, and the references of pronouns—and then employs a large set of verb patterns to determine the appropriate event code.

The events were coded from Reuters News Service lead sentences obtained from the NEXIS data service for the period April 1979 through May 1997 and the Reuters Business Briefing service for June 1997 through September 1999. The lead is usually a simple declarative sentence that summarizes the article, e.g., "The United Arab Emirates welcomed a resumption of formal diplomatic ties between Egypt and Syria after a 12-year rift." For closely reported crisis areas such as the Middle East and the Balkans, lead sentence coding provides thorough coverage of political events. The coding software, coding dictionaries and data developed by the project are available at the KEDS web site, `http://www.ku.edu/~keds`.

The cases evaluated are the Israel-Lebanon and Israel-Palestinian conflicts in the Levant, and the Serbia-Croatia and Serbia-Bosnia conflicts in the Balkans (Serbia-Kosovo is included in some of the proportional hazard models). The Levant data covers April 1979 to September 1999; the Balkans data cover January 1991 to April 1999.

The scaled data uses the Goldstein (1992) scale at monthly aggregations—the scaled events are totaled for each dyad-month. When event counts are analyzed, we use the following categories based on the WEIS 2-digit "cue categories":

---

[8] Discussions of machine coding can be found in Bond et al 1997, Gerner et al 1994, Schrodt & Gerner 1994, Huxtable & Pevehouse 1996, and Schrodt, Weddle & Davis 1994. Refereed research employing machine-based event data include studies of triangulation and reciprocity in the Balkans (Goldstein & Pevehouse 1997) and Middle East (Goldstein et al forthcoming), foreign policy decision making (Wood & Peake 1998), early warning systems of political instability (Schrodt & Gerner 1997, 2000, Schrodt 2000) and studies intrastate civil conflicts (Bond, et al 1997, Huxtable 1997, Thomas 1999a)

vercp: Verbal cooperation—WEIS categories 02, 03, 04, 05, 08, 09, 10

matcp: Material cooperation—WEIS categories 01, 06, 07

vercf: Verbal conflict—WEIS categories 11, 12, 13, 14, 15, 16, 17

matcf: Material conflict—WEIS categories 18, 19, 20, 21, 22

This reduces the total number of event categories that can be used as independent variables to something manageable. It is also likely to reduce the effects of coding error somewhat: Several of the "verbal conflict" codes in WEIS are ambiguous even for human coders, and the automated coding probably generates some misclassification in those categories.

We consider ten different cases of third-party mediation:

Abbreviation	Actor A	Actor B	Mediators
ISRLEB	Israel	Lebanon	USA, EU, UN
ISRPAL	Israel	Palestinians	USA, EU, UN
ALLBFR	Serbia	Bosnia	[all mediators listed below]
USABFR	Serbia	Bosnia	USA
EURBFR	Serbia	Bosnia	EU, France, Germany, Italy, United Kingdom
UNOBFR	Serbia	Bosnia	UN
ALLCRO USACRO EURCRO UNOCRO	Serbia	Croatia	[same sets of mediators as Bosnia cases]

The dyadic variables have the prefix *m2a* for events with any of the mediators as the source and actor A as the target; *m2b* for events with any of the mediators as the source and actor B as the target; and *a2b* and *b2a* are events from Actor A to Actor B, and Actor B to Actor A respectively. So, for example, in the EURBFR case, *m2amatcp* is the number of events with either the EU, France, Germany, Italy, or United Kingdom as the actor, Serbia as the target, and events in the "material cooperation" categories. Additional pattern-based variables will be defined below.

4.2. Cross-Correlation

Our first set of tests involves cross-correlation (see Appendix I) of the mediation measure with the total level of conflict, defined as the negative of the sum of the Goldstein-scaled net-cooperation scores $A \rightarrow B$ and $B \rightarrow A$ (i.e. high values imply high levels of conflict). Event scores are aggregated by month. The objective of the cross-correlation test is two-fold. First, we initially used it as an empirical "plausibility probe" to demonstrate that non-trivial results can be obtained from this event data using a pattern-based definition of mediation (Gerner and Schrodt 2001).[9] Second—and more generally—cross-correlation should be one of the first steps one used when examining data where the timing of the effect of a variable is not clear from the theory.

In our case, the mediation literature certainly allows for the possibility that there will be some lag between use of mediation and the impact of that mediation on the level of conflict, but "common sense" would allow this lag to be anywhere from a few weeks to a few months, and it might also differ between regions. While some agreements—notably ceasefires—are supposed to be implemented immediately, many others—for example, disarmament, territorial disengagement and deployment of peacekeeping forces—involve substantial negotiated delays, and these are of varying lengths. In addition, the qualitative literature is full of assertions about de-escalation processes taking time to "take hold" due to factors such as continued hostilities by groups opposed to the peace process, wariness by the population that hostilities have actually ended, gradual repatriation of refugees and reconstruction of infrastructure, and other time-consuming processes. Many of these elements have a strong stochastic element.

The indicator of mediation: *the number of instances where the mediator has a cooperative interaction (WEIS categories 01 through 10, excluding comments) with both sides of the conflict within a period of 7 days*.[10] This pattern does not guarantee that the third party is actually engaged in mediation—and our future work will use more precise measures—but almost all

[9] The figures below are slightly different than those in Gerner & Schrodt (2001). Between that analysis and this one, we (a) eliminated NATO and added individual European states to the ALL… mediator group; (b) modified the mediation pattern to exclude comments (WEIS 02 cue); and (c) found the inevitable minor bug in the program used to count mediation events. The Goldstein measure was also changed from net cooperation to net conflict, so the signs of the correlations reverse. None of these differences turn out to be important, but this list does give one a sense of the [frightening] number of more or less arbitrary decisions that go into such an analysis. The assortment of C, perl, and Stata 6.0 programs used to generate these figures are available from the authors.

mediation activities will satisfy this criterion. In other words, this measure provides a necessary but not sufficient indicator of mediation activity.

In the cross-correlation diagram, the values to the left of zero (the center of the graph) are the correlations with mediation activity and cooperation between the antagonists *prior* to the mediation; the values to the right of zero are the correlations with mediation activity and cooperation *following* the mediation. If mediation is successful at reducing conflict, we would expect to see a positive correlation between mediation events at time t and cooperation at time t+k in these figures. The dotted lines on the correlograms show the critical values at the two-tailed 5% significance level; these were determined by Monte Carlo simulation. The cross-correlation approach is discussed in greater detail in Appendix II.

Levant

In the analysis of the Levant case, we looked at mediation efforts involving either the USA, UN or European Community/Union. Most of this activity, unsurprisingly, involves the USA: of the 95,464 events in the data set, 22,752 (23.8%) involve the USA as actor or target; 6,186 (6.5%) involve the UN, and only 579 (0.6%) involve the European Community or European Union.[11] Because we looked at interactions involving any of these actors, a meeting between UN officials with Palestinians followed five days later by a meeting between US officials and Israelis would count as a mediation effort. This is imprecise but probably still a reasonable approximation. UN involvement is far more likely in the Israel-Lebanon case than in the Israel-Palestinian case.

Figure 4.1 shows the cross-correlation of the mediation indicator with Israel-Palestinian conflict measure. The correlogram shows a very distinct pattern of positive correlations for lagged values of cooperation and zero or negative correlations for cooperation in the period following the mediation. In other words, mediation correlates with the level of conflict in the months before the mediation, and correlates with increased cooperation following the mediation. The levels of correlation are relatively low and the highest correlations are barely significant at the

---

[10] We did a few tests using an interval of 4 days; this made no discernible difference in the results.

[11] In retrospect, there was little point in including the EU as a possible mediator, but the analysis had already been done by the time these aggregate statistics were calculated.

5% level (see Appendix), but the overall pattern is quite regular.[12] While the individual cross-correlations in the period following mediation are not significant at the two-tailed 5% significance level, the overall pattern is significant: For example, the 5% critical value for the minimum absolute value of three *consecutive* cross-correlations is around ±0.05—this value was determined by Monte Carlo approximation—and the correlations satisfy this for k>7.

Figure 4.1. Cross-correlation of mediation and conflict in the Levant

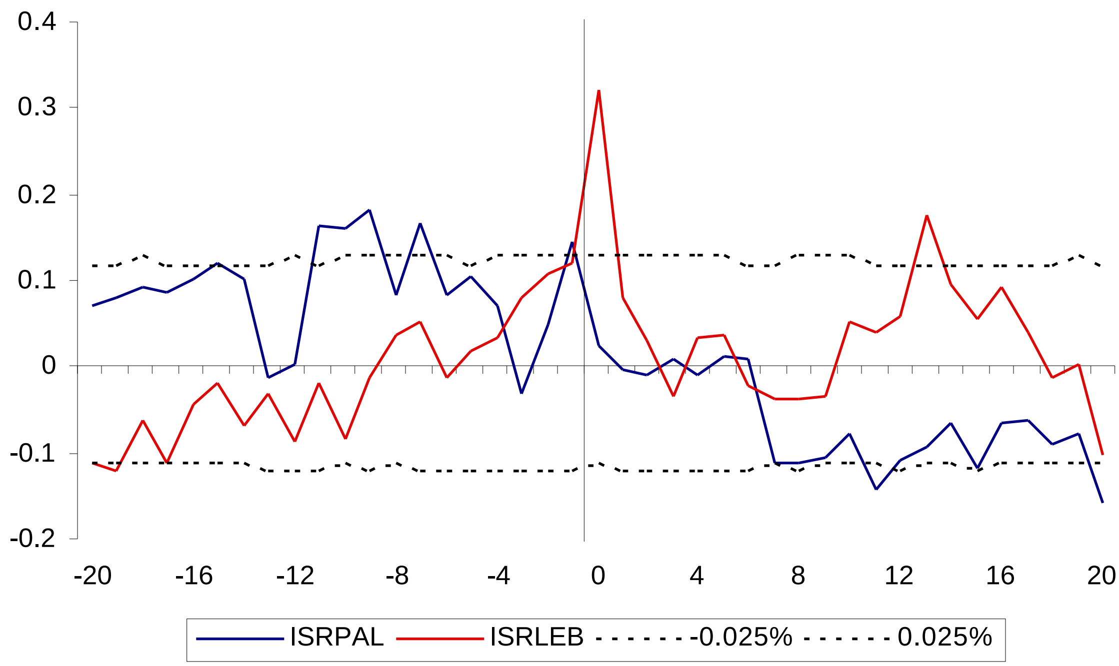


Figure 4.1 also shows the correlogram for mediation and Israel-Lebanon cooperation. This shows a very different pattern that the Israel-Palestinian case. The strongest correlations are contemporaneous—roughly 2 months before and after the mediation—and positive, indicating the mediation is most likely to occur when the level of conflict is high. However, the correlogram

[12] As a check that this pattern is actually measuring mediation and not just interactions, we also ran a cross-correlation between Goldstein-scaled cooperation from the USA→Israel and Israel→Palestinian cooperation. The Goldstein scaled score differs from the mediation score because it only measures interactions between the USA and Israel, without adjusting for whether the US is talking (or otherwise cooperating) with both sides, and also takes into account both positive and negative interactions (e.g. US criticism of Israel). The resulting correlogram—which can be viewed on the KEDS web site (http://www.ukans.edu/~keds/ISA01.supplement/ISA01.Supplement.html) —is quite different than Figure 4: it shows the typical positive spike of contemporaneous correlation at -1, 0 and +1 months, but otherwise the correlation is flat and close to zero. We conclude from this that the mediation indicator is picking up something more than simple interaction.

gives no evidence that the mediation is effective: the correlations between mediation and subsequent cooperation remain near zero or slightly positive. Conflict in the Israel-Lebanon produces mediation efforts, but these have no results.

Balkans

Our analysis of the conflict in the former Yugoslavia looked at three different sets of mediators: the United Nations, the United States, and Europe (operationalized as the EU, France, Germany, Italy and the United Kingdom).[13] The measures for the actors include the activities of the various ethnic factions combined with those of the governments. In other words, "Serbia" includes the actions of ethnic Serbs in Bosnia and Croatia as well as the actions of the Serbian government. As with all event data, the identification of the ethnicity of individuals or groups responsible for actions was dependent on how the event was reported in the news story.

The anecdotal accounts of the conflict suggest that the effectiveness of these efforts varied substantially depending on who was doing the mediation (see Kaldor 1999: 31-68; Weiss 1999: 97-136), a proposition supported by our cross-correlation analysis. Figure 4.2 shows the cross-correlogram of the various mediators and the level of conflict. The three mediators show quite different patterns.

The correlations for the UN are significantly positive both prior to and following the mediation. In other words, the UN mediation increased during periods of increased conflict in the dyad, but in contrast to the pattern seen for mediation in the Levant, the level of conflict actually become *greater* following the mediation. US mediation efforts, in contrast, had a positive effect on cooperation: there is a positive correlation with conflict prior to the mediation, but a significant negative correlation—that is, mediation correlates with decreased conflict—in the period following the mediation.[14]

---

[13] We also analyzed mediation by Russia and Ukraine. This series has substantially less variance than the European mediation series (17.8 versus 97.6) but generally shows a pattern similar to that of Europe.

[14] While the individual correlations for the US are barely significant at the 0.05 level, the critical value for the minimum of three consecutive correlations is ±0.07—see Appendix I—and the U.S. pattern clearly satisfies this criterion.

Figure 4.2. Cross-correlations of mediation and Serbia-Bosnia conflict

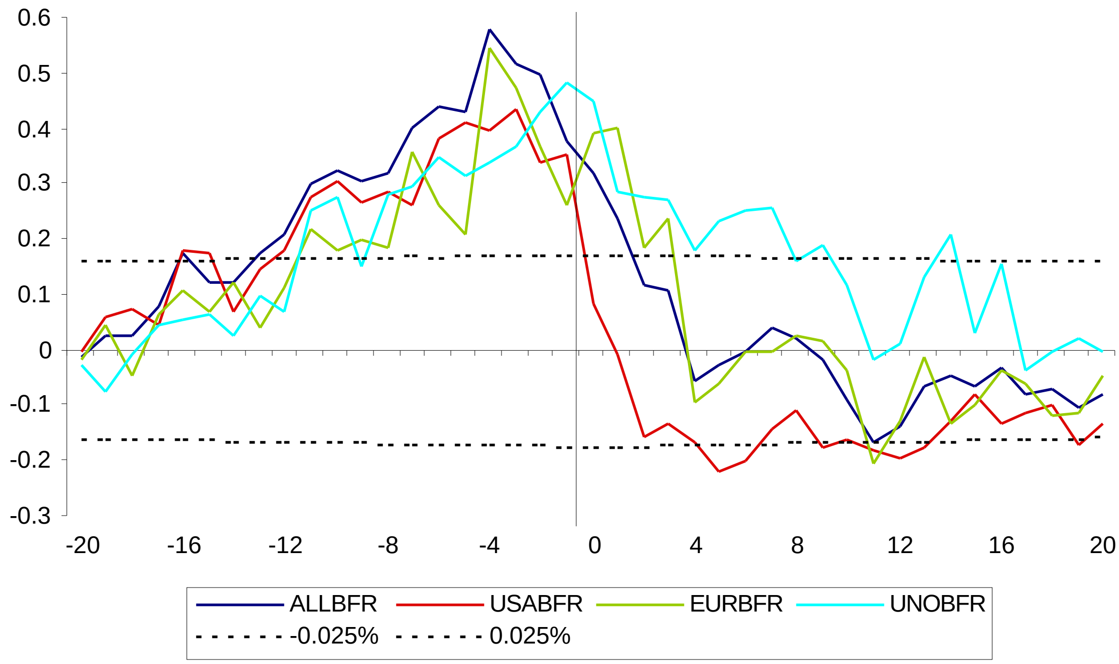
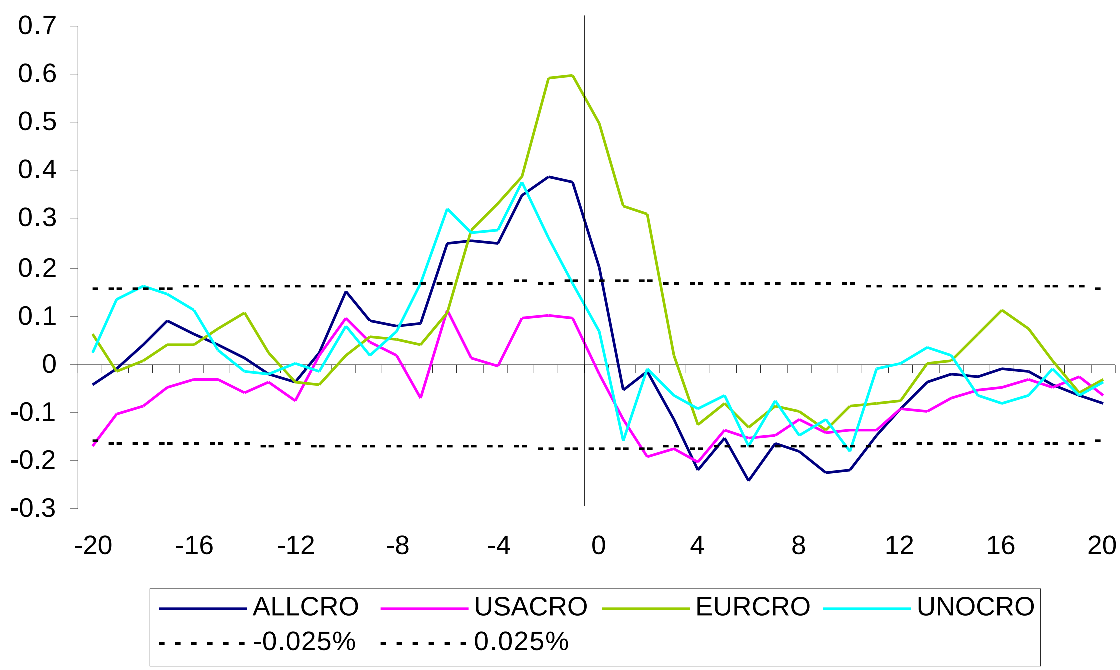


Figure 4.3. Cross-correlations of mediation and Serbia-Croatia conflict



Finally, European mediation efforts have no effect. It shows the usual positive correlation in the lagged period, but most of the correlations are close to zero for periods following the mediation. Unlike the UN efforts, European mediation does no harm, but it does no good either.

The pattern for the Serbia-Croatia pattern—Figure 4.3—is generally similar, but with a couple of differences. First, the period of significant lagged correlations (that is, mediation responding to increased conflict) is about half the length of the comparable period for Bosnia; this may be due in part to the more concentrated character of fighting in the Serbia-Croatia conflict, which generally occurred in a few months in 1991 and 1995. Second, the period of positive correlations following UN mediation is shorter, and unlike the Bosnia case, there are no positive correlations following European mediation. The most significant negative correlations are found with the "ALL" measure of mediation, which could either indicate successful coordination of mediation efforts or simply be an artifact.

4.3. Time Series Analysis of the "Sticks-or-Carrots" Model

The next series of tests will look at the "sticks-or-carrots" issue: is mediation more likely to be effective when it is accompanied by material cooperation or conflict? We will test this using the mediation and event counts of cooperative and conflictual behavior between the mediator and antagonists as the independent variables, and several measures of mediation effectiveness as the dependent variable.

Figure 4.4 summarizes a number of experiments with different formulations of the "conflict variable." It shows the average values across the ten cases of the z-score on the mediation variable for various lags (k = 0 to 10) of the independent variables when the complete sticks-and-carrots model is estimated. The dependent variables tested were

totconf diff:	totconf(t+k) - totconf(t)
totmatcf diff:	totmatcf (t+k) - totmatcf (t)
totconf lags:	totconf(t+k)
totmatcf lags:	totmatcf (t+k)

Effective mediation, in the sense of violence reduction, should result in negative z-scores for all formulations. The pattern we find here is in fact quite consistent. First, the shape of the curve over time is quite similar for the four formulations, with a high positive contemporaneous value (no lag in the independent variables), and then a decline to zero or negative values with the increasing lag, which levels off around k>4. When mediation success is measured by the *change* in conflict levels, the z-score on mediation is consistently stronger than when conflict is measured by the level, and the z-scores are consistently stronger when conflict is measured by the *matcf* event-frequency measure than with the *totconf* Goldstein-scaled measure. Note that these are averages for the ten cases, including the two UN cases that have poor mediation success, so while the average in *best* case—difference in conflict measured by event counts—is barely significant, the z-scores in several of the individual cases are quite significant.

Figure 4.4. Comparision of lagged and differenced measures with conflict dependent variable measured with scaled and frequency totals

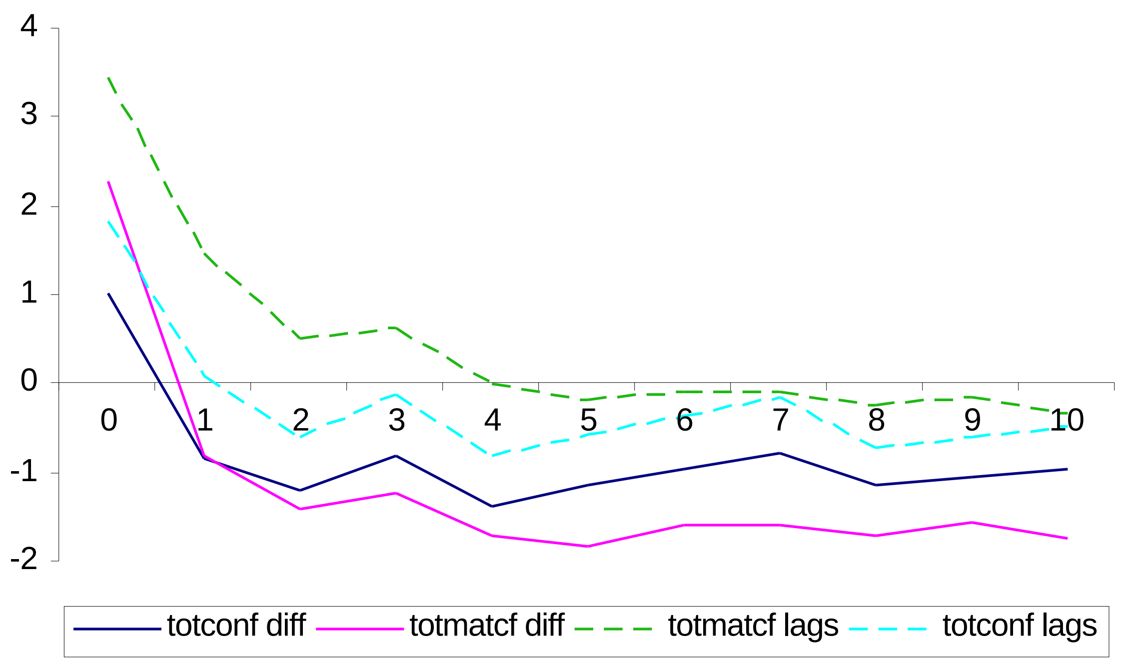


Figure 4.4 is arguably treading a fine line between exploratory analysis and a statistical fishing expedition. We argue that this exploration is necessary, particularly at this early stage of the research, for at least three reasons. First, we know from the cross-correlation analysis that there is a substantial lag between mediation efforts and changes in the level of conflict.

Second, the problem of lags is complicated by the presence of a strong—but theoretically plausible—*positive* contemporaneous correlation between mediation and violence. In the post-WWII period, outbreaks of violence invoke almost immediate attempts at mediation; in Schrodt

(1990)—a sequence-recognition exercise using the BCOW data set (Leng 1987)—mediation was the primary behavior distinguishing pre-WWII and post-WWII crises. Almost all of these variables also have some auto-correlation—for example in the ALLBFR case, both the scaled *totconf* and the frequency *totmatcf* dependent variables have significant (5%-level) auto-correlation to lag 3, and *mediatn* has significant auto-correlation to lag 7.[15] Consequently, sorting out the effects of violence correlating contemporaneously with mediation, but mediation [potentially] correlating at a lag with reduced violence is problematic. Such is the nature of social science statistics when they are applied to real data.

Finally, we are still unclear as to whether it is better to study these behaviors using scaled (*totconf*) or frequency (*totmatcf*) measures of conflict. There is clearly not a whole lot of difference between the two, though in a series of additional experiments we will not report here, we found that the frequency measure almost always produces slightly stronger relationships with the mediation variable, whether measured through the R^2 of the entire equation, or the z-score on the *mediatn* variable. This may be due to the fact that *mediatn* is itself an event-frequency measure, or it may be additional evidence reinforcing our skepticism about the utility of scaled event data. In the end we analyzed both formulations.

Tables 4.1 and 4.2 show the tests of the sticks-or-carrots model for the change in the *totconf* and *totmatcf* variables for differences of 4 and 6 months. The models are of the form

$$\begin{aligned} y(t+k) - y(t) \quad = \quad & a + b_1 \text{ mediatn}(t) + b_2 \text{ m2amatcp}(t) + b_3 \text{ m2amatcf}(t) + \\ & b_4 \text{ m2bmatcp} + b_5 \text{ m2bmatcf}(t) \end{aligned}$$

When OLS regression was used, about half of the cases had significant Durbin-Watson statistics indicating the presence of first-order serially-correlated residuals. Consequently the estimates in these table use the Prais-Winsten (1954) transformed regression estimator (Stata *prais*), though in general the pattern of significant coefficients is the same in the OLS and Prais-Winsten estimates. In order to reduce the size of the tables, only coefficients that were significant a level of $p < 0.10$ are reported; full results are available from the authors.

---

[15] However, *partial* aurto-correlation is significant only at a lag of 1—in other words, the extended auto-correlation is due primarily to strong month-to-month correlation between x(t) and x(t-1).

Table 4.1. Material conflict event frequency, Prais-Winston regression

	Lag 4	Lag 6
ISRPAL -- R^2 (prob)	.077 (.001)	.043 (.063)
mediatn	-.11	-.10
t (prob)	-2.55 (.011)	-2.41 (.017)
m2amatcp		1.43
t (prob)		2.14 (.033)
m2bmatcf	-2.15	
t (prob)	-2.72 (.007)	
ISRLEB -- R^2 (prob)	.125 (.000)	.121 (.000)
mediatn		-.19
t (prob)		-2.85 (.005)
m2amatcf	-3.40	-2.18
t (prob)	-3.90 (.000)	-2.55 (.011)
ALLBFR -- R^2 (prob)	.164 (.001)	.145 (.003)
mediatn	-.07	-.06
t (prob)	-3.73 (.000)	-3.12 (.002)
m2amatcp	.37	
t (prob)	1.80 (.073)	
m2bmatcf		.16
t (prob)		1.66 (.099)
USABFR -- R^2 (prob)	.271 (.000)	.245 (.000)
mediatn	-.14	-.09
t (prob)	-4.67 (.000)	-2.71 (.008)
m2amatcp	2.06	
t (prob)	3.18 (.002)	
EURBFR -- R^2 (prob)	.107 (.025)	.144 (.004)
m2amatcp	-.96	
t (prob)	-1.97 (.051)	
m2amatcf	-.47	
t (prob)	-1.82 (.071)	
m2bmatcp		-.64
t (prob)		-2.70 (.008)
m2bmatcf		.46
t (prob)		2.66 (.009)
UNOBFR -- R^2 (prob)	.032 (.599)	.080 (.097)
mediatn	-.10	-.16
t (prob)	-1.77 (.079)	-2.58 (.011)
m2bmatcp		.30
t (prob)		2.29 (.024)
ALLCRO -- R^2 (prob)	.135 (.005)	.089 (.065)
m2bmatcp		-1.57
t (prob)		-2.27 (.025)
USACRO -- R^2 (prob)	.063 (.194)	.037 (.521)
m2bmatcp	-2.59	-2.52
t (prob)	-1.74 (.085)	-1.68 (.095)

Table 4.1. Material conflict event frequency, Prais-Winston regression, continued

EURCRO -- R^2 (prob)	.164 (.001)	.056 (.266)
mediatn	-.53	
t (prob)	-2.88 (.005)	
m2bmatcf	-1.31	
t (prob)	-1.97 (.050)	
UNOCRO -- R^2 (prob)	.077 (.105)	.065 (.184)
m2bmatcp		-2.20
t (prob)		-1.89 (.061)

Table 4.2. Goldstein-scaled conflict, Prais-Winston regression

	Lag 4	**Lag 6**
ISRPAL -- R^2 (prob)	.042 (.068)	.040 (.085)
m2bmatcp	15.73	24.51
t (prob)	1.63 (.104)	2.40 (.017)
m2bmatcf	-23.67	
t (prob)	-2.77 (.006)	
ISRLEB -- R^2 (prob)	.104 (.000)	.110 (.000)
mediatn		-1.69
t (prob)		-2.75 (.006)
m2amatcf	-27.44	-19.18
t (prob)	-3.45 (.001)	-2.48 (.014)
ALLBFR -- R^2 (prob)	.098 (.040)	.164 (.001)
mediatn	-.46	
t (prob)	-2.27 (.025)	
m2amatcf		-5.56
t (prob)		-3.90 (.000)
m2bmatcf		1.71
t (prob)		1.91 (.058)
USABFR -- R^2 (prob)	.029 (.650)	.099 (.040)
m2amatcf		-8.60
t (prob)		-2.37 (.019)
EURBFR -- R^2 (prob)	.172 (.000)	.157 (.002)
mediatn	-1.25	
t (prob)	-2.99 (.003)	
m2amatcf	-4.64	-5.33
t (prob)	-1.82 (.071)	-1.90 (.060)
m2bmatcp		-4.95
t (prob)		-2.17 (.032)
m2bmatcf		3.87
t (prob)		2.32 (.022)
UNOBFR -- R^2 (prob)	.055 (.270)	.112 (.022)
m2amatcf		-3.74
t (prob)		-1.73 (.086)
m2bmatcp		2.89
t (prob)		2.34 (.021)

Table 4.2. Goldstein-scaled conflict, Prais-Winston regression, continued

ALLCRO -- R^2 (prob)	.112 (.019)	.141 (.004)
m2bmatcp	-12.05	-21.68
t (prob)	-1.77 (.078)	-2.99 (.003)
USACRO -- R^2 (prob)	.074 (.122)	.053 (.297)
m2bmatcp	-30.15	-36.12
t (prob)	-2.08 (.040)	-2.25 (.026)
EURCRO -- R^2 (prob)	.164 (.001)	.111 (.022)
mediatn	-5.34	-4.02
t (prob)	-3.03 (.003)	-1.99 (.049)
m2amatcf	-10.89	
t (prob)	-1.95 (.053)	
UNOCRO -- R^2 (prob)	.085 (.074)	.117 (.017)
m2bmatcp	-19.43	-31.68
t (prob)	-1.71 (.090)	-2.62 (.010)

Three general patterns are evident from these results. First, generally the results at the lag of 4 and the lag of 6 are similar, though there are several exceptions to this. However, these differences involve only the presence of a significant coefficient; in no cases do we observe the sign of a coefficient changing with the change in the lag time. This is consistent with Figure 4,1, which suggests that mediation takes hold after a lag time of about four months and then has a generally consistent effect.

Second, the correlations are significant on almost all of the regressions. Most of the exceptions involve either the UN—as expected—and the USA mediation on Croatia. In contrast to the cross-correlation analysis, there are significant correlations in all of the Levant cases.

Finally, the mediation variable—when significant—is *always* negative: there are no exceptions to this pattern. In the tables as a whole, about two-thirds of the significant coefficients on the behavior variables are negative (21 out of 30); the exception is the conflict frequency variable at lag 4 where there are equal numbers of positive and negative coefficients.

The analysis,however, is less clear on the “sticks-or-carrots” question. The scaled measure provides a relatively clear pattern, with *m2amatcf* always negative (conflict with the stronger antagonist reduces conflict) and usually positive coefficients on the *m2bmatcp* variable (rewards to the weaker antagonist), though this variable also has several negative coefficients. *m2bmatcf* is significant in only three cases, but the coefficients are inconsistent; and *m2amatcp*—material aid to the stronger antagonist—is never significant. Consequently the message from this analysis is that mediation is most likely to reduce violence when it is combined with conflict towards the stronger antagonist and rewards to the weaker. However, there is a lot of variability among the cases on this.

These results do not hold up when the frequency measure is used. The consistent finding of negative coefficients on *m2amatcf* remains, but it is only significant in three cases. The remaining variables are all found with both positive and negative signs depending on the case, although consistent with the scaled results, *m2bmatcp* is negative in 5 out of 6 of the cases where it is significant.

Table 4.3 shows the analysis for all of the cases combined. This data set was created by concatenating all of the data files, manually creating the differenced variables in Excel, then eliminating the final cases in each set because these are actually creating a "difference" using the next data series; the total sample size is 1,400. Because this is a pooled time-series rather than a single series, Prais-Winsten could not be used, so the estimation method is OLS.

Table 4.3. All cases combined

	Lag 4	**Lag 6**
matcf- conflict frequency		
R^2 (prob)	.0.04 (<.001)	.0.04 (<.001)
mediatn	-0.95	-0.12
t (prob)	-4.54 (<.001)	-4.87 (<.001)
m2amatcf	-0.489	-0.39
t (prob)	-2.46 (.014)	-1.85 (.064)
totconf - scaled conflict		
R^2 (prob)	.022 (<.001)	.023 (<.001)
mediatn	-0.43	-0.52
t (prob)	-2.16 (.031)	-2.42 (.016)
m2amatcf	-5.30	-6.51
t (prob)	-2.79 (.005)	-3.22 (.001)
m2bmatcf	-2.51	
t (prob)	-1.87 (.061)	

The combined case analysis is generally consistent with the strongest results in the individual cases. The coefficients for mediation and *m2amatcf* are consistently negative in sign and significantly different from zero. With one exception, the results are consistent across the lag 4 and lag 6 cases, and the fit of the overall model is highly significant, though the R^2 is quite small (although this is not uncommon with such a large sample size.) The frequency measure generally produces stronger results than the scaled measure, but the differences are not dramatic. The "carrot" effects of the *m2bmatcp* variable do not show up anywhere in this analysis, and in fact the estimated coefficients (significance levels around 0.3) are positive.

4.4. Proportional Hazards Models

Our final analysis uses duration models[16], specifically the Cox proportional hazard model. In this approach, the variable of interest is the expected amount of time required for an event to occur, but this is modeled explicitly as a stochastic process rather than as a deterministic process. In other words, the independent variables increase or decrease the probability of an event occurring, but the model does not attempt to predict exactly when the event will occur. This approach is consistent with the theoretical expectations of the mediation literature, which suggests that there is a large random component to the timing of negotiation phases. It also has the distinct advantage of not requiring arbitrary aggregation of the behavior into a period such as a month, which Thomas (1999b) has shown to be potentially problematic in event data.

As an exploratory effort, we will test this using the simple Cox proportion hazard model. For each of the dependent variables, we look at the length of time between the beginning of the "at risk" period and the time of "failure."[17] There are multiple instances of these periods—in some cases hundreds of instances—in each of our cases, and these multiple instances become of the observations of our analysis. The "treatments" in each case will be the average daily frequency of each of the aggregated interactions *vercp, matcp, vercf* and *matcf* within the "at risk" period.[18] These are tabulated between the antagonists (a separate set of variables for each directed dyad) and between the mediator(s) and the both of the antagonists combined (note that this last failure

[16] see Allison 1984; Blossfeld, Hamerle, & Mayer 1989; Blossfield & Rohwer 1995; Maller & Zhou 1996; Box-Steffensmeier & Jones 1997; Bennett 1999.

[17] We are using this term in the technical sense employed in the survival time literature: it is the time that the event defining the end of the activity being studied occurs. For two of the three indicators, the "failure" is in fact a success in terms of mediation.

[18] We also tried estimating the model using the total number of event rather than average daily event frequency, but because these event counts are generally proportional to the length of the survival time, all of the coefficients are negative and there is no coherent pattern to the set of coefficients.

is different than the "sticks and carrots" model, where m2a and m2b behaviors were measured separately). We will analyze the cases both separately and collectively.

We operationalize the core hypotheses of our project using the following patterns:

- ❖ Do the disputants openly agree to mediation?
  - At risk pattern: WEIS 22 event between antagonists
  - Failure pattern: Mediation event (defined in section 4.2)
- ❖ Do the parties formally reach an agreement?
  - At risk pattern: Mediation event
  - Failure pattern: Agreement events (WEIS 05 or WEIS 08) in both directions in the dyad within a period of 7 days
- ❖ Does the agreement reduce violence?
  - At risk pattern: Agreement as defined above
  - Failure pattern: Eight WEIS 22 events between antagonists

In the cases where multiple events are required to match the pattern, the failure date is the day of the event that completes the pattern. We have specified that at least one day must occur between the beginning of the risk period and the failure, and new risk periods begin at least one day after the previous failure. While these behaviors could form the cyclical pattern

violence → mediation → agreement → violence

using the definitions, it is possible to get two consecutive periods of mediation and agreement without having violence following the agreement (this in fact occurs frequently in the data).

The Cox proportional hazard model was estimated using the *stcox* routine in Stata 6.0; default options were used.[19] These results should be considered tentative: we have only spot-checked for the effects of collinearity causing a high correlation between the coefficient estimates in those places where we found anomalous coefficients; we have not checked the extent to which our data are consistent with the assumptions of the Cox model, nor have we looked at alternative model specifications such as using the Weibull or Gompertz distributions. Hazard rates are reported along with z-scores and significance level; when the hazard rate is greater than 1.0 ($z > 0$), higher values of the event type is associated with a shorter survival time; a hazard rate less than 1.0 ($z < 0$) means that the activity is associated with a longer survival time. In mediation and agreement tests, short survival times indicate successful third-party mediation; in the violence test short survival time indicates unsuccessful mediation.

The results of this analysis are reported in Tables 4.4 to 4.13. In order to reduce the size of the tables, only coefficients that were significant a level of $p < 0.10$ are reported; full results are available from the authors. In several of the cases where the sample case is small, extremely large ($HR > 10^6$) or extremely small ($HR < 10^{-6}$) coefficients were estimated; these are reported as "+++" and "---" respectively. The model for "violence" excludes the *a2bmatcf* and *b2amatcf* measures, since these events are used to define the failure point.

The two Levant cases are reported in Tables 4.4 to 4.6. Five of the six models have significant fit—the exception is the Israel-Palestinian mediation model—and generally the coefficients are plausible. The time between violence and mediation in the Israel-Lebanon case is shortened by material and verbal cooperation from the mediator, and by verbal cooperation from Israel; none of the measured indicators have a significant effect on lengthening the period between violence and mediation.

The model for agreement in the Israel-Lebanon case has the highest number of significant coefficients, but also several problematic estimates that are very high, very low, or have signs that are inconsistent with expectations (e.g. Israel material conflict, unless this is actually forcing the Lebanese side to an agreement); the small sample size may be problematic here. The Israel-Palestinian case has only a single significant coefficient—Israel's verbal cooperation, which shortens the period—but since the significance of the entire model is substantially greater than that of the coefficient, the effects of other factors may be being masked by collinearity.

---

[19] the Breslow method for dealing with ties was used, and the standard rather than the Stata "robust" method was used to calculate the variance-covariance matrix

Table 4.4. Proportional Hazards Estimates for Levant Mediation

	Israel-Lebanon	Israel-Palestinian
MED verbal coop	.551	
z	-2.89	
prob	.005	
MED material coop	2.034	
z	2.80	
prob	.005	
ISR verbal coop	1.80	
z	1.80	
prob	..071	
OPP verbal coop		1.31
z		1.95
prob		.051
N	379	621
LR chi2 (prob)	30.97 (.002)	8.64 (.733)

Table 4.5. Proportional Hazards Estimates for Levant Violence

	Israel-Lebanon	Israel-Palestinian
MED material coop	+++	
z	2.98	
prob	.003	
MED verbal conflict		17.46
z		2.35
prob		.018
MED material conflict	---	
z	-2.80	
prob	.005	
ISR verbal conflict		9.09
z		2.33
prob		.020
OPP verbal coop	.0002	
z	-2.11	
prob	0.035	
N	24	98
LR chi2 (prob)	26.44 (.003)	20.29 (.026)

Table 4.6. Proportional Hazards Estimates for Levant Agreement

	Israel-Lebanon	Israel-Palestinian
MED verbal coop	5.91	
z	2.28	
prob	.023	
MED material coop	---	
z	-1.98	
prob	0.048	
MED verbal conflict	---	
z	-2.53	
prob	.011	
ISR verbal coop		2.18
z		2.29
prob		.022
ISR material conflict	81.81	
z	2.14	
prob	0.032	
OPP material coop	+++	
z	2.34	
prob	0.019	
OPP verbal conflict	---	
z	-1.72	
prob	.085	
N	29	157
LR chi2	28.04	32.13
Prob	.005	.001

Finally, the coefficients for the violence measure in Israel-Lebanon are again problematic—they have extreme magnitudes and implausible signs, this is probably a consequence of the small sample size. The Israel-Palestinian case, in contrast, presents a very consistent story: when the mediator or Israel engage in verbal conflict, the agreement is about to break down.

Tables 4.7 to 4.12 show the analysis of the Balkans, first with the cases aggregated, and then with Bosnia and Croatia treated separately. The aggregated case ("Balkans") includes cases involving Kosovo; these were initially going to be included in the analysis but were found to have too few cases to analyze separately. Similarly, there were too few cases of periods between agreement and violence to analyze separately for Bosnia and Croatia, and too few agreements to analyze separately for Bosnia.

The patterns of coefficient estimates are less consistent than those in the “sticks-or-carrots” model, but we would note the following general patterns.

1. Except for the low-sample cases, most of the significant coefficients are positive—that is, they indicate behaviors that reduce the amount of time before mediation or agreement. In some of the cases where there are negative estimates that are inconsistent with theoretical expectations, we have found collinearity (i.e. relatively high correlations between the coefficient estimates as reported by the Stata *vce, corr* command) to be an issue. For example, the positive coefficient on opposition verbal conflict in the Balkans Agreement/Europe case has a negative correlation of –0.68 with the coefficient estimate of opposition material cooperation, and may be masking the effect of that variable.
2. The “Balkans violence” results make no sense whatsoever. Verbal cooperation appears to be the most important variable, but it is consistently of the wrong sign. Collinearity does not appear to explain this; the small sample size might.
3. Consistent with the cross-correlation analysis, the fit of the UN cases are generally weaker than those of the USA and European cases. As expected, the coefficients for the individual cases are frequently quite different than those of the collective (“ALL”) case, though some of this may be due to quirks in the pattern-recognition.

Table 4.7. Proportional Hazards Estimates for Balkans Mediation

	Mediators			
	All	**USA**	**Europe**	**UN**
MED verbal coop	1.210		1.903	6.42
z	1.70		3.33	4.71
prob	0.088		0.001	<0.001
MED verbal conf		7.68		
z		2.34		
prob		0.019		
MED material conflict			3.28	
z			3.28	
prob			0.001	
SER material coop		3.13		
z		2.46		
prob		0.014		
SER verbal conflict		0.0176		
z		-1.74		
prob		0.082		
OPP material coop		8.29		
z		1.92		
prob		0.055		
OPP verbal conflict			4.51	
z			2.39	
prob			0.018	
OPP material conflict		6.18		2.90
z		2.64		1.69
prob		0.008		<0.000
N	189	104	131	108
LR chi2	15.42	31.63	32.28	44.90
Prob	0.21	0.002	0.001	<0.000

Table 4.8. Proportional Hazards Estimates for Balkans Agreements

	Mediators			
	All	**USA**	**Europe**	**UN**
MED verbal coop				176.84
z				4.83
prob				<0.000
MED material coop	6.78		18.00	
z	2.02		2.02	
prob	0.044		0.043	
MED verbal conf		.0003		0.0019
z		-1.75		-2.12
prob		0.081		0.003
MED material conf				.00002
z				-2.92
prob				0.003
SER verbal coop	41.41			
z	2.42			
prob	0.015			
SER material coop		579.15		
z		2.06		
prob		0.039		
SER verbal conflict			.019	
z			-1.71	
prob			0.088	
OPP verbal coop	19.20	265.86		163.2
z	1.89	3.80		2.02
prob	0.059	<0.001		0.044
OPP verbal conflict			45.37	153.57
z			2.46	3.30
prob			0.014	0.001
N	68	41	51	44
LR chi2	25.32	33.62	25.18	48.20
Prob	0.013	0.001	0.014	<0.000

Table 4.9. Proportional Hazards Estimates for Balkans Violence

	All	Mediators USA	Europe	UN
MED material coop	57.18			
z	1.94			
prob	0.052			
SER verbal coop	3.47	3.04	2.80	3.62
z	2.79	2.34	2.18	2.67
prob	.005	.019	.029	.008
OPP verbal coop	.28	5.70	4.87	5.09
z	1.90	2.08	1.97	1.67
prob	.057	.038	.049	.094
N	44	44	44	44
LR chi2	23.35	22.04	22.82	23.63
Prob	0.009	0.015	0.011	0.008

Table 4.11. Proportional Hazards Estimates for Croatia Mediation

	All	Mediators USA	Europe	UN
MED verbal coop				70.21
z (prob)				3.684 (<0.001)
MED material coop			---	
z (prob)			-1.763 (0.078)	
SER material coop				6.62
z (prob)				2.189 (0.029)
SER material conflict			2.0455	
z (prob)			1.684 (0.092)	
OPP material coop				---
z (prob)				-1.91(0.056)
OPP verbal conflict	2.31	37.85	3.74	
z (prob)	2.809 (0.005)	1.907 (0.057)	1.689 (0.091)	
N	85	31	54	44
LR chi2	16.41	14.54	20.39	36.26
Prob	.1732	.267	.060	<.001

Table 4.10. Proportional Hazards Estimates for Bosnia Mediation

		Mediators		
	All	**USA**	**Europe**	**UN**
MED verbal coop			1.86	4.04
z			1.85	2.71
prob			.064	.007
MED material conf		12.63	8.16	
z		2.05	2.62	
prob		.041	.009	
SER verbal coop			19.00	
z			2.82	
prob			.023	
SER verbal conflict			37.36	
z			3.09	
prob			.002	
SER material confl	5.69		37.36	12.75
z	2.02		3.09	2.34
prob	.043		.002	.007
OPP verbal coop		7.88	.064	
z		1.83	-3.03	
prob		.066	.002	
OPP verbal conflict		4.30	25.20	
z		1.85	3.47	
prob		.065	.001	
OPP material conflict	9.05			
z		2.07		
prob		.038		
N	66	45	49	52
LR chi2	12.79	27.96	28.62	17.46
Prob	.384	.005	.004	.133

Table 4.12. Proportional Hazards Estimates for Croatia Agreement

	Mediators			
	All	**USA**	**Europe**	**UN**
MED material coop			51.75	
z			2.19	
prob			0.029	
MED verbal conflict	82.58			
z	2.81			
prob	.066			
MED material conf	0.00493			
z	-1.84			
prob	0.066			
SER verbal coop				325.42
z				2.802
prob				0.005
SER verbal conflict				+++
z				1.81
prob				0.07
SER material conflict	69.72			
z	3.068			
prob	0.002			
OPP verbal coop	1150.0	114.05		+++
z	3.19	2.47		2.41
prob	0.001	0.013		0.016
OPP verbal conflict	2848.3	10404	555.89	----
z	3.153	1.665	1.802	-1.9
prob	0.002	0.096	0.072	0.057
OPP material conflict	0.00039			
z	-2.293			
prob	0.022			
N	41	21	28	22
LR chi2	32.39	20.58	19.09	24.07
Prob	.001	.057	.086	.020

Finally, Table 4.13 shows the results when all of the cases—the two Levant cases and the eight Balkans cases—combined in a single analysis. The combined results have the advantage of a large sample size and, in fact, make more sense than several of the individual cases. The Mediation and Agreement cases are straightforward: The time between violence and mediation is shortened by material conflict involving the mediator and the smaller actor; the time from mediation to agreement is lengthened by verbal conflict by the mediator (yelling apparently doesn't work here) and material conflict by the smaller antagonist. The standard Violence model has some coefficients that are the opposite of theoretical expectations, but these could be affected by a strong negative correlation (-0.91) between the coefficients between the antagonists verbal cooperation. If the *a2bvercp* variable is eliminated ("Violence2"), the resulting model shows positive coefficients (that is, shorter duration times between agreement and subsequent violence) result from verbal conflict by the antagonists.

Table 4.13. Proportional Hazards Estimates for All Cases Combined

	Behavior			
	Mediation	**Agreement**	**Violence**	**Violence2**
MED material coop			9.77	
z (prob)			1.96 (.050)	
MED verbal conflict		.282		
z (prob)		-1.86 (.063)		
MED material conflict	1.51			
z (prob)	2.73			
Actor A verbal coop			2.79	
z (prob)			2.32 (.021)	
Actor A verbal confict			2.79	5.46
z (prob)			2.32 (.021)	1.22 (.034)
Actor B verbal conflict			8.00	8.19
z (prob)			2.47 (.013)	2.57 (.010)
Actor B material confl	1.33	1.42		
z (prob)	2.76 (.006)	-1.66 (.096)		
N	1189	254	166	166
LR chi2	27.65	57.75	46.78	42.35
Prob	<.006	.001	<.001	<.001

5. Conclusions

This analysis was intended to illustrate three main points. First, it is possible to formulate meaningful hypotheses about the dynamics of mediation processes —as distinct from their structural characteristics—and test these hypotheses using conventional statistical methods. Second, by using a combination of machine-coded event data and relatively simple definitions of event sequences, it is possible to derive plausible measures of behavior relevant to this study in a completely transparent and reproducible manner that does not “judgement calls” by human coders. Finally, we have shown that the results found from analyzing individual cases may be quite different from those of aggregated samples. While this last point presumably is not surprising, it does run against the grain of much of the statistical research in international politics.

This is the first major analytical work from our project, and we regard these results as illustrative rather than conclusive. The hypotheses that we have studied here do not capture many of the nuances (or inconsistencies) in the existing theoretical literature and we have done only a few of the necessary diagnostic tests on the statistical results. As discussed in more detail below, we are still using data coded in the WEIS framework, which we do not think is ideal for the studying mediation, or post-Cold War political behavior more generally. In this concluding section, we will discuss briefly where the project is going from here.

5.1. MEDB—Yet Another Event Coding Scheme

Machine coding allows researchers to experiment with alternative coding rules that reflect a particular theoretical perspective or interest in a specific set of issues. Both COPDAB (Azar 1982) and WEIS were both developed during the Cold War and assume a "Westphalian-Clausewitzian" political world view of sovereign states reacting to each other through diplomacy and military threats. Consequently these coding systems are ill-suited to dealing with contemporary issues such as ethnic conflict, low-intensity conflict, organized criminal activity, or multilateral intervention.[20] These systems have other problems as well: for example WEIS has only a single category of “military engagement” that must encompass everything from a shot fired at a border patrol to the strategic bombing of cities. COPDAB contains only 16 event categories, and these are intended to span a single conflict-cooperation continuum that many researchers consider inappropriate. WEIS was considered only a "first draft" by Charles McClelland, its creator, and he certainly did not anticipate that WEIS would continue to be used, with only minor modifications, for four decades (McClelland 1983).

The "lock-in" of these early coding systems is readily explained by the time-consuming nature of human event coding from paper and microfilm sources. Because human coders typically produce between five and ten events per hour, and a large data set contains tens of thousands of events, experimental re-coding was not possible. Established protocols for training and maintaining consistency among coders presumably further constrained efforts to modify WEIS and COPDAB once these were institutionalized. As a consequence, only marginal changes were made in these schemes such as Tomlinson's (1993) incremental extensions of WEIS or the GEDS project extensions of COPDAB (Davies and McDaniel 1993). Automating coding, in contrast, allows even a long series of texts spanning multiple decades to be recoded in a few minutes and allows a researcher to focus his or her efforts on maximizing the *validity* of a coding scheme for a particular problem, since the automated coding process itself guarantees the reliability of the system.

Despite the obvious drawbacks of WEIS, we have used that coding system for all of our earlier work with KEDS. WEIS was good enough, and in the early stages of our development of automated coding, it was important for us to be implement an existing system so that we could directly compare human-coded and machine-coded data (Schrodt & Gerner 1994).

However, we have recently decided to abandon WEIS. Three considerations motivated this decision. First and foremost were long-standing criticisms of the WEIS cue categories. Most conspicuously, the "Warn" (16) category overlaps almost completely with either the "Threat" (17) category or the "Demonstration—Armed force display" (182) category, and a "Promise" (05) is almost impossible to distinguish from an "Agree" (08) except for the idiosyncratic use of the English word "promise." In addition, the distribution of events in WEIS is quite irregular, and several of the cue categories generate almost no events.

---

[20] There have been some efforts to extend the WEIS and COPDAB—most notably Leng's (1987) Behavioral Correlates of War (BCOW) and the Bond et al (1997) Protocol for the Analysis of Nonviolent Direct Action

The result is the coding scheme we are tentatively calling the Mediation Event Data Base, or MEDB.[21] MEDB is specifically designed to code events relevant to the mediation of violent conflict—its tertiary categories involve objects such as cease-fires and peacekeeping—and many of the categories would not work for a trade negotiation or labor dispute. The current (23 August 2001) draft of this is found in Appendix III. Please note that this will definitely be changed over the next couple of months—for example, we do not intend to retain the empty categories 09 and 15 and unless we fill these in with new categories, this will result in a major renumbering of the categories.

We are the first to acknowledge that development of a new coding framework may or may not be a good thing. The clear disadvantage is that MEDB introduces yet another event coding scheme into the discipline. But given the known ambiguities in WEIS, and a large number of behaviors that WEIS does not differentiate, perhaps it is appropriate to experiment with a variety of new schemes in order to determine what types of categories can most effectively be used in event data analysis.

More generally, we contend that the patterns of mediation behaviors (or any other political behavior) have a significant empirical component that is distinct from the theoretical considerations of the academic literature on the subject, and therefore it is important to experiment with coding systems rather than trying to establish these *a priori*. Due to the strong selectivity of news reports, the fact that a behavior may be important in a case study (the analytical approach that still informs most of the mediation literature) does not mean this behavior will necessarily show up as a useful *statistical indicator*. Similarly, good exploratory analysis of the event data may reveal indicators not found in the theoretical literature, often because these are surrogates for other variables. This is not be say that statistical studies should be atheoretical, but the development of useful statistical models will, in part, be an empirical exercise of matching methods to data.

Finally, the effort involved in implementing a new coding system—even one that involves a radical rearrangement of several of the WEIS categories—is relatively small because most of this

(PANDA)—but WEIS and COPDAB still dominate the published literature.

[21] This apparently innocent acronym is, in fact, the name of the Celtic goddess of war, also known as Maeve and Medhbh. See http://www.geocities.com/cas111jd/celts_table/majordeities/celts_medb.htm for additional information. Third-party conflict mediation does not figure prominently in the mythology surrounding Medb—her interests ran more towards sex, violence, and excessive consumption of alcoholic beverages—and consequently

can be done within the dictionary of verb phrases. In most cases the verb phrase can be unambiguously assigned to the appropriate new category. If a phrase *cannot* be unambiguously assigned to a code, it will be eliminated or modified, and this itself is an improvement in the coding system. We anticipate that implementing the new system in our existing dictionaries (which use WEIS codes) will require only a person-month or so of work in the dictionaries, followed by a longer period of further dictionary development that evaluates the system on sentences from news reports and adds new phrases as necessary. As long as dictionaries are preserved along with the data, future researchers can determine precisely the verb phrases that were used in each coding category.

We are currently in the process of developing a formal codebook for this system. Following the lead of the IDEA codebook from the PANDA project (see http://vranet.com/idea/), this will exist in both a printed and a web-based format, and the description of each coding category contain several examples of sentences illustrating the use of the category. We have also followed the lead of IDEA in introducing 4-digit tertiary coding categories that focus on very specific types of behavior, for example differentiating agreement to, or rejection of, ceasefires, peacekeeping, and conflict settlement. We anticipate that in most of our analysis the tertiary categories will not be used—we will instead aggregate the data to the secondary or primary level—but this framework retains distinct code for very specific behaviors that might be useful in defining patterns.

In contrast to IDEA, we have not retained strict backwards-compatibility with WEIS. Instead, we have combined some of the WEIS categories, and provided far more detail on others (notably our expansion of the WEIS 23 “force” category into three distinct primary categories. Nonetheless, it should be possible to map most of the MEDB secondary categories into a clear WEIS secondary category, and in fact over half of our primary categories are identical to those in WEIS.

In the long run, we anticipate that event data coding schemes could evolve using a “mix-and-match” framework whereby a researcher could adopt most of his or her coding categories from a standard set, and then elaborate on a smaller number of new categories. For example, a data set

---

this name should be considered merely a working title. Our project should also not be confused with that of the dealing with trade negotiation would not require any of the detail MEDB has on ceasefires and peacekeeping, and it would require substantially more detail on imposition of tariffs, non-tariff barriers, and appeals to the World Trade Organization. However, primary categories such as Consult, Agree, and Reject would be the same in both systems, and many of the secondary categories that deal with behaviors not specific to mediation or trade would also be the same. Common vocabulary of dictionaries could also be shared, and the focus of the new dictionary development could be on the behaviors specific to a particular theoretical issue.

5.2. Additional Conflicts

As the analysis in this paper has shown, the effects of mediation can vary across conflicts. In addition to continuing our analysis on the Levant and Balkans conflicts (and probably adding the Serbia-Kosovo and Macedonia-Kosovo conflicts to the Balkans case), we expect to expand our analysis using KEDS data to additional cases in the Middle East, a series of cases in West Africa, and possibly some additional cases based on data sets that have been coded by other projects.

The Middle East is our longest time series (beginning 15 April 1979, a few weeks after the start of the Iran-Iraq War) and it is the region where we have invested the greatest amount of effort in refining our coding dictionaries, often with coders who have had field experience in the Levant. Conveniently for us—if rather inconveniently for the local populations—this area has experienced a number of conflicts that have been subject to a variety of different mediation efforts and degrees of success. In addition to the Israel-Lebanon and Israel-Palestinian dyads we have examined in this paper, mediated dyads include:

- Israel and Jordan
- Israel and Syria
- Syria and Lebanon
- Iraq and various international organizations
- various parties in the Lebanese civil war
- Iran and Iraq
- Iran and the United States

This region has been intensely covered by the international news media and a detailed record of political activity is available. It has also been the subject of numerous case studies of international mediation: An informal survey of the books at the University of Kansas library

Maui Economic Development Board, another entity with whom we suspect we have strong disagreements.

listed under the subject heading “Mediation, international” found that about a third of the case studies dealt with the Middle East.

West Africa is another region that has been subject to extensive conflict and mediation efforts. We will focus on the period 1990 to the present and we already have regionally-specific coding dictionaries available for these areas. The civil conflicts in West Africa, in contrast to those of the Balkans, have been dealt with primarily through regional intervention by ECOWAS, although more recently there has been some United Nations involvement. We will focus primarily on the civil wars in Liberia and Sierra Leone, although if sufficient data are available, we will also try to look at Senegal-Mauritania, Nigeria-Cameroon, and possibly international efforts to mediate ethnic conflicts within Nigeria. Unlike the Middle East and Balkans cases, West Africa is only sporadically covered by the international media (Huxtable & Pevehouse 1996), and case studies of mediation in this region are rare.

If we find some hypotheses that are strongly supported (or produce contradictory results) in these three areas, data are available on other regions. These additional tests will not be identical to our core tests because of differences in coding systems and the operationalization of some variables, but they will expand the temporal and geographical scope of our analysis.

At the present time, KEDS-coded data sets are available on the conflicts between North and South Korea, China-Taiwan, and the civil conflict in Northern Ireland. All of these disputes have involved extensive international mediation. The Behavioral Correlates of War data set (BCOW; Leng 1987) provides another a dense, high-quality event data set that focuses on about forty crises over the past two centuries, and employs an extensive set of codes involving mediation activities. While some of the crises coded in BCOW involve very little third-party mediation, quite a few were mediated—successfully and unsuccessfully—and could be analyzed. BCOW would considerably extend the temporal range of our analysis.

5.3. Modeling Lagged Responses

This analysis has made us acutely aware of the problem of analyzing processes that are known to have a time lag between the action and the effect, but where the length of that time lag is stochastic. This has emerged as a major complication when we are using conventional time series methods involving the correlation of variables at fixed lag (t-k). Because we do not know of any “natural” time lag or set of lags to apply in these models, we are instead are left with at least three options, none of which we find wholly palatable.

The first method, which we employed in the time-series analysis here, is to use exploratory methods such as cross-correlation and experimentation with alternative lag structures (within some plausible range) to get a general idea of the lag where the effects of the relationship seem to be strongest, and do the analysis with a small number of lags. This runs the risk, however, of over-fitting the data, and the choice of the time lag is somewhat arbitrary, particularly (as we have seen in this case) where the effects are spread across a number of months. There is also no guarantee that results will be stable across multiple lags.

The second method would be to use a range of credible lags for all of the independent variables. This is the approach used in VAR, which has been employed in a number of studies that use event data (e.g. Goldstein & Freeman 1990; Goldstein and Pevehouse 1997). The disadvantage of VAR is that it puts one back into the realm of diffuse parameter structures with indeterminant values: When the independent variables are auto-correlated (as the sequences studied here generally are), collinearity expands the standard errors of the VAR coefficients to the point where they cannot be interpreted substantively, which is the same situation one finds with computational methods such as neural networks and hidden Markov models. On the positive side—at least for the analyst—the decades-long crises that we are examining provide sufficient degrees of freedom, even at monthly levels of aggregation, that fairly elaborate VAR models can be estimated.

The final method would be to use a method such as duration models or Poisson regression where the stochastic delay between the “treatment” and “response” is explicitly part of the model. Despite the rather mixed experience that we have had in this exploratory analysis using the Cox proportional hazards model, this is probably the most appropriate method, even though it takes one into territory that is not wholly familiar to most political methodologists.

We suspect that the issue of specifying lagged responses has not received a great deal of attention in the existing quantitative literature in international politics because most of these studies have used data aggregated by year. That time period is sufficiently long that most responses will appear to occur either contemporaneously or, at most, with a lag of one period. Event data, in contrast, can effectively be aggregated to a month or even a week, and at this level of detail, there is usually a substantial difference between the time that a change in behavior occurs and the time its effects are observed.

5.4. Formal Specification of Patterns

It would be nice to come up with a means of formally and consistently specifying date-ordered sequences that is comparable to that of a perl regular expression. Natural language is not a particularly good way of describing patterns, and while procedural programming languages such as C can be used to unambiguously define (and implement) a pattern, computer languages are probably little better than natural language in expressing patterns in a form that can be easily understood, manipulated and compared. Patterns expressed as regular expressions are often sufficiently simple that they fit into the "5 ± 2" limit of human working memory; patterns expressed in C do not.

The existing regular expression notation of perl goes part of the way to accomplishing this. For example, that if one inserted a "new day" indicator "99" into a sequence and used perl's {n,m} function ("at least n occurrences and less than m occurrences"), then could we specify a pattern

```
AB and BA meet (WEIS 03) within 7 days
```

using the regular expression

```
A03B 99{0,7} B03A
```

This still doesn't deal with the problem of partial ordering, however, and a useful notation would probably need to explicitly deal with the issue of dates rather than trying to simulate them as part of the event sequence.[22]

Notation may or may not be important—again, one can always specify patterns using computer programs. However, one cannot help but notice the extent to which a robust pattern-specification makes it easier to get work done. In our project, perl programs are generally about one-tenth the length of their C counterparts, and most of that difference comes from the ability to use regular expressions. An analogy on the advantages of getting a good notation can also be seen

[22] And while we're working on wish-lists, a facility for dealing with the hierarchical event coding structure found in WEIS, BCOW, IDEA, and MEDB would also be helpful.

in the comparison between classical Greek geometry and analytical geometry. A unit circle drawn with a compass describes the same object as the equation $1 = x^2 + y^2$, but one can do a lot more with the equation.[23]

[23] Proponents of the computer language LISP, another sophisticated formalism for working with lists and strings, also made claims about the ten-to-one ratio of LISP to C code that accomplished comparable tasks. And for yet another analogy, most historians of science contend that in the late 18th century, the development of mathematics in continental Europe dramatically outpaced that in England in part because of the superiority of Leibniz's notation for the differential calculus over that of Newton, despite the mathematical equivalence of the two systems. In this frame of reference, our notation of dealing with event sequences is probably about at the level of Roman numerals.

Bibliography

Alker, Hayward R., Jr., and Frank L. Sherman. 1982. "Collective Security-Seeking Practices Since 1945," In *Managing International Crises*, ed. Daniel Frei. Beverly Hills, CA.: Sage.

Allison, Paul D. 1984. *Event History Analysis*. Beverly Hills: Sage

Azar, Edward E. 1982. *The Codebook of the Conflict and Peace Data Bank (COPDAB).* College Park, MD: Center for International Development, University of Maryland.

Beck, Nathaniel, Jonathan N. Katz and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42, 4: 1260-88.

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What To Do (and Not To Do) with Time-Series Cross-Section Data." *American Political Science Review* 89, 3: 634-47.

Beck, Nathaniel, Gary King, and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture," *American Political Science Review* 94, 1: 21-36.

Bennett, D. Scott. 1997. "Testing Alternative Models of Alliance Duration, 1816-1984." *American Journal of Political Science* 41, 4: 846-878.

Bennett, D. Scott. 1999. "Parametric Models, Duration Dependence, and Time-Varying Data Revisited." *American Journal of Political Science* 43, 1: 256-270.

Bercovitch, Jacob. 1996a. "Understanding Mediation's Role in Preventative Diplomacy." *Negotiation Journal* 12, 3: 241-258.

Bercovitch, Jacob, ed. 1996b. *Resolving International Conflict: The Theory and Practice of Mediation.* Boulder, CO: Lynne Rienner Publishers.

Bercovitch, Jacob and Allison Houston. 2000. "Why Do They Do It Like This? An Analysis of the Factors Influencing Mediation Behavior in International Crises." *Journal of Conflict Resolution* 44, 2: 170-202.

Bercovitch, Jacob, J. and Gerald Schneider. 2000. "Who mediates? The Political Economy of International Conflict Management." *Journal of Peace Research* 37, 2: 145-165.

Bercovitch, Jacob, and Richard Wells. 1993. "Evaluating mediating strategies: A theoretical and empirical analysis." *Peace and Change* 18, 1: 3-25.

Bloomfield, Lincoln P., and Amelia C. Leiss. 1969. *Controlling small wars; a strategy for the 1970's*. New York: Knopf.

Bloomfield, Lincold P., and Allan Moulton. 1997. *Managing International Conflict.* New York: St. Martin's Press.

Blossfeld, Hans-Peter, Alfred Hamerle, and Karl Ulrich Mayer. 1989. *Event History Analysis.* Hillsdale, NJ: Lawrence Erlbaum.

Blossfeld, Hans-Peter and Götz Rohwer. 1995. *Techniques of Event History Modeling.* Mahwah, NJ: Lawrence Erlbaum.

Bond, Doug, J. Craig Jenkins, Charles L. Taylor and Kurt Schock. 1997. "Mapping Mass Political Conflict and Civil Society: The Automated Development of Event Data." *Journal of Conflict Resolution* 41, 4: 553-579.

Box-Steffensmeier, Janet M., and Bradford S. Jones. 1997. "Time is of the Essence: Event History Models in Political Science. *American Journal of Political Science* 41, 4: 1414-1461

Box-Steffensmeier, Janet M. and Christopher Zorn. 1999. "Modeling Heterogeneity in Duration Models." Paper presented at the 1999 Summer Meeting of the Political Methodology Society, July15-17, 1999, Texas A&M University, College Station, TX.

Bueno de Mesquita, Bruce. 2001. "International Studies Association Presidential Address, Domestic Politics and International Relations." Text of speech delivered at the International Studies Association, Chicago, February 22, 2001 (used with permission of the author).

Butterworth, Robert Lyle, with Margaret E. Scranton. 1976. *Managing Interstate Conflict, 1945-74: Data With Synopses.* Pittsburgh: University Center for International Studies.

Carment, David and Dane Rowlands. 1998. "Evaluating Third-Party Intervention in Intrastate Conflict." *Journal of Conflict Resolution* 42,5: 572-599.

Chatfield, C. 1989. *The Analysis of Time Series: An Introduction.* London: Chapman and Hall.

Conrad, Burkhard and Klaus Schlichte. 2001. "The Hamburg experience or—Quantitative research: four limits and one alternative." Paper presented at 'The Uppsala Conflict Data Conference' 8-9 June 2000, Uppsala/Sweden

Davies, John L., and Chad K. McDaniel. 1993. "The Global Event-Data System." In *International Event-Data Developments: DDIR Phase II*, ed. Richard L. Merritt, Robert G. Muncaster, and Dina A. Zinnes. Ann Arbor: University of Michigan Press.

Dixon, William J. 1996. "Third-party technique for preventing conflict escalation and promoting peaceful settlement." *International Organization* 50, 4: 653-81.

Fisher, Roger, and William Ury. 1978. *International mediation, a working guide: ideas for the practitioner*. New York: International Peace Academy.

Fisher, Roger, Andrea Kupfer Schneider, Elizabeth Borgwardt, and Brian Ganson. 1997. *Coping with International Conflict: A Systematic Approach to Influence in International Negotiation.* New York: Prentice-Hall.

Gerner, Deborah J., Philip A. Schrodt, Ronald A. Francisco, and Judith L. Weddle. 1994. "The Machine Coding of Events from Regional and International Sources." *International Studies Quarterly* 38, 1: 91-119.

Gerner, Deborah J.and Philip A. Schrodt. 1998. "The Effects of Media Coverage on Crisis Assessment and Early Warning in the Middle East." In *Early Warning and Early Response,* ed. Susanne Schmeidl and Howard Adelman. New York: Columbia University Press-Columbia International Affairs Online.

Gerner, Deborah J. and Philip A. Schrodt. 2001. "Analyzing the Dynamics of International Mediation Processes in the Middle East and the former Yugoslavia." Paper presented at the annual meeting of the International Studies Association, Chicago.

Gerner, Deborah J., and Ian S. Wilbur. 2000. “Semantics or Substance? Showdown between the United States and the Palestine Liberation Organization.” Pew Case Studies in International Affairs, Institute for the Study of Diplomacy, Georgetown University.

Goldstein, Joshua S., and John R. Freeman. 1990. *Three-Way Street: Strategic Reciprocity in World Politics*. Chicago: University of Chicago Press.

Goldstein, Joshua S. 1992. “A Conflict-Cooperation Scale for WEIS Events Data.” *Journal of Conflict Resolution* 36, 3: 369-385.

Goldstein, Joshua S., and Jon C. Pevehouse. 1997. “Reciprocity, Bullying and International Cooperation: A Time-Series Analysis of the Bosnia Conflict.” *American Political Science Review* 91, 3: 515-530.

Goldstein, Joshua S. , Jon C. Pevehouse, Deborah J. Gerner, and Shebli Telhami. forthcoming. “Dynamics of Middle East Conflict and US Influence.” *Journal of Conflict Resolution*

Haas, Ernst B. 1967. *Collective Security and The Future International System.* Denver: University of Denver Monograph Series in World Affairs 5, 1.

Haas, Ernst B. 1986. *Why We Still Need The United Nations.* Berkeley, CA: Institute of International Studies.

Haas, Ernst B, Robert Lyle Butterworth, and Joseph S. Nye. 1972. *Conflict Management by International Organizations.* Morristown, N.J.: General Learning Press.

Hamilton, James D. 1994. *Time Series Analysis.* Princeton: Princeton University Press.

Hudson, Valerie, ed. 1991. *Artificial Intelligence and International Politics*. Boulder: Westview

Huxtable, Phillip A. 1997. *Uncertainty and Foreign Policy-Making: Conflict and Cooperation in West Africa.* Ph.D. dissertation, University of Kansas.

Huxtable, Phillip A. and Jon C. Pevehouse. 1996. “Potential Validity Problems in Events Data Collection.” *International Studies Notes* 21, 2: 8-19.

Laurance, Edward J. 1990. “Events Data and Policy Analysis.” *Policy Sciences* 23:111-132.

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

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* Cambridge: Cambridge University Press.

King, Robert O. Keohane and Sidney Verba. 1994. *Designing Social Inquiry*. Princeton: Princeton University Press.

Kaldor, Mary. 1999. *New and Old Wars*. Stanford: Stanford University Press.

Kleiboer, Marieke. 1996. “Understanding Success and Failure in International Mediation.” *Journal of Conflict Resolution* 40, 2: 360-389.

Kovar, K., J. Fürnkranz, J. Petrak, B. Pfahringer, R. Trappl, and G. Widmer. 2000. “Searching for Patterns in Political Event Sequences.” *Cybernetics and Systems* 31, 6.

Maller, R. A., and S. Zhou. 1996. *Survival Analysis with Long-Term Survivors*. New York: Wiley.

McClelland, Charles A. 1970. "Some Effects on Theory from the International Event Analysis Movement." mimeo, University of Southern California, February, 1970

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

McClelland, Charles A. 1983. Let the User Beware. *International Studies Quarterly* 27,2 :169-177.

Mooradian, Moorad and Daniel Druckman. 1999. "Hurting Stalemate or Mediation? The Conflict over Nagorno-Karabakh, 1990-1995." *Journal of Peace Research* 36, 6: 709-727.

Moore, Will H. 1995. "Action, Reaction or Rational Expectations? Reciprocity and the Domestic: International Conflict Nexus during the Rhodesia Problem." *Journal of Conflict Resolution* 39, 1: 129-167.

Nye, Joseph S. 1968. *Peace in Parts.* Boston: Little, Brown and Company.

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, 4: 538-546.

Schrodt, Philip A. 1989. "Short Term Prediction of International Events using a Holland Classifier." *Mathematical and Computer Modeling* 12: 589-600.

Schrodt, Philip A. 1990. "Parallel Event Sequences in International Crises." *Political Behavior* 12:97-123.

Schrodt, Philip A. 1991a. "Classification of Interstate Conflict Outcomes using a Bootstrapped ID3 Algorithm." *Political Analysis* 2: 31-56.

Schrodt, Philip A. 1991b. "Prediction of Interstate Conflict Outcomes Using a Neural Network." *Social Science Computer Review* 9, 4: 359-380.

Schrodt, Philip A. 1999. "Early Warning of Conflict in Southern Lebanon using Hidden Markov Models." In *The Understanding and Management of Global Violence,* ed. Harvey Starr. Pp. 131-162. New York: St. Martin's Press.

Schrodt, Philip A. 2000. "Pattern Recognition of International Crises using Hidden Markov Models." In *Political Complexity: Nonlinear Models of Politics,* ed. Diana Richards. Pp. 296-328. Ann Arbor: University of Michigan Press.

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

Schrodt, Philip A., and Deborah J. Gerner. 1997. "Empirical Indicators of Crisis Phase in the Middle East, 1982-1995." *Journal of Conflict Resolution* 41, 4: 529-552.

Schrodt, Philip A., and Deborah J. Gerner. 2000. "Cluster-Based Early Warning Indicators for Political Change in the Contemporary Levant." *American Political Science Review* 94,4.

Schrodt, Philip A., Shannon G. Davis, and Judith L. Weddle. 1994. "Political Science: KEDS—A Program for the Machine Coding of Event Data." *Social Science Computer Review* 12, 3: 561-588.

Sekhon, Jasjeet, and Walter R. Mebane, Jr. 1998. "Genetic Optimization Using Derivatives: Theory and Application to Nonlinear Models." *Political Analysis* 7.

Sherman, Frank L. 1994. "SHERFACS: A Cross-Paradigm, Hierarchical and Contextually Sensitive Conflict Management Data Set." *International Interactions* 20, 1-2: 79-100.

Sherman, Frank L. 1987. *Partway to Peace: The United Nations and the Road to Nowhere*. Ph.D. Dissertation, The Pennsylvania State University.

Sherman, Frank L., and Laura Neack. 1993. "Imagining the Possibilities: The Possibilities of Isolating the Genome of International Conflict From the SHERFACS Dataset." In *Theory and Management of International Event Data: DDIR Phase II*, ed. Richard L. Merritt, Robert G. Muncaster, and Dina A. Zinnes. Pp. 87-112. Ann Arbor: University of Michigan Press.

Sommer, H. and J. Scarritt. 1998. "The Utility of Reuters for Events Analysis in Area Studies: The Case of Zambia-Zimbabwe Interactions, 1982-1993." *International Interactions* 25: 29-59.

Thomas, G. Dale. 1999a. *The "Strange Attractiveness" of Protracted Social Conflict in Northern Ireland.* Ph.D. dissertation, University of South Carolina.

Thomas, G. Dale. 1999b. "Data Aggregation Intervals: A Key Factor Or Convenient Choice?" Manuscript, Bowling Green State University (daletho@bgnet.bgsu.edu)

Tomlinson, Rodney G. 1993. World Event/Interaction Survey (WEIS) Coding Manual. Mimeo, Department of Political Science, United States Naval Academy, Annapolis, MD.

Trappl, Robert, J. Fürnkatz, J. Petrak, and J. Bercovitch. 1997. "Machine Learning and Case-Based Reasoning." In *Learning, Networks and Statistics*, ed. Riccia, G.D., R. Kruse and H-J. Lenz. Vienna: Springer-Verlag.

Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science.* Ithaca: Cornell University Press.

Wall, Larry, Tom Christiansen, and Jon Orwant. 2000. *Programming Perl* (3rd. ed.). Sebastpol, CA: O'Reilly.

Ward, Michael and Sheen Rajmaira. 1992. "Reciprocity and Norms in U.S.-Soviet Foreign Policy." *Journal of Conflict Resolution* 36, 2: 342-368.

Weiss, Thomas G. 1999. *Military–Civilian Interactions.* Lanham, MD: Rowman & Littlefield.

Wickbolt, A., J. Bercovitch and S. Piramuthu. 1999. "Dynamics of International Mediation: Analysis using Machine-Learning Methods." *Conflict Management and Peace Science* 17, 1: 49-68.

Wood, B. Dan, and Jeffrey S. Peake. 1998. "The Dynamics of Foreign Policy Agenda Setting." *American Political Science Review* 92, 1: 173-184.

Appendix I: Cross-Correlation

Cross-correlation is useful in determining if a behavior has a long-term effect when the likely timing of that effect is not specified by the theory. The technique is not a widely used in political science and some explanation is in order.

The cross-correlation function is similar—but not identical—to computing the Pearson product moment "r" between x_t and $y_{\pm k}$ for various values of k. Both statistics have the form

$$r = \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}(x)\mathrm{Var}(y)}}$$

In a cross-correlation, Var(x) and Var(y) are estimated from the entire sample, whereas in a Pearson product moment these variances are computed only on the cases that were used to compute the covariance. Note that the "cross-correlograms" are *not* a time series giving the effect of a single mediation on subsequent behavior; they are a correlation of the mediation with prior and future behavior for the entire time period. For additional information on cross-correlation, see Kendall (1973: 129), Chatfield (1989: 136), and Gottman (1981: 318).

The *approximate* critical value of the cross-correlation coefficient at the 5% two-tailed significance level is $\pm 2/\sqrt{N}$, which is roughly 0.13 for the Levant case and 0.18 for the Balkans case. However, these correlograms have been computed on the raw series rather than the detrended and pre-whitened series (see Chatfield 1989: 137-140) so the correlation may be over-estimated. Consequently, these statistics should be interpreted as primarily descriptive rather than inferential. Figures 4.1, 4.2 and 4.3 show Monte-Carlo estimates of the 0.025% confidence bands for N=128 that were computed by the authors.

We ran cross-correlations on detrended variables, and the results are generally consistent with those found in computations using the raw data; those supplementary correlograms can be found at the KEDS web site. The detrended series, however, still contain autocorrelation at a lags of one month, and sometimes two months, so detrending alone is insufficient to produce a white noise process.

At this point, the cook-book approach would be to continue to process the data until we had "whitened" it. This can be done, but every step in the sequence of standard time-series transformations that improve the statistical characteristics of the estimators—removal of trend, removal of autocorrelation, and the like—also take the data and the analysis further from anything that an analyst can actually understand. For example, when trend and autocorrelation are removed from the time series for US mediation and Serbia-Bosnia conflict (using detrending, then first-difference), then resulting correlogram still has significant negative correlations at lags of -17, -10, -7. -4, -2, -1 and a lead of +1, and positive correlations at leads of +6 and +13. These results are generally tell the same story as the un-transformed data—U.S. mediation responds to past period of high conflict, and has a positive effect on later cooperation.

But the two series on which the statistically correlogram was computed are almost impossible to explain (try it in English…) and one *cannot* say that the correlogram implies that U.S. mediation has a positive effect only at six months and thirteen months. The correlogram implies this is true from detrended and differenced values of that series, a set of transformations that is nearly meaningless from the perspective of figuring out the underlying behavior. The only advantages gained from the transformations are improved analytical properties of the estimators (and even these are just asymptotic approximations). From the perspective of figuring out what was happening in the Balkans during 1990-1999—US mediation improved the situation, UN mediation made it worse—the original data are more useful.

Given the advances in computing power, an alternative approach to this problem is to estimate the significance levels using Monte Carlo simulation. Some examples of this are shown in Table A.1, which shows approximations of the 5% significance bounds. These are based on 30,000 Monte Carlo trials, but required only a minute or so of computing time.

The $2/\sqrt{N}$ and "no autocorrelation" rows compare the standard approximation to the Monte Carlo estimate under the assumption that there is no autocorrelation in either of the series. As it turns out, the approximation is conservative by about 0.02. The next two rows show Monte Carlo estimates for the estimated autocorrelation structure in the Bosnian and Israel-Palestinian cases. Despite significant autocorrelation, the bounds for the Israel-Palestinian case are just the same as those in the $2/\sqrt{N}$ approximation. The much higher autocorrelation in the Bosnian case, in contrast, widens the bounds substantially.

A further advantage of the Monte Carlo approximation is that it is possible to empirically estimation significance bounds for conditions that would be virtually impossible to estimate analytically. The final two rows empirically confirm our assertion in the discussion that while the individual cross-correlations for the USA mediation in Bosnia are only barely significant, the likelihood that multiple, consecutive cross-correlations would all have a high value is much less likely. The bounds for $\min_{j=-1,0,1}(CCF_{k+j})$ show the 5% critical value for the *minimum* of three consecutive cross-correlations. As expected, the USA-Bosnia case is well above this level. By way of contrast, $\max_{j=-1,0,1}(CCF_{k+j})$, shows the 5% critical value for the *maximum* of three consecutive cross-correlations; none of our correlations pass this much stricter criterion.

Table A.1. Monte-Carlo critical value estimates of 5%, two-tailed significance level for cross-correlation under various assumptions

	N=128	N=256
$2/\sqrt{N}$	0.18	0.13
no autocorrelation	0.16	0.11
$y_t = 0.85\ y_{t-1} + e$ [ALLBFR] $x_t = 0.38\ x_{t-1} + e$	0.22	0.16
$y_t = 0.23\ y_{t-1} + e$ [ISRPAL] $x_t = 0.57\ x_{t-1} + e$	0.18	0.14
$\min_{j=-1,0,1}(CCF_{k+j})$	0.07	0.05
$\max_{j=-1,0,1}(CCF_{k+j})$	0.20	0.14

In short, specification of the null model for this data is going to be a complicated process,. While the tendency in statistical analysis in the past has been to pound on the data until it fits some analytical model with known properties (because in the absence of such pounding, one would know nothing about the estimators), contemporary computationally-intensive statistical analysis allow one to empirically approximate the properties of estimators based on a much broader set of assumptions. These empirical approximations may, in fact, be more accurate than the analytical approximations, which often are only valid asymptotically.

Appendix II: World Event Interaction Survey Events (WEIS)

Cue code	Secondary code		Goldstein scale value
01	YIELD		
	011	Surrender, yield to order, submit to arrest, etc.	0.6
	012	Yield position; arrest; evacuate	0.6
	013	Admit wrongdoing; retract statement	2.0
02	COMMENT		
	021	Explicit decline to comment	-0.1
	022	Comment on situation-pessimistic	-0.4
	023	Comment on situation-neutral	-0.2
	024	Comment on situation-optimistic	0.4
	025	Explain policy or future position	0.0
03	CONSULT		
	031	Meet with; at neutral site; or send note	1.0
	032	Visit; go to	1.9
	033	Receive visit; host	2.8
04	APPROVE		
	041	Praise, hail, applaud, condolences	3.4
	042	Endorse others policy or position give verbal support	3.6
05	PROMISE		
	051	Promise own policy support	4.5
	052	Promise material support	5.2
	053	Promise other future support action	4.5
	054	Assure; reassure	2.8
06	GRANT		
	061	Express regret; apologize	1.8
	062	Give state invitation	2.5
	063	Grant asylum	-1.1
	064	Grant privilege, diplomatic recognition; etc	5.4
	065	Suspend negative sanctions; truce	2.9
	066	Release and/or return persons or property	1.9
07	REWARD		
	071	Extend economic aid (for gift and/or loan)	7.4
	072	Extend military assistance	8.3
	073	Give other assistance	6.5
08	AGREE		
	081	Make substantive agreement	6.5
	082	Agree to future action, agree to meet, to negotiate	3.0
09	REQUEST		
	091	Ask for information	0.1
	092	Ask for policy assistance	3.4
	093	Ask for material assistance	3.4
	094	Request action; call for	-0.1
	095	Entreat; plead; appeal to; help me	1.2

10	PROPOSE		
	101	Offer proposal	1.5
	102	Urge or suggest action or policy	-0.1
11	REJECT		
	111	Turn down proposal; reject protest, threat, etc.	-4.0
	112	Refuse; oppose; refuse to allow	-4.0
12	ACCUSE		
	121	Charge; criticize; blame; disapprove	-2.2
	122	Denounce; denigrate; abuse	-3.4
13	PROTEST		
	131	Make complaint (not formal)	-1.9
	132	Make formal complaint or or protest	-2.4
14	DENY		
	141	Deny an accusation	-0.9
	142	Deny an attributed policy, action, or position	-1.1
15	DEMAND		
	151	Issue order or command, insist; demand compliance, etc	-4.0
16	WARN		
	161	Give warning	-3.0
17	THREATEN		
	171	Threat without specific negative sanctions	-4.4
	172	Threat with specific nonmilitary sanctions	-5.8
	173	Threat with force specified	-7.0
	174	Ultimatum; threat with negative sanctions and time limit	-6.9
18	DEMONSTRATE		
	181	Nonmilitary demonstration; walk-out on	-5.2
	182	Armed force mobilization, exercise and/or display	-7.6
19	REDUCE RELATIONSHIP (as negative sanctions)		
	191	Cancel or postpone planned event	-2.2
	192	Reduce routine international activity	-4.1
	193	Reduce or cut off aid or assistance	-5.6
	194	Halt negotiations	-3.8
	195	Break diplomatic relations	-7.0
20	EXPEL		
	201	Order personnel out of country	-5.0
	202	Expel organization or group	-4.9
21	SEIZE		
	211	Seize position or possessions	-9.2
	212	Detain or arrest person(s)	-4.4
22	FORCE		
	221	Non-injury destructive act	-8.3
	222	Nonmilitary injury; destruction	-8.7
	223	Military engagement	-10.0

Source: McClelland and Young (1969:29); Goldstein 1993