

Project Description

This project will investigate the dynamics of third-party international mediation in the Middle East, the former Yugoslavia, and West Africa over the past two decades using statistical time-series models of political event data. We will also extend this analysis to a number of historical and contemporary conflicts that are represented in data sets already collected by other researchers.

Following the end of the Cold War the international system has become increasingly vulnerable to sudden outbreaks of serious systematic violence, both intrastate and international. Iraq's invasion of Kuwait, the conflict between Armenia and Azerbaijan, the genocidal violence observed in Bosnia and Rwanda, and the violent internal conflicts in Somalia, Chechnya, Haiti, Algeria, and Liberia are all examples of this. The disappearance of communism as an ideological principle for organizing conflict (and the disappearance of coercive institutions devoted to suppressing ethnic conflict) appears to have "removed the lid" from long-simmering regional and ethnic tensions and stimulated lethal disputes.

One of the most conspicuous responses of the international community to this violence has been mediation. *Mediation is a specific type of political activity that highlights the role of a third party in facilitating a negotiation process.* This "third party" may be a government official whose country is not a direct party to the dispute, someone associated with an international body like the United Nations, or, as in the case of Track II or unofficial diplomacy, representatives of a nongovernmental organization such as the Quakers or an individual such as Jimmy Carter who may or may not be a citizen of one of the conflictual states (Bercovitch & Houston 2000; Irani 1999). Mediation tends to be less formal than arbitration and adjudication. While there are a variety of roles that a mediator may play, s/he generally does not attempt to *impose* a settlement (unlike some other forms of third-party intervention).

Mediation is an extremely common and often but not always successful, form of conflict management. For instance, Butterworth's (1976) study of 310 international conflicts between 1945 and 1974 found that in 82% of the cases there was some form of mediation. In 1983, Kal Holsti reported results of a study that found mediation occurred in 45 percent of 94 post-WWII disputes examined (cited by Bercovitch 1997, 131). In 70% of the disputes that involved mediation, the outcome of the mediation was at least partially "successful." More recently Bercovitch & Houston (1996) identified 241 international disputes in the 1945-1990 period, of which 137 (57%) were mediated at least once.

The literature on mediation generally encompasses three approaches (for general surveys, see Bercovitch & Rubin 1992; Kleibor 1996; Bercovitch & Houston 1996; Starkey, Boyer, & Wilkenfeld 1999). First, there are many, many case studies (Princen 1992; Zartman 1995; Crocker, Hampson, & Aall 1999; Greenberg, Barton & McGuinness 2000). While some are theoretically-driven (e.g., Wehr & Lederach 1991, Mitchell and Webb 1988), many are fascinating but purely historical narratives. Second, there is a "wisdom literature" on mediation (which typically spans the gamut from negotiating with five-years-olds, through the massive literature on legal, marital and industrial negotiations, and international negotiation) that is also long on antecedents and short on consequences. Fisher's work (Fisher & Ury 1978; Fisher et al. 1997) is probably the best known example of this genre; others include Lall (1985); Brus (1995); Zartman & Rasmussen (1997); and Mitchell (2000). Finally, there is a small quantitative literature that systematically collects data on mediation and tests hypotheses using inferential statistics; this will be discussed below.

Conflicting Claims about Successful Mediation

Finding testable theories in the literature on the correlates of successful mediation can be a bit frustrating. Authors will readily identify what they view to be important variables, but they are much more reluctant to commit to specifying directional relationships. Furthermore, taken as a whole, the field is particularly rife with contradictory assertions:

- ❖ The best mediators are neutral *vs.* the best mediators have a stake in the issue;
- ❖ Successful mediation is most likely when power is equal because parties have maximum flexibility *vs.* success is most likely when power is unequal because uncertainty is minimized;
- ❖ Mediated negotiation works best early in the process, before the parties have taken firm positions *vs.* mediated negotiation works best after the parties have grown tired of the unresolved conflict;
- ❖ Mediation is more likely to be successful when there is a positive long-term relationship between the mediator and the disputant(s) *vs.* friendship between the mediator and disputant(s) may make mediation more difficult;
- ❖ A mediator's leverage over the disputant(s) is an important asset to the mediation task *vs.* a mediator's lack of leverage will enhance the prospects for successful mediation;
- ❖ Active mediation strategies are more effective in international mediation *vs.* mediation will be most successful when the mediator acts as a facilitator and channel of communication rather than using directive strategies.

This ambiguity is not unique to the mediation literature—look at the debate within balance of power theories about whether alliances inhibit or exacerbate conflict or whether extended deterrence actually works—but because of the limited number of systematic studies, the field has not exhibited a great deal of closure.

Quantitative Studies of International Mediation

The quantitative study of international mediation dates back to the 1960s. The initial work was done by Haas (1968, 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 (1971) 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; Bercovitch & Houston 1996; Trapp 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.

In addition to these major projects, several other studies have been recently published based on smaller data sets that focus on a limited number of crises. For example, Ayres (1997) looks at the quantitative dynamics of image change in three conflicts; Mooradian & Druckman (1999) do the same for six mediation attempts in the Nagorno-Karabakh dispute. Carment and Rowlands (1998) develop a game theoretic model based on salience, intensity, capability and expected gains for the belligerent, and assess six 1990s crises against the model.

The objective of our proposed 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 on mediation 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 variable emphasized in the existing quantitative studies and the micro-level advice to individual negotiators that is found in the “wisdom literature” and the case studies.

Event Data and the KEDS Project

This project will use political “event data” to study dynamic mediation processes. Event data—nominal or ordinal codes recording the interactions between international actors as reported in the open press—break down complex political activities into a sequence of basic building blocks (e.g., comments, visits, rewards, protests, demands, threats, military engagements). When aggregated, these data provide summary measures of political activity. Event data will not substitute for a traditional understanding of a political situation, any more than knowing only the price of a company's stock substitutes for knowing about its product, management, and markets. However, while a stock price does not tell *everything* about a company, it tells *something*. If a company is experiencing problems, its stock price will probably fall; if the company's products are in high demand, the stock price will probably rise. The stock prices quoted in daily papers provide summary measures of thousands for companies that an equivalent amount of text could not convey. Event data serve a similar function. The advantages and disadvantages of event data are discussed in greater detail by Azar, Brody & McClelland (1972); Burgess & Lawton (1972); Sigler, Field, & Adelman (1972); Azar & Ben-Dak (1975); and Merritt, Muncaster, & Zinnes (1993).

The KEDS Project

For the past decade, we have been working on the development of the **Kansas Event Data System (KEDS)**, a computer program that creates event data from machine-readable text (Schrodt, Davis, & Weddle 1994). KEDS is a pattern-matching system using 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. Unlike more complex full parsing, sparse parsing techniques

can be used successfully on unedited news wire text. We have experimented with coding a variety of texts, including specialized regional sources in English and German (Gerner et al. 1994). Most of our work, however, has been with Reuters News Service lead sentences. 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, but KEDS has also been successfully used to code complete stories in regions that are less well reported, such as West Africa (Huxtable 1997).

Although we have used KEDS to create several data sets, most of our work has dealt with the Middle East. The Middle East exhibits some of the most complicated political behavior in the world, with a variety of state and non-state actors vying for influence in the context of the ongoing Arab-Israeli conflict and, until 1990, US-Soviet competition. This extensive foreign policy activity presents a realistic challenge to any system of automated coding because of the quantity and variety of material. We have validated KEDS against both the textual record and human-coded events and found no systematic biases in machine coding (Gerner et al. 1994; Schrod & Gerner 1994). Thomas (1999) found similar results in an independent validation for his KEDS-coded event data on Northern Ireland. When the human and machine-coded data are used in statistical tests, the results are almost indistinguishable except for differences due to the higher number of events in the machine-coded data. An independent test of KEDS by the Protocol for the Analysis of Nonviolent Direct Action (PANDA) project of the Program on Nonviolent Sanctions and Cultural Survival at Harvard's Center for International Affairs found that when coding dictionaries have been optimized for a set of data, the sparse-parsing methods of KEDS can assign event codes to Reuters leads with a 91% accuracy. Typical accuracy is in the range of 80% to 85%, similar to the accuracy of human-coded data (Burgess & Lawton 1972).

In the spring of 2000, Schrod produced a new automated coding system named TABARI—**T**extual **A**nalysis **B**y **A**ugmented **R**eplacement **I**nstructions—that is based on the same sparse-parsing principles as KEDS (and hence can use dictionaries developed for KEDS) but is far faster and more flexible. KEDS was written in Pascal and worked only on the Macintosh operating system; TABARI is written as “open-source” code in ANSI C++ and is available on the Linux, Macintosh, and Windows operating systems. TABARI eliminates some deep-seated idiosyncrasies of KEDS and is about 70-times faster, reducing the time required to recode a data set from hours to minutes or even seconds.

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; this has been part of our collaboration with the PANDA project.

In working with KEDS, we discovered two additional advantages to machine coding. First, it is free of non-reproducible coding biases and is therefore both reliable and transparent. Human coding is 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 its vocabulary and can be modified by the researcher; there is no such record in human coding and thus no ability to address this potential problem.

Second, machine coding allows the researcher to experiment with alternative coding rules that reflect a particular theoretical perspective or interest in a specific set of issues. Using contemporary equipment and software (e.g., TABARI running on a 600 Mhz Pentium III), our 100,000-event Arab-Israeli data set can be completely recoded in about thirty seconds. Historically, the most commonly used event data sets for international relations research have been Azar's (1982) Conflict and Peace Data Bank (COPDAB) and McClelland's (1976) World Event Interaction Survey (WEIS). These 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; they are ill-suited to dealing with ethnic conflict, low-intensity conflict, or multilateral intervention. With machine coding, alternative coding schemes can be implemented and refined with relative ease, as the PANDA project has already demonstrated.

When the KEDS project began in the late 1980s, accurate machine coding was regarded as something that could only be achieved in the distant future. As recently as 1998, an article on early warning dismissed automated coding as something beyond "our current (or foreseeable) knowledge" (Davies & Harff 1998:81). These pessimistic assessments, however, did not take into account "Moore's Law"—the doubling of computer capacity every 18 months—which has made a desktop computer in 2000 roughly 250-times more powerful than a computer in 1988, the year discussions began on NSF's Data Development in International Relations event data project.

With high-capacity computers, automated coded proved to be an imminently tractable problem. During the past five years, machine-coded data have become completely accepted in the political science community:

- ❖ Articles using KEDS-coded data have been published in the top peer-reviewed journals in political science, including the *American Political Science Review* (Goldstein & Pevehouse 1997; Edwards & Wood 1998; Wood & Peak 1999; Schrodt & Gerner 2000), *American Journal of Political Science* (Schrodt & Gerner 1994), *Journal of Conflict Resolution* (Bond et al. 1997; Schrodt & Gerner 1997; Pevehouse & Goldstein 1999), and *International Studies Quarterly* (Gerner et al. 1994).
- ❖ At least four NSF-funded projects have used KEDS-coded event data: Ronald Francisco (Kansas), "New Methods for Collecting Data on Domestic Conflict" (SES-9631229); Joshua Goldstein (American University), "Testing Theories of International Cooperation in Regional Conflicts 1990-1996 (SES-9617157); J. Craig Jenkins (Ohio State), "Mass Political Conflict: Origins and Impact on Political Change" (SES-9710958), and Kelly Kadera (Iowa), "Testing the Power-Conflict Model" (SES-9806123).
- ❖ At least two dissertations have been completed that develop new event data sets using the KEDS system (Huxtable 1997, regional focus West Africa; Thomas 1999, regional focus Northern Ireland), and we know of additional dissertations in progress at the University of Michigan, The Ohio State University, and Texas A&M.
- ❖ Several government projects in the United States and Europe, as well as a United Nations project, have employed machine coding systems (either KEDS or systems derived from KEDS such as TABARI and the commercial VRA coder) in crisis early warning systems.

In a word, the automated coding of event data has been accepted as both a viable—and in most cases, preferable—alternative to traditional human coding.

Proposed New Research

Extensions to TABARI

The current version of TABARI is basically KEDS translated into C++, cleaned up, internally documented, and extended to multiple operating systems. Under the funding of this project, we will work on the following new features:

- ❖ A systematic means of dealing with attribution (SAID, REPORTED, etc.) using a module that will allow the user to specify general parameters for handling this situation;
- ❖ Extensive use of sets of synonyms, so that the coding dictionaries can be organized around politically-equivalent expressions rather than around individual words and phrases;
- ❖ Facilities for measuring the "affect" of a sentence—for example, determining whether a report indicates that an actor thinks the situation is getting better or worse;
- ❖ Integration of automatic actor detection into the main program (it is currently a separate module).

Because TABARI has been developed under the GPL open-source license, the source code as well as the compiled program will be made available to the research community as soon as we are confident that these changes function as documented. We will also modify the coding dictionaries to take advantage of the new features. We have a small grant from the University of Kansas that will allow us to begin some of this work in the fall semester of 2000.

Event Data Coding

As noted above, most work in event data analysis has been done with either the WEIS or COPDAB event coding schemes, which were originally developed in the 1960s. The actors and activities that characterize contemporary crises are frequently quite different than those assumed by these coding systems. For example, almost all post-Cold War crises involve important actors at the sub-national and trans-national level; many involve substantial legitimate and criminal economic activities in addition to military-diplomatic activity. There is a great deal of interest in the event data community in developing new coding schemes to deal with these behaviors.

While we have done all of our earlier work with WEIS, in this project we will work with the new "Integrated Data for Events Analysis" (IDEA) framework under development by Charles Taylor, Craig Jenkins, Joe Bond and Doug Bond (Taylor, Jenkins & Bond 1999; Taylor et al. 1999). IDEA incorporates the event forms of several existing systems—most conspicuously WEIS and *World Handbook of Political and Social Indications* (Taylor & Hudson 1972; Taylor & Jodice 1983)—and provides a cross-mapping back to the original protocols. However, the system also benefits from the experience of 40 years of event coding, incorporates resources such as Princeton's "WordNet" synonym set (<http://www.cogsci.princeton.edu/~wn/>), and is far more flexible and comprehensive than the existing event coding schemes.

We anticipate that if properly constructed and validated, the IDEA system could be a stable system rather than yet-another event coding scheme. Event data research has suffered at least as much from "lock-in" as from scheme proliferation—most international relations data are coded in either WEIS or WEIS-derived schemes, despite the fact that WEIS was developed nearly four decades ago, was intended as merely a first try, and has several well-known weaknesses. Our hope is that IDEA will reset the baseline and provide a new, and better, standard for future work. (Jenkins, personal correspondence, 20 February 1999)

At present time, the IDEA researchers have focused primarily on the extension of the coding system to handle internal political activities such as protest. While IDEA, like WEIS, contains some categories relevant to mediation, we anticipate that these will also need to be expanded. IDEA uses a multi-level framework that allows events to be coded at increasing levels of detail, and consequently, we can provide customized dictionaries that provide substantial additional information involving mediation activities without requiring modifications in the overall IDEA scheme. We will coordinate with the Taylor et al. group during this process to insure that we are not introducing incompatibilities; Schrodt is already an outside consultant to this project.

In order to use the IDEA coding system, we will need to revise the coding dictionaries for the TABARI program. This involves two steps. First, we will take our existing dictionaries for coding into WEIS and simply make the appropriate changes in the codes: because IDEA is cross-mapped into WEIS, this should be relatively straightforward. Second, we will incorporate any new vocabulary that has been developed by the IDEA project into the TABARI system. IDEA is currently being developed using a proprietary coder from VRA, Inc., but the dictionaries from the NSF-funded work on IDEA are available. We will translate these to the open-source format of TABARI. Most of this translation can be done automatically. If there are important features of the IDEA system that cannot be readily translated into TABARI, we will add these to TABARI whenever possible within TABARI's sparse-parsing framework.

Variables

Event data provides an extremely rich set of potential variables for the analysis of mediation activities. 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. 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 news-wire sources), as can quite a few of the "contextual" variables.

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. The one potentially important type of behavior that we cannot study is secret negotiations such as those between Israel and the Palestinians that led to the Oslo Accords. Information about such negotiations is usually (if not always reliably) available in the case study literature and we can use that information to cross-check our models. For example, it will be interesting to see if our models predict that something like Oslo should be happening even though the negotiations are not part of the public event record.

As noted above, the literature abounds with suggestions for independent variables. Many of these correspond directly to information coded with traditional event data, including offers of mediation, pressure from allies to support or reject mediation, the promise of economic assistance or imposition of economic sanctions, and the presence or absence of military activity.

As discussed below, we will be using both discrete and scaled event data. In discrete event data, we focus on the occurrence of specific categories of events, e.g. meeting or agreements. Scaled event data, in contrast, aggregate events for a set interval of time (e.g. monthly or biweekly) into a composite measure using a weighted scale such as the one developed by Goldstein (1992) for the WEIS system, or being developed as part of the IDEA system (Taylor et al 1999). Part of our research will also involve experimenting with alternative scales.

Our initial analysis will use event data almost exclusively unless a structural variable is absolutely required to specify a model. Our rationale is three-fold. First, most of the hypotheses that address structural factors are already ably handled by the Bercovitch group; others can be examined through more thorough analysis of data sets such as Butterworth-Scranton, CASCON, or SHERFACS. Second, event data can be inexpensively, consistently, and transparently coded, which allows us to do comparable analyses across multiple cases. (We still have the problem of inconsistent news coverage, particularly in Africa, but this also affects human-coded data.) Finally, the event data are free of the problem of “hind-sight bias” which is an unavoidable risk in human-coded data: knowing the outcome of a mediation can potentially affect how informed coders assign values to the independent variables. This has not necessarily been a problem in previous research—coding rules and coder training were presumably designed to avoid hind-sight bias, and hind-sight is less likely to be a problem in structural studies than time-series studies—but it would be useful to have a check on this.

That said, we will leave open the possibility of using some structural information when it is clearly dictated by theoretical considerations and can be unambiguously coded. Geographical proximity is one obvious variable; gross indicators of military and economic power (e.g., sufficient to distinguish the U.S. from Kuwait or Nigeria from Liberia) might be another. We will certainly consider the distinction between conflicts that are largely internal (civil wars in Lebanon, Liberia, and Sierra Leone), conflicts involving sovereign states (Israel-Jordan, Iran-Iraq, Senegal-Mauritania), and conflicts involving actors who would like to become sovereign (Palestinians, Kurds, Kosovo). We may be able to accommodate these with a few variables, but it may also be necessary to create separate models. We will try to avoid variables that require judgement calls (e.g. is Russia a “superpower” in the Balkans? Nigeria in West Africa?)

Regional Focus

Given the amorphous state of theories of international mediation, we intend to first develop our models in regions where we have a substantial knowledge base, then test the more successful of those models in regions where we will depend on other sources for validation.

Middle East

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. In addition to our quantitative analysis of conflict in the region, Gerner has done field work in the Middle East for close to twenty years and has published both qualitative and quantitative books and articles on the conflict; she also has access to an extensive network of other scholars who do research on the region.

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. Mediated dyads include:

- Israel and Palestinians
- Israel and Jordan
- Israel and Lebanon
- various parties in the Lebanese civil war

- Israel and Syria
- Syria and Lebanon
- Iraq and various international organizations
- 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 listed under the subject heading “Mediation, international” found that about a third of the case studies dealt with the Middle East.

The Balkans and West Africa

These two contemporary crisis areas have been subject to extensive—but quite different—conflict and mediation efforts. The conflict in the former Yugoslavia was subject to intense international scrutiny from the beginning. Various methods were attempted to mediate or otherwise control instability in the region, including United Nations, regional European, and super-power (U.S.) intervention. Like the Middle East, this conflict contains a variety of dyadic disputes that will probably exhibit somewhat different dynamics. The international news media have covered the Balkans extensively, and a number of case studies are available (e.g. Greenberg and McGuiness 2000; Hanson 2000).

The civil conflicts in West Africa, in contrast, 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 other cases we will study, West Africa is only sporadically covered by the international media (Huxtable & Pevehouse 1996), and case studies of mediation are rare; getting adequate data will be more of a challenge here.

We have done no fieldwork in the Balkans and have only limited experience in West Africa (albeit we inadvertently found ourselves in the middle of a Senegal-Mauritania border dispute earlier this summer). However, the University of Kansas has Department of Education Title VI area studies centers for both Africa and Russia/Eastern Europe and there is substantial expertise on campus upon which we can draw. We also worked earlier with Phillip Huxtable on a dissertation that included an event-data analysis of regional intervention in West Africa. Based on Huxtable’s experience, we know that credible events can be coded for this region, even if the noise level is higher than it is in the Middle East. We already have regionally-specific coding dictionaries available for these areas.

Additional Areas

After we have refined our statistical models on these three regions, we will try to test the models that produced the most coherent results on several other data sets. 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. We are currently considering the following sources:

Other Machine-Coded Data Sets

At the moment, 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. These data sets have been produced through various research projects and dissertation work, and use the WEIS framework or minor variations on WEIS. If the

costs of acquiring news-wire text drops substantially—this could occur at any time due to competition in the commercial market for information—we might also code some additional crises on our own: India-Pakistan, Somalia, and the “Great Lakes” (Rwanda, Zaire, Congo, etc.) are potential targets.

BCOW

The **Behavioral Correlates of War** data set (BCOW; Leng 1987) is a dense, high-quality event data set that focuses on about forty crises over the past two centuries. BCOW uses an extensive set of codes involving mediation activities and comes close to the level of detail found in IDEA. 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.

Episodic Data Sets

As noted earlier, a number of data sets are available that focus primarily on the contextual characteristics of international mediation. Unlike BCOW, these do not contain explicit event data, but they do include some information involving the dynamics of mediation. Depending on the models that we find to be most effective in explaining mediation, we may be able to do some ancillary tests on these data sets, which would considerably extend the temporal and geographical scope of our study. The data sets that appear most promising in this regard are CASCON, SHERFACS, and the Bercovitch data set.

The **Conflict Early Warning Systems** (CEWS) chronologies (Alker, Gurr, & Rupesingh, forthcoming; <http://www.usc.edu/dept/LAS/ir/cis/cews/>) are yet another potential source of new data. CEWS provides a series of narratives describing 20 of the major conflicts that were active in the 1990s such as Chechnya, Chiapas, Kashmir, Sudan, and Tibet.

Designed to illustrate feasibility and usefulness, this prototype reports on 20 conflict cases, using narratives or chronologies provided by an international team of experts in conflict prevention. They were written from a violence diminishing perspective, so that fruitful lessons might be drawn from comparisons of a relatively equal number of relatively successful and unsuccessful cases in the area of the experts' special competence.

(http://www.usc.edu/dept/LAS/ir/cis/cews/html_pages/aboutcews.htm;
accessed 7 August 2000)

While these chronologies are not directly comparable to news-wire based event data—they do not contain multiple reports of events, and generally have a much lower density of events—they still appear to be a potentially valuable resource. In the case of long-running crises such as Kashmir, they contain information going back to the 1940s. The narratives appear quite straightforward to code using TABARI and are already available in machine-readable form on the CEWS web site.

Statistical Models

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

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 existing statistical methods used with event data prior to 1990 were very crude, often little more than contingency table analyses. While there has been some additional use of these methods to analyze political behavior—for example neural networks are used by King and Zeng (e.g. Beck, King & Zeng 2000; King and 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. In addition, pattern recognition algorithms lack a clearly defined inferential model, frequently have poorly-understood properties, and, due to the required investment in specialized or custom-written software, the approach is difficult to use without a substantial knowledge of computer programming.

This project, in contrast, will rely exclusively on the application of statistical techniques. This is motivated partly by the issue of accessibility, but also by the recognition that 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 1999). Consequently, while time series models are still not a perfect fit 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. Because these methods can be implemented with existing statistical packages such as *Stata* and *S-Plus*, we will be able to focus most of our efforts on analysis rather than software development.

This emphasis on inferential statistics is also appropriate given the differences between our earlier focus on prediction and the explanatory nature of this project. Good predictive models do not necessarily involve good explanations; in fact when models with diffuse parameter structures are used (for example hidden Markov models, neural networks, or VAR), there may actually be a tradeoff between effective prediction and coherent explanation. Because this project will be evaluating hypotheses from a rich, if inconsistent, theoretical literature, inference is our primary concern.

We are planning to employ a number of different techniques to explore this data. This is *not* intended as a statistical fishing expedition, but rather a recognition that the current state of the theoretical and empirical literature on this topic is not sufficiently focused to dictate a single statistical model. We are well aware that by employing multiple techniques, we are doing multiple statistical comparisons and hence will “beat the significance test” by mere chance on occasion. Our objective is not to find a single best model, but rather to look at the data from a number of different perspectives and, we hope, derive a coherent, statistically-supported explanation of the mediation process.

Interval-level Time Series

A variety of well-understood time series techniques are available for the forecasting of interval-level data: econometricians have been working on these for decades (e.g. Chatfield 1989; Hamilton 1994; Greene 2000). These methods use lagged values of the independent (and some of the dependent) variables to predict current values. In contrast to many economic time series, aggregated event data are very irregular, but this has not prevented the successful application of time-series methods: Goldstein & Freeman (1990); Ward & Rajmaira (1992); Moore (1995); and Goldstein & Pevehouse (1997) are examples of this approach. Because our work is primarily focused on testing hypotheses suggested by the mediation literature, we anticipate specifying our models based on theoretical considerations, rather than using some of the more general techniques (notably VAR) found in earlier event-based studies of international reciprocity.

To use these methods, the events must first be scaled and aggregated. In the work we have done on forecasting, we have obtained credible results using Goldstein’s (1992) scale for WEIS data. As noted earlier, the development of scaling is an integral part of the development of the IDEA

system, so this scale will be available for the data we code using that system. Following the lead of Goldstein & Pevehouse (1997), we will probably aggregate our data bi-weekly. This gives about 500 data points for the Middle East data, and 250 for the Balkans and West Africa, a sample size that is sufficiently large that most time series estimation methods can be used with a reasonable level of confidence.

We also will explore the effects of the choice of scales. In a suggestive experiment using alternative scales in a forecasting model employing cluster analysis (Gerner & Schrod 1998) we found our results to be remarkably insensitive to the scale used to aggregate the data: the simple presence or absence of events provided roughly 50% of the explanatory power of the model. We want to determine whether this result generalizes to other methods. We will apply for super-computer time at the National Center for Supercomputer Applications to work on this problem because this analysis is computationally intense. Schrod has previously worked with the NCSA systems and we anticipate that we will be able to obtain time on the NCSA systems for this research. If not, the work can be done (albeit more slowly) on existing computers available to the project.

Time Series Cross-Section Logit

We will use logit analysis to estimate models predicting the success or failure of our three dependent variables (mediation, agreement, and behavioral change resulting from the agreement) as a function of various independent variables. This approach is similar to that employed over the past decade in many of the democratic peace studies: Beck, Katz, & Tucker (1998:1260) list 18 such studies and additional analyses employing time series cross-section logit have appeared more recently. Dixon (1996) applies this method in his examination of mediation using the SHERFACS data. Recent work by Beck, Katz & Tucker (1998), King & Zeng (1999) and others has identified a number of ways to modify the standard logit techniques that can improve the properties of the estimators; we will these suggestions into account in designing our studies.

Event History and Duration Models

The final set of analyses will focus on event history and duration models (see Allison 1984; Blossfeld, Hamerle, & Mayer 1989; Blossfeld & Rohwer 1995; Maller & Zhou 1996; Box-Steffensmeier & Jones 1997; Bennett 1999). In these models, 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. Duration models have recently been applied by Bennett to study war duration (Bennett 1997; Bennett & Stam 1996) and similar techniques could be used for the study of mediation.

Linkage with Other Projects

This proposal grows out of work that was initiated in the summer of 1998 during the "Multiple Paths to Knowledge Project" sponsored by the James A. Baker III Institute for Public Policy, Rice University, and the Program in Foreign Policy Decision Making, Texas A&M University. The overall focus of that project was the challenge of studying international mediation. Charles Taber (SUNY-Stony Brook), Jonathan Wilkenfeld (Maryland), and Schrod realized that their current research methods—computer simulation using artificial intelligence algorithms, experimental methods, and event data respectively—presented a natural hierarchy of experimental control. Gerner and Schrod's event data analysis could look at the problem of mediation in a real-world

context, but with no controls on the data. Wilkenfeld and his research group (Wilkenfeld et al. 2000) could use a formal experimental design that could control some aspects of the situation (e.g. the tactics used by the mediator and the presentation of the mediation problem) and measure other aspects, such as the "cognitive complexity" of his subjects. But he could not actually find out *why* the subjects were behaving as they did. Finally, Taber's artificial intelligence simulations (Taber 1992, 1999; Taber and Timpone 1994) would give him complete control of every aspect of the situation, including the problem-solving methods used by the simulated actors.

Papers discussing these three approaches were presented at the 1999 International Studies Association meetings in Washington and 2000 ISA meetings in Los Angeles. We originally considered submitting a joint NSF proposal in this funding cycle. However, Taber and Wilkenfeld are currently involved in other funded research so we decided not to write a joint proposal at this point (although we may do so in the future). While we have designed our project so that it can stand alone, in all likelihood we will be comparing our results with these other projects as well as developing tests that can be directly compared with the experimental and simulation approaches.

Project Timeline

We expect to complete this research in approximately 18 months. The first semester (Spring 2001) will be used primarily for data preparation, including the development of dictionaries for coding the IDEA system, updating the Middle East, Balkans and West Africa data sets, and creating any new IDEA coding categories required for the coding of international mediation. We have received funds from the University of Kansas General Research Fund for a graduate research assistant and one undergraduate coder for the Spring 2001 semester. Both of these students are experienced with KEDS/TABARI dictionary development and will train and supervise the undergraduate coders if this project is funded. Based on our prior experience developing a number of coding dictionaries, we are quite confident that we can have a working set of data by the end of the summer (if not earlier). Gerner will be teaching a graduate seminar on international mediation during the spring semester and will try to interest some of the students in that class to experiment with quantitative methods in their seminar papers.

In summer 2001 we will identify the graduate research assistant for this project and work with this student on locating testable hypotheses in the mediation literature and precisely specifying the statistical models that can be used to test these. If appropriate courses are available, we may also send this student to ICPSR or APSA short courses on time series analysis or event history analysis; Political Science department funds are available for this.

During the fall 2001 semester we will start the statistical analysis, beginning with the Middle East cases and then extending models that appear theoretically and empirically credible to the Balkans and West Africa cases. During this time we will also further refine the coding dictionaries, and extend those dictionaries to work with other cases. The spring 2002 semester and summer 2002 will be used for the overall evaluation of the models, as well as adapting the analysis to alternative data sets such as BCOW, SHERFACS and the CEWS chronologies. During the spring we can also re-run earlier analyses to take advantage of improvements in the dictionaries and the additional data; this will require relatively little time because both the coding and the statistical analysis can be re-done quickly once the initial analysis has been set up.

We anticipate presenting the results of this research at academic conferences, specifically the 2001 and 2002 American Political Science Association meetings, the 2001 and 2002 North American meetings of the International Studies Association, and the Hong Kong meeting of the International Studies Association in summer 2001. We will endeavor to publish the work in peer-reviewed academic publications, as well as the edited books on mediation which are more likely to be read by practitioners the policy community.

Results of Prior NSF Support

Development of Machine-Coded Event Data Techniques for the Analysis of Political Behavior SBR-9410023, Philip A. Schrodt and Deborah J. Gerner

In this project, the Kansas Event Data System (KEDS) automated coding program—originally developed with funds from the Data Development in International Relations project (Merritt, Muncaster and Zinnes 1993)—was enhanced to include more sophisticated handling of verbs, filtering and modification of sentences, facilities for identifying sub-state political agents, and linking information contained in multiple sentences in an article so that complete news stories could be coded. The machine-coded data were used to study conflict in the Levant and West Africa. Our analysis employed both conventional statistical techniques such as factor analysis and regression, and sequential cluster analysis methods developed specifically for the forecasting of political events. These models were very successful in delineating phases of political activity and showed promise in providing early warning indicators for those changes.

Research Products:

In 1996, we established a web site for the KEDS project—<http://www.ukans.edu/~keds>. This site has become the primary mode of outreach for the project and has dramatically increased the accessibility of our work in the academic and policy communities. Roughly 80% of the visits the site are from academic institutions or government users; about 25% are from outside the United States. While some of these are casual browsers (and yes, a few misguided souls looking for information on tennis shoes), the site has been used in political science courses at other institutions, and has proven to be an effective means of assisting graduate students to use the KEDS software and data in seminar and thesis research.

The KEDS computer program was brought to final form during this grant; it has been freely available to the academic community since its inception. KEDS has been used in a number of peer-reviewed articles, dissertations and government projects, as well as four NSF-sponsored projects—Ronald Francisco (Kansas) SES-9631229; Joshua Goldstein (American) SES-9617157; J. Craig Jenkins (Ohio State) SES-9710958; Kelly Kadera (Iowa) SES-9806123. In 1995, KEDS was recognized with the "Best Research Software Award" by the American Political Science Association Computers and Multimedia Section. The web site also contains a number of utility programs that facilitate the coding and analysis of event data, as well as the customized cluster analysis programs that we developed. We have made available both the compiled and source code for these programs.

We have continued to maintain the Middle East data sets in near-real time, and using other sources of funding, extended our data collection to two additional crisis areas, the Balkans and Central Asia. To the best of our knowledge, these data sets and the PANDA data (produced at Harvard using the KEDS program) have been the only large event data sets currently available to the general academic community that cover post-Cold War political behavior.

Publications resulting from NSF support:

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.

Huxtable, Phillip A. 1997. *Uncertainty and Foreign Policy-Making: Conflict and Cooperation in West Africa*. Ph.D. dissertation, University of Kansas. [project provided partial support for the event data collected and analyzed in the dissertation.]

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

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

Schrodt, Philip A., and Deborah J. Gerner. 1998. "Cluster Analysis as an Early Warning Technique for the Middle East, 1979-1996." In *Risk Assessment and Crisis Early Warning Systems*. ed. John Davies and Ted R. Gurr. Pp. 95-107. New York: Rowman and Littlefield.

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

[In addition, we are nearly finished with a book-length manuscript tentatively titled *Analyzing International Event Data: A Handbook of Computer-Based Techniques*. Three chapters are available on our web site and have already been used for graduate instruction at the University of Kansas and elsewhere.]