

# Chapter One

## International Event Data

---

International conduct, expressed in terms of event data, is the chief dependent variable of international relations research. ... [This] starting point is provided as readily by the ordering principle of classical diplomatic history as by the basic concepts of general system analysis. Thus, we may assert that the prime intellectual task in the study of international relations is to account for actions and responses of states in international politics by relating these to the purposes of statecraft or, alternatively, we can say that the problem is to account for the relations among components of the international system by analyzing the characteristics of the various components of that system by tracing recurring processes within these components. [Both definitions] carry about the same information and involve nearly the same range of choices of inquiry and analysis.

Charles McClelland 1970:6

Much of political behavior can be characterized by interactions between political actors, whether individuals or institutions. Politics primarily consists of “who did what to whom”, and most of our information about politics comes in the form of narratives about actions, reactions, and activities. Patterns of political interactions are commonly both used to characterize existing situations—for example in distinguishing peaceful from conflictual periods in the relationship of two states—and are also used as indicators of possible future situations, as when relations between two states are characterized as “deteriorating” or “improving.”

The importance of interactions in political behavior has been recognized since the beginnings of the efforts to systematically study international political behavior using “scientific” approaches. This research has primarily taken the form of the analysis of *event data*—nominal or ordinal codes recording the interactions among international actors as reported in the open press—which break down complex political activities into a sequence of basic building blocks (e.g., comments, visits, rewards, protests, demands, threats, and military engagements). Event data sets were a major focus of quantitative international relations research in the 1960s and

1970s—work that generated Edward Azar's (1982) Conflict and Peace Data Bank (COPDAB) and Charles McClelland's (1976) World Event Interaction Survey (WEIS). Over the past decade, interest in event data analysis has increased as the combination of machine-readable news reports and automated coding have dramatically reduced the costs of generating, customizing, and analyzing event data.

This volume is directed at researchers who are interested in employing event data to study international behavior, though most of the information contained here is equally relevant to the study of domestic behavior. We will cover two general topics. First, we provide a general introduction and survey of the potentials and problems of event data analysis as it exists at the beginning of the 21st century, with a particular emphasis on methods for automated coding of events from electronic sources (data sets developed prior to the 1990s, in contrast, were coded by humans from paper or microfilm sources).

Second, we will discuss a variety of techniques specifically designed for the analysis of event data. These include an assortment of clustering algorithms, the Levenshtein metric, parallel event sequences, and hidden Markov models. These methods, while generally well-developed in fields distant from political science (for example speech recognition and the analysis of protein coding sequences), are relatively inaccessible to the typical political analyst whose primary training has been in econometric techniques. We believe that these methods can be a useful supplement to conventional time-series analysis techniques for characterizing and forecasting event sequences.

The bulk of this book deals with our research over the past ten years with the Kansas Events Data System project (KEDS), or with closely related work such as that of the Protocol for the Analysis of Nonviolent Direct Action (PANDA) project at Harvard. While much of this work has been reported in journal-length articles in various political science publications (Schrodt 1991, 1993, 1999; Gerner et al 1994; Schrodt and Gerner 1994, 1996, 1997; Schrodt, Davis and Weddle 1994), this volume provides the opportunity to put all of the material in one place, and to update some of our earlier observations in light of subsequent experience. For example, while we

initially saw machine coding as an inexpensive though qualitatively inferior alternative to human coding, but ten years of experience has convinced us that the consistency and stability of machine coding make it superior to human-coded data.

The methods we have developed are, necessarily, illustrative rather than definitive. For instance, virtually all of our work has focused on international behavior in the Middle East, a region that is not necessarily typical of international politics in general, but which we've studied and visited for many years. Nonetheless, the methods that we have used in this research, as well as many of the problems that we have encountered, are likely to apply to many other studies.

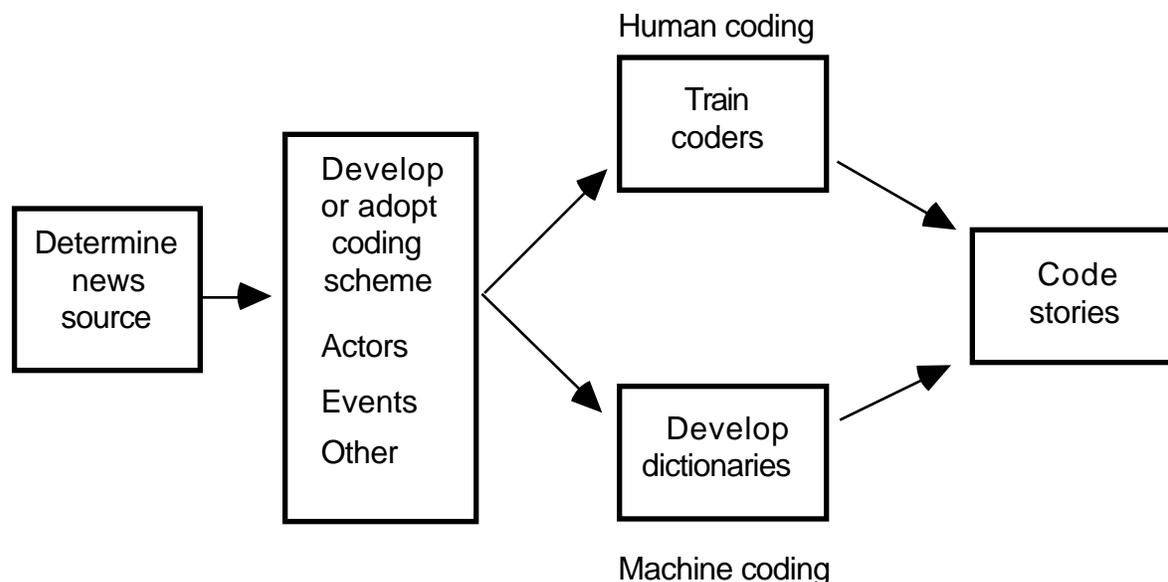
## 1.1. What are event data?

In their simplest form, event data convert natural language reports of political activity to a data set where each entry has the form

date      source      target      event

The "date" is a calendar date; typically this is accurate to within a day or two of when the event occurred. The "source" and "target" are political actors such as nation-states, national leaders, political parties, non-governmental organizations or guerrilla movements. The "event" is a specific code for the type of activity that was reported. Event coding systems may have anywhere from fewer than 20 to more than 100 distinct event categories.

Event data is created through a process of *content analysis* (see Krippendorff 1980; Weber 1990) and involves four steps; these are illustrated schematically in Figure 1.1. In the first step, the researcher identifies a source of news about political interactions. This could be a newswire service such as Reuters or *Agence France Presse* (AFP), an internationally-oriented newspaper such as *The New York Times* or *Times of London*, a set of regional newspapers and news magazines, or a news summary such as *Facts on File* or *Deadline Data on World Affairs*. As we will discuss below, the choice of the event source can have a substantial effect on the number and types of events reported.



**Figure 1.1. Creating an event data set**

Second, a coding system is developed, or a researcher may decide to adopt an existing coding system such as WEIS or COPDAB. Table 1.1 shows a sample of the lead sentences of reports on the Reuters newswire that preceded Iraq's invasion of Kuwait in August 1990. (The full set of reports is considerably more extensive, particularly during the week prior to the invasion.) In most cases, each lead corresponds to a single event, although some sentences generate multiple events. For example, the report "July 23, 1990: Iraqi newspapers denounced Kuwait's foreign minister as a U.S. agent Monday" corresponds to an event in the WEIS event coding scheme: the WEIS category 122 is defined as "Denounce; denigrate; abuse." In this event, Iraq is the *source* of the action and Kuwait is the *target*. Together, this information generates the event record "900723 IRQ KUW 122" where "900723" is the date of the event, IRQ is a standard code for Iraq, KUW is the code for Kuwait, and 122 is the WEIS category. Table 1.2 shows the Reuters stories converted to WEIS events.

---

**Table 1.1. Reuters Chronology of 1990 Iraq-Kuwait Crisis**

---

**July 17, 1990: RESURGENT IRAQ SENDS SHOCK WAVES THROUGH GULF ARAB STATES**

Iraq President Saddam Hussein launched an attack on Kuwait and the United Arab Emirates (UAE) Tuesday, charging they had conspired with the United States to depress world oil prices through overproduction.

**July 23, 1990: IRAQ STEPS UP GULF CRISIS WITH ATTACK ON KUWAITI MINISTER**

Iraqi newspapers denounced Kuwait's foreign minister as a U.S. agent Monday, pouring oil on the flames of a Persian Gulf crisis Arab leaders are struggling to stifle with a flurry of diplomacy.

**July 24, 1990: IRAQ WANTS GULF ARAB AID DONORS TO WRITE OFF WAR CREDITS**

Debt-burdened Iraq's conflict with Kuwait is partly aimed at persuading Gulf Arab creditors to write off billions of dollars lent during the war with Iran, Gulf-based bankers and diplomats said.

**July 24, 1990: IRAQ, TROOPS MASSED IN GULF, DEMANDS \$25 OPEC OIL PRICE**

Iraq's oil minister hit the OPEC cartel Tuesday with a demand that it must choke supplies until petroleum prices soar to \$25 a barrel.

**July 25, 1990: IRAQ TELLS EGYPT IT WILL NOT ATTACK KUWAIT**

Iraq has given Egypt assurances that it would not attack Kuwait in their current dispute over oil and territory, Arab diplomats said Wednesday.

**July 27, 1990: IRAQ WARNS IT WON'T BACK DOWN IN TALKS WITH KUWAIT**

Iraq made clear Friday it would take an uncompromising stand at conciliation talks with Kuwait, saying its Persian Gulf neighbor must respond to Baghdad's "legitimate rights" and repair the economic damage it caused.

**July 31, 1990: IRAQ INCREASES TROOP LEVELS ON KUWAIT BORDER**

Iraq has concentrated nearly 100,000 troops close to the Kuwaiti border, more than triple the number reported a week ago, the Washington Post said in its Tuesday editions.

**August 1, 1990: CRISIS TALKS IN JEDDAH BETWEEN IRAQ AND KUWAIT COLLAPSE**

Talks on defusing an explosive crisis in the Gulf collapsed Wednesday when Kuwait refused to give in to Iraqi demands for money and territory, a Kuwaiti official said.

**August 2, 1990: IRAQ INVADES KUWAIT, OIL PRICES SOAR AS WAR HITS PERSIAN GULF**

Iraq invaded Kuwait, ousted its leaders and set up a pro-Baghdad government Thursday in a lightning pre-dawn strike that sent oil prices soaring and world leaders scrambling to douse the flames of war in the strategic Persian Gulf.

Source: Reuters

---

Table 1.2. WEIS Coding of 1990 Iraq-Kuwait Crisis

Date	Source	Target	WEIS Code	Type of Action
900717	IRQ	KUW	121	CHARGE
900717	IRQ	UAE	121	CHARGE
900723	IRQ	KUW	122	DENOUNCE
900724	IRQ	ARB	150	DEMAND
900724	IRQ	OPC	150	DEMAND
900725	IRQ	EGY	054	ASSURE
900727	IRQ	KUW	160	WARN
900731	IRQ	KUW	182	MOBILIZATION
900801	KUW	IRQ112	REFUSE	
900802	IRQ	KUW	223	MILITARY FORCE

The coding system must specify the set of political interactions that constitute an "event," identify the political actors that will be coded (for example, whether nonstate actors such as international organizations and guerrilla movements will be included in the data set) and establish the categories of events and their codes. Some systems also specify additional information to be coded about the event: For instance, the COPDAB data set codes whether an event is primarily military, economic, diplomatic, or one of five other types of relationship. WEIS codes for specific "issue arenas" such as the Vietnam War, Arab-Israeli conflict, and SALT negotiations. The complete list of event categories for four systems—WEIS, COPDAB, BCOW and IDEA—are found in Appendix 1.

The coding systems in COPDAB and WEIS are comprehensive: they attempt to code all political interactions by all states and some non-state actors during a period of time. In contrast, specialized event data sets such as Hermann and Hermann's (1973) CREON (Comparative Research on the Events of Nations) and Leng's (1987) BCOW (Behavioral Correlates of War) focus on specific subsets of behavior, foreign policy and crises respectively. A variety of domestic and international event data collections, usually focusing on a limited set of actions such as uses of force, domestic violence, or changes of government, are embedded in other data sets

such as Rummel's (1972) DON (Dimensionality of Nations), the *World Handbook* (Taylor and Hudson 1972), and various data sets collected by the research projects of Ted Gurr (for example Gurr 1974). (Peterson (1975) and McGowan et al. (1988) provide a general discussion of these data sets.)

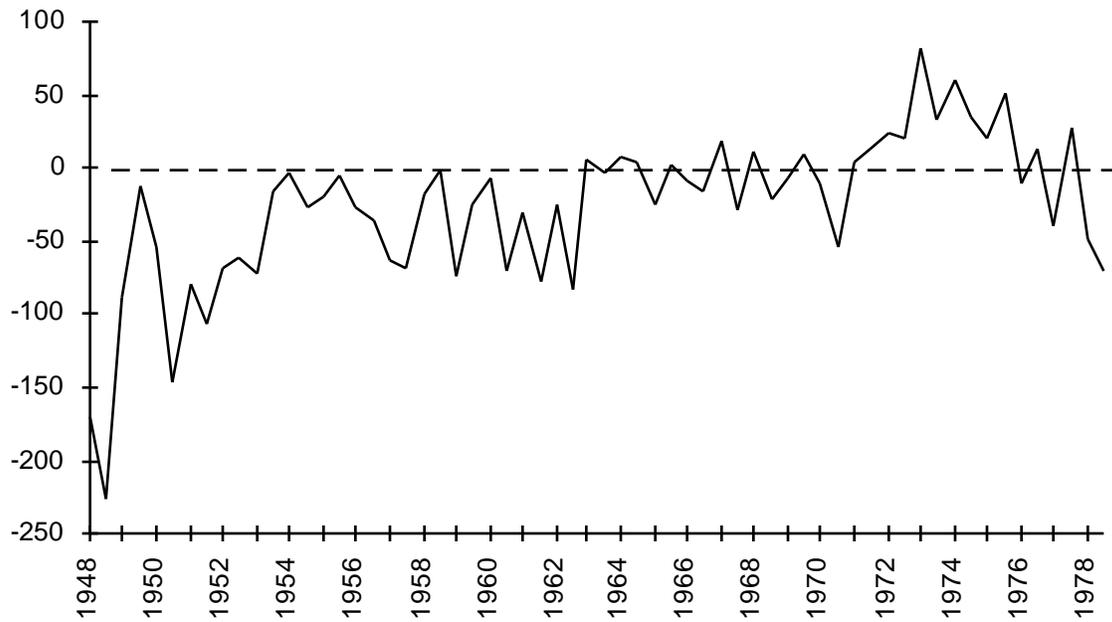
News stories can be coded either by human coders, or with specialized software; the relative merits of these two approaches will be discussed in chapter 2. In a project using human coders, these coding rules are collected into a formal coding manual. In an extended project, these manuals are often fifty or more pages in length and deal with a variety of contingencies that coders may encounter. Coders—typically graduate students or advanced undergraduates in political science—need to be trained so that a news story will be assigned the same codes by any individual coding it. The training stage is frequently quite time consuming but, with sufficient training, most projects train coders to the point where two coders will assign the same code to a news report in 80% to 90% of the cases (see Burgess and Lawton 1972: 58). Human coding is a relatively slow process and most human-coding projects have been maintained over a number of years, with intermittent periods of re-training as coders enter and leave the project.

In contrast, in a machine-coding project, coding rules are implemented through a computer program that employs customized dictionaries of phrases that identify actors and events and associate these with specific codes (see Lehnert and Sundheim 1991; Schrodt, Davis, and Weddle 1994). These dictionaries are typically constructed by having a dictionary developer monitor the codes assigned by the automated system as it goes through a large number of test sentences. Appropriate new vocabulary is added or modified by the developer whenever the program makes a coding error. When the level of coding accuracy reaches a level sufficient for the intended application, the dictionaries are fixed and the entire data set (including the test cases) is re-coded, thus insuring that the same accuracy level is found in the entire data set. The fully-automated coding could take from a few minutes to several hours of computation, depending on the number of stories coded.

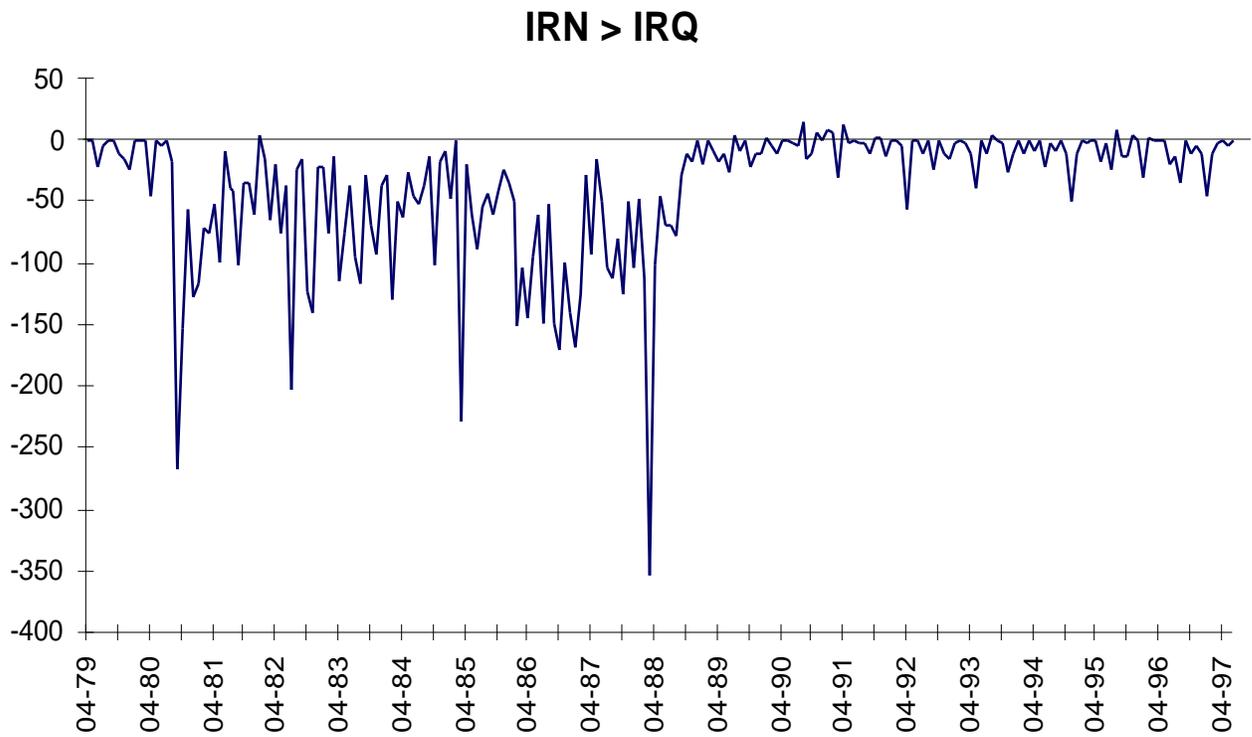
A completed event data set will cover a period of years and typically contain tens of thousands of events. These are then used in statistical or computational analyses. Most statistical studies do not work with the individual events, but instead work with weekly, monthly or yearly aggregations of the events. An example of this type of analysis is found in Chapter 4. This aggregation is done by assigning a numerical value to each type of event using a standardized scale, and then summing those values. Computational analyses, in contrast, work directly on the event sequences without the intermediate step of numerical aggregation; some of these methods will be discussed in Chapters 5 and 6.

Figure 1.2 shows the actions that the United States directed towards the Soviet Union for the period 1948-78 based on the COPDAB scores reported in Goldstein and Freeman (1990:162). (In the COPDAB coding system negative numbers indicate conflictual behavior while positive numbers indicate cooperation.) The events are coded from *The New York Times* and a variety of regional newspaper sources. The COPDAB time series shows three general periods. The early Cold War (1948-62) is characterized by uniformly negative relations, although relations are more stable in the late 1950s than in the early 1950s. A partial thaw occurs in 1962-70 following the Cuban Missile Crisis, with the relationship centered at zero, hence neutral. Finally, the 1970-78 period shows the rise and fall of the Nixon-Kissinger détente policy.

The event data record of U.S.-Soviet interactions corresponds closely to the patterns one would expect from an historical study. Moreover, the event data can also be used to fine-tune that chronology. For example, while Nixon clearly intended to implement a détente policy from the beginning of his administration in 1969, there was continued disagreement between the U.S.A. and U.S.S.R. over the U.S. involvement in Vietnam, the 1968 Soviet invasion of Czechoslovakia, and other issues. Thus the aggregated interactions between the U.S.A. and U.S.S.R. do not become positive until 1971. Positive interactions peak about the time of Nixon's resignation in 1974; the event data scores then decline during the two years of the Ford administration and return to post-Cuban Missile Crisis levels by 1976.



**Figure 1.2. Azar-Sloan-scaled series for USA USSR, 1948-78**



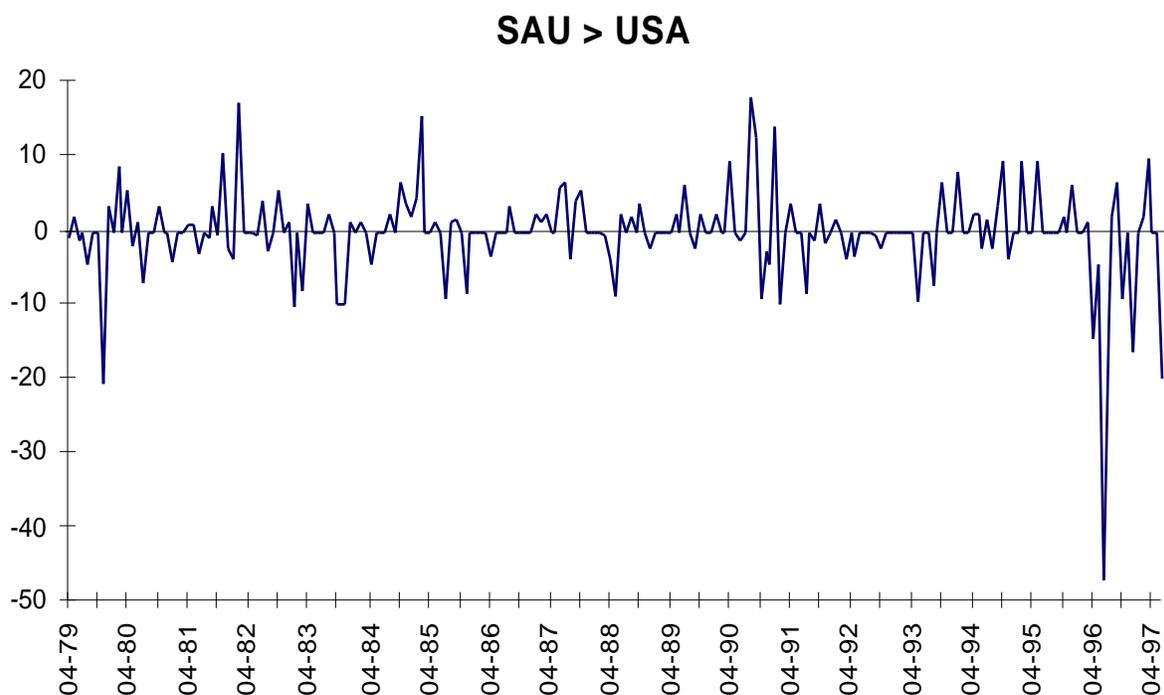
**Figure 1.3. Goldstein-scaled series for Iran Iraq, 1979-97**

Figure 1.3 shows the events initiated by Iran and directed to Iraq, coded with the WEIS system and aggregated using Goldstein's (1992) scale.<sup>1</sup> The major offensives of the Iran-Iraq war are clearly visible, as are the negotiations and overtures that Iraq made to Iran following Iraq's invasion of Kuwait. The 1990s, as expected, are characterized by sporadic disagreements but only limited activity compared to the 1980s.

In contrast to the largely conflictual behavior in Figure 1.3, the graph of the Saudi Arabia to U.S.A. relationship in Figure 1.4 shows a sequence of ebbs and flows in the relations between the two states (also note that the vertical scale is about one-tenth that of Figure 1.3). An examination of the individual events shows that the positive values primarily correspond to meetings, agreements and statements of policy support. Negative values correspond to diplomatic disagreements between the two governments, for example on the controversy over the sale of AWACS aircraft in the early 1980s, and over U.S. policy towards Iran and Israel later in that decade. The very large negative spike in 1996 is due to the Khobar Towers bombing in 1996. This data set did not differentiate sub-state actors, and therefore the bombing was coded as a Saudi attack on the U.S.; additional negative events were generated by the diplomatic disagreements between the two governments over the investigation of that incident. The only large anomalous negative value occurs in January and February 1991, where there are a number of incorrectly-coded uses of force. These come from reports about the Second Gulf War where the machine coding system failed to accurately determine the appropriate actors.<sup>2</sup>

<sup>1</sup> The data in Figures 1.3 and 1.4 were coded by the authors from Reuters lead sentences using KEDS. The Iraq to Iran sequence is similar: the two correlate with  $r = 0.84$

<sup>2</sup> As we will discuss later, the KEDS coding dictionaries, which were developed to code general international behavior, were prone to errors when coding stories dealing with military activities in the Second Gulf War. Descriptions of these events involved vocabulary and sentence constructions not encountered in other contexts, and sentences were sometimes misinterpreted.



**Figure 1.4. Goldstein series for Saudi Arabia U.S.A., 1979-97**

As these figures illustrate, event data can be used to summarize the overall relationship between two countries over time. The patterns shown by event data generally correspond to the narrative summaries of the interactions found in historical sources. But unlike narrative accounts, event data can be subjected to statistical analysis and other computerized analytical techniques.

Event data analysis relies on a large number of events to produce meaningful patterns of interaction. The information provided by any single event is very limited, and single events are sometimes affected by erroneous reports and coding errors. However, important events trigger other interactions throughout the system. For example while Iraq's invasion of Kuwait by itself generates only a single event with WEIS code 223 ("military force") the invasion triggers an avalanche of additional activities throughout the international system as states and international organizations denounce, approve or comment on the invasion, so the crisis is very prominent in the event record.

## 1.2. The Development of Event Data Analysis in International Relations Research

The historical development of event data concepts and collections dates back to nearly the beginning of "scientific" studies of international politics. The early theoretical development of WEIS is thoroughly discussed in a series of papers by McClelland (1967a, 1967b, 1968a, 1968b, 1969, 1970)<sup>3</sup>; the foundations of the COPDAB project are fairly well documented in a series of papers coming out of Azar's Michigan State University event data conferences during 1969-71 (Azar, Brody and McClelland 1972; Azar and Ben-Dak 1975; Azar et al. 1972; Azar and Sloan 1975).

Most of these early efforts were motivated by attempts to develop statistical early warning indicators of international and domestic instability. The Department of State experimented with coding event data for a small set of states in 1971 in its Foreign Relations Indicator Project (see Lanphier 1975). The Pentagon's Defense Advanced Research Project Agency (DARPA) sponsored a large-scale project in the 1970s to develop event data models for crisis forecasting and management. In the early years of the Reagan administration, the National Security Council staff in the White House undertook a major event data collection and analysis effort. MIT's case-oriented CASCON project was also receiving considerable attention in the policy community during this period, including the U.S. Department of Defense, State Department and Arms Control and Disarmament Agency, and the United Nations (Bloomfield and Moulton 1998: chapter 8).

These efforts apparently had little long-term impact on the formulation of foreign policy, although the data produced continued to be used in academic research. Laurance (1990) analyzes the reasons for the limited impact of event data on policy, including the failure to coordinate the event data projects with the analysts and policy-makers who were supposed to use the data, the absence of user-friendly analytical tools, and the absence of guidelines on how event data could

<sup>3</sup> We are indebted to Harold Guetzkow for an extensive collection of the early WEIS memoranda.

be used with traditional, non-statistical sources of information. Because of these problems, new global event data collection efforts ceased in the 1980s, although the COPDAB and WEIS data continued to be refined, and data sets such as CREON were used in academic research. A small number of new data sets focusing on international crises—notably Leng's BCOW (Leng 1987) and Sherman's SHERFACS (Sherman and Neack 1993)—were developed during this time.

Large-scale event data efforts were revived in the early 1990s in the second phase of the National Science Foundation's "Data Development in International Relations" project (DDIR), directed by Dina Zinnes and Richard Merritt (Merritt, Muncaster, and Zinnes 1993). Rather than simply extending the work of the 1970s, DDIR emphasized the development of new approaches, with particular emphasis on exploiting the computing power available in personal computers and using machine-readable news sources. The Global Event Data System (GEDS) at the University of Maryland grew out of this project, as did the early work on KEDS and several other experimental projects. The DDIR project marked a transition between the DARPA-style event data research and contemporary approaches, and the various articles in Merritt, Zinnes and Muncaster (1993) show a mix of old and new techniques. These methods slowly diffused into the policy community, and by the end of the decade, event data were employed in the development of experimental early warning systems at the U.S. Department of Defense, in the dynamic modeling phase of the State Failures Project, and in Switzerland by the FAST project (Krummenacher and Schmeidl 2000).<sup>4</sup>

By the late 1990s, three general changes were apparent in event data analysis. First, there was a move away from newspaper sources to wire-service sources, particularly Reuters, and

<sup>4</sup> We are currently aware of event data projects using the VRA automated coder—a commercial spin-off from the PANDA project—at the Joint Warfare Analysis Center (JWAC) and CINCPAC at the U.S. Department of Defense, UNICEF in the United Nations, and FAST in Switzerland. KEDS-based data are being used at JWAC and a U.S. government project on forecasting genocide/politicide; GEDS-based data on “accelerators” was used in the U.S. State Failures II project (Esty et al 1998). All of these projects have focused on the forecasting of state breakdowns and humanitarian crises. It is quite possible that additional projects exist that we would not know about, particularly if they used software from the MUC experiments (Lehnert and Sundheim 1991) rather than software derived from DDIR projects.

projects moved from relying on paper to using electronic databases, notably NEXIS.<sup>5</sup> The electronic systems were more convenient than the paper sources; for example the NEXIS system provided an elaborate Boolean search facility that allowed a researcher to narrowly focus a request for information. Wire service reports were more comprehensive than reports from newspapers, particularly as newspapers such as the *New York Times* began to place greater emphasis on soft news such as lifestyle and popular culture, at the expense of their international coverage.

Second, machine-coding replaced human coding for almost all applications. There were some initial experiments in the early 1990s to determine whether machine-*assisted* coding systems—which employed user-friendly software to simplify the clerical aspects of event coding—could produce dramatic improvements in efficiency compared to the paper-and-pen systems of the 1970s. This did not prove to be the case, and the computer-assisted human coding projects experienced the same inefficiencies caused by coder turnover, boredom, and training costs that had plagued earlier efforts. In the meantime, dramatic reductions in the cost and increases in the capacity of personal computers gave an overwhelming economic advantage to the machine coding projects. As the decade progressed, other advantages to machine coding became apparent—for example its transparency, stability and flexibility—and consequently machine coding became the preferred approach except in a few specialized applications.

Third, the focus of most event data collections shifted from the global approach to regional approaches. New data sets tended to focus on a specific geographical area, and decisions were made to code specific sub-state actors based on the political circumstances. The end of the Cold War also shifted the focus of most event data efforts from superpower conflict (Ashley 1980; Goldstein and Freeman 1990; Ward and Rajmaira 1992) to regional and even substate conflicts (Huxtable and Pevehouse 1996; Kinsella 1995, 1998; Reuveny and Kang 1996a, 1996b; Bond et al. 1997; Goldstein and Pevehouse 1997; Schrodts and Gerner 1994, 1997; Schrodts 1999, 2000).

<sup>5</sup> Prior to 10 June 1997, the NEXIS service, which was available to academic institutions at a relatively inexpensive rate, contained daily updates from various Reuters services.

While the COPDAB coding system was maintained and expanded by the GEDS project (<http://www.bsos.umd.edu/cidcm/geds/>), WEIS continues to be the most widely employed system, albeit it has often been extended to provide greater detail in the coding of domestic conflict events. The most notable of these WEIS extensions was done by the Protocol for the Analysis of Nonviolent Direct Action (PANDA) project at Harvard in the mid-1990s (<http://data.fas.harvard.edu/cfia/pnscs/panda.htm>), which produced a global, Reuters-based event data set covering 1984 through early 1995. To accommodate domestic events, PANDA more than doubled the number of WEIS categories, while providing a systematic table for translating PANDA codes to WEIS codes. More recently, this effort has been extended to the IDEA coding system (Taylor, Jenkins, and Bond 1999), which is designed to be used in the next edition of the *World Handbook*.

### 1.3. Event Data Sets

Event data sets fall into two general categories. *Actor-oriented* data sets record all interactions among a set of actors for a specific period of time, for example the Middle East 1979-99. *Episode-oriented* sets look only at the events involved in a specific historical incident, usually an international crisis or use of force.

The objective of most academic event data research is to find theoretically-informed statistical regularities, so event coding systems are closely linked to a theory or set of theories about international behavior. The detailed reports of the event data collection efforts (for example Azar, Brody and McClelland 1972; Azar and Ben-Dak 1975; Burgess and Lawton 1972; Hermann et al 1973; Merritt, Muncaster and Zinnes, 1993)—although not necessarily the codebooks for those data—show a deep awareness of the linkage between theory, coding and data collection.

The WEIS and COPDAB schemes, for example, were constructed in the milieu of the international relations theory of "realism" that placed primary emphasis on diplomatic and military behavior. In contrast, the CREON data set focuses on the elements of foreign policy behavior identified by the theories developed in James Rosenau's "Inter-University Comparative

Foreign Policy Project” (see Hermann et al 1973:8-15). While both of these data sets provide good indicators of conflict behavior, neither are particularly useful in studying contemporary international economic or environmental issues. Similarly, when the PANDA project sought to study “non-violent direct action” in domestic conflict, they found that they needed to make substantial extensions to the WEIS coding scheme.

### 1.3.1. Actor-Oriented Data Sets

#### WEIS

The WEIS coding scheme classifies events into 63 specific categories; these are organized into 22 general categories such as "Consult", "Reward", "Protest" and "Force" (see Table 1.3). The general categories form a very rough cooperation-conflict continuum. WEIS coding was the *de facto* standard used by the U.S. government-sponsored projects during the 1970s, and consequently a number of the data sets available in the ICPSR archive use the WEIS scheme.

The WEIS data set available at the ICPSR covers only eleven years (1966-77) and contains only about 90,000 events; the source text is *The New York Times*. Data after 1977 have continued to be coded by McClelland and several of his students—most recently Rodney Tomlinson at the US Naval Academy (Tomlinson 1993)—but the full series is not available in the public domain at the present time.

Because most common statistical routines, such as regression analysis, use numerical rather than categorical data, WEIS events are often aggregated into numerical scores before being analyzed. Vincent (1979) and Goldstein (1992) provide two such scales that assign numbers on a cooperation-conflict continuum to each WEIS category; Table 1.3 shows examples of Goldstein scores for several WEIS categories. As we will demonstrate in Chapter 3, widely-varying scales can produce similar analytical results in many problems, because much of the variance in event data is due simply to the presence or absence of an event. WEIS codes can also be translated into the COPDAB scale, although one cannot translate from COPDAB to WEIS because COPDAB makes fewer distinctions in the type of event.

Table 1.3. Examples Of WEIS Event Codes

WEIS Category	Goldstein value
<b>11. REJECT</b>	<b>-4.0</b>
111 Turn down proposal; reject protest, demand or threat	-4.0
112 Refuse; oppose; refuse to allow	-4.0
<b>12. ACCUSE</b>	<b>-2.0</b>
121 Charge, criticize, blame, disapprove	-2.2
122 Denounce, denigrate, abuse	-3.4
<b>13. PROTEST</b>	<b>-2.0</b>
131 Make complaint (not formal)	-1.9
132 Make formal complaint or protest	-2.4
<b>17. THREATEN</b>	<b>-6.0</b>
171 Threat without specific negative sanctions	-4.4
172 Threat with specific nonmilitary negative sanctions	-5.8
173 Threat with force specified	-7.0
174 Ultimatum: threat with negative sanctions and time limit specified	-6.9
<b>18. DEMONSTRATE</b>	<b>-6.0</b>
181 Non-military demonstration; walk out on	-5.2
182 Armed force mobilization, exercise and/or display	-7.9

## COPDAB

The COPDAB data set is substantially larger in size and scope than WEIS, with about 350,000 international events for the period 1948-78. COPDAB uses a number of different news sources rather than depending solely on *The New York Times*; in particular it uses a variety of regional sources to cover events outside of North America and Europe.<sup>6</sup> In contrast to the categories in WEIS, COPDAB uses an ordinal coding scheme that goes from 1 to 16 (see Table 4) supplemented by a numerical cooperation-conflict intensity scale developed by Azar and Sloan (1975). COPDAB coding also classifies an event into one of eight types—for example symbolic,

<sup>6</sup> Because WEIS and COPDAB are based on different sources, they do not have a high degree of overlap: *International Studies Quarterly* (1983) contains two analyses of this problem along with a commentary by McClelland; Reuveny and Kang (1996b) deal with the issue of “splicing” the two data sets.

political, military, economic or cultural. The GEDS project has been augmenting the original COPDAB data with more recent data on selected dyads during the 1990s, using machine-assisted coding methods.

---

Table 1.4. Examples Of COPDAB Event Codes

---

- |           |   |
|-----------|---|
| <b>09</b> | Nation A expressed mild disaffection toward B's policies, objectives, goals, behaviors with A's government objection to these protestations; A's communiqué or note dissatisfied with B's policies in third party. [scaled vale: 6]   |
| <b>10</b> | Nation A engages in verbal threats, warning, demands and accusations against B; verbal, hostile behavior. [scaled vale: 16]   |
| <b>11</b> | Nation A increases its military capabilities and politico-economic resources to counter Nation B's actions or the latter's contemplated actions; A places sanctions on B or hinders B's movement in waterways or on land and attempts to cause economic problems for B. [scaled vale: 29] |
- 

## CREON

The Comparative Research on the Events of Nations data set (Hermann et al 1977; East, Salmore and Hermann 1978) is designed for the study of foreign policy interactions. Its basic event coding scheme is similar to that of WEIS, but CREON also codes over 150 variables dealing with the context of the event, related actions, and internal decision-making processes. Unlike WEIS and COPDAB, CREON does not code all interactions during a period of time: instead it covers a random sample of time periods during 1959-68 and a stratified sample of 36 nation-states that contains a disproportionate number of developed and English-speaking countries. The purpose of CREON is to study the foreign policy *process*, rather than foreign policy output. In practice this means that CREON is better suited than WEIS or COPDAB to studying the linkages between the foreign policy decision-making environment and foreign-policy outputs for specific decisions, but it cannot be used to study policy outputs over a continuous period of time or for countries not in the sample. CREON is no longer maintained, but is available through the ICPSR.

## PANDA

The most current global event data set comes from the PANDA project at the Program for Nonviolent Sanctions and Cultural Survival at the Center for International Affairs at Harvard (Bond, Bennett and Vogele, 1994). PANDA data set was automatically coded from Reuters lead sentences using the KEDS computer program and employs a superset of the WEIS coding scheme that provides greater detail in internal political events. It contains about 500,000 events covering the entire world for the period 1984 to early 1995; the data are available on the Web at <http://data.fas.harvard.edu/cfia/pnscs/DOCS/datafiles.htm>.

Table 1.5. PANDA versus WEIS codes for the WEIS "Reject" category

WEIS		PANDA	
111	Turn down proposal; reject protest, demand or threat	411	turn down proposal, reject non-judicial appeal
112	refuse; oppose; refuse to allow	412	refuse to allow; ban or prohibit action
		413	defy law, customs or norms
		414	flee, protest emigration
		415	ignore, isolate, ostracize, disclose information, outing
		421	voice opposition, disagreement

## Other Actor-Oriented Event Data Sets

While WEIS, COPDAB, CREON and PANDA are the largest actor-oriented data sets, a variety of smaller sets exist. The ICPSR has several regionally-specific, WEIS-coded data sets dating from the 1970s, and additional regional data sets are being collected at the present time. The South Africa Event Data set (SAFED; van Wyk and Radloff 1993) is a WEIS-coded collection focusing on southern Africa for the period 1977-88; it has unusually dense coverage of non-state actors such as guerrilla movements. Ashley (1980) assembled a data set focusing only on the interactions of the superpowers—the USA, USSR and PRC—for 1950-72; this contains about 15,000 events and is coded with a COPDAB-like scale.

### 1.3.2. Episode-Oriented Data Sets

#### BCOW

The Behavioral Correlates of War data set (Leng 1987) codes a sample of 45 major international crises over the period 1816-1979; roughly half of these crises culminated in war and the other half were resolved without war. Most of the crises are in the 20th century; about a third are post-WWII; and many of the crises preceding WWI and WWII are included in the sample. BCOW's event codes are an expanded version of the WEIS scheme containing about 100 categories and differentiating more clearly between verbal, economic and military behavior. Leng (1993b) contains an extensive analysis of this data set.

BCOW uses multiple sources of information, including newspaper accounts, diplomatic histories, and chronologies (Leng 1987:1). The number of events in each crisis range from 120 events in the 1889-90 British-Portugal crisis in southern Africa to 2352 events in the 1956 Suez crisis. The ICPSR data set is accompanied by a very extensive coding manual that would allow a researcher to code additional crises in a manner consistent with the original data; it also includes some specialized software that can be used to analyze the data.

---

Table 1.6. Examples Of BCOW Event Codes

---

#### **Military Actions (sample from a total of 36 categories)**

11212	International Peacekeeping Force
11333	Alert
21143	Change in Combat Force Level
31133	Fortify Occupied Territory

#### **Diplomatic Actions (sample from a total of 35 categories)**

12121	Negotiate
12362	Declare Neutrality
12213	Punish of Restrict Foreign Nationals
32151	Grant Independence to Colony

#### **Economic Actions (sample from a total of 20 categories)**

13121	Economic Negotiation
23121	Sell or Trade
23231	Pay for Goods or Services

#### **Unofficial Actions (sample from a total of 11 categories)**

14251	Proforeign Demonstration
14213	Antiforeign Demonstration
14152	Hostage Taking

---

## **CASCON**

The Computer-Aided System for the Analysis of Local Conflicts system (CASCON; Bloomfield and Moulton 1989, 1997) codes the characteristics of 85 internal and international conflicts during the post-World War II period. The analytical framework is based on a study by Bloomfield and Leiss (1969) and is organized around six predefined conflict phases ranging from the issues leading to the initiation of the dispute to the resolution of the dispute. CASCON codes 571 "factors" for each crisis; some of these describe specific types of events, others describe contextual characteristics of the crisis such as whether the parties to the conflict are dependent on outside aid.

The current version of CASCON is an integrated "decision support system" designed to help decision-makers and students compare current crises with the historical data on the crises; it accompanies by an undergraduate textbook on international conflict. The system runs on personal computers and an earlier version won a prestigious EDUCOM/NCRIPTAL award for excellence in educational software. The system contains the conflict data set, a variety of analytical tools that can be used to compare conflicts, and a subsystem for entering new cases into the database.

## **SHERFACS**

The SHERFACS data set (Sherman and Neack 1993) codes over 700 international disputes and almost 1,000 domestic disputes in the 1945-1984 period. It combines several different coding schemes, including COPDAB event codes, the CASCON crisis phase structure, and a variety of conflict management variables originally used in the Butterworth (1976) data set on crisis mediation. SHERFACS is particularly strong on coding non-state actors such as ethnic

groups, transnational actors such as intergovernmental organizations, and non-national actors such as multinational corporations.

An early version of SHERFACS is available from the ICPSR (Alker and Sherman 1982, 1986). A much more extensive version of the data set was planned and partially implemented, but this project was cut short by Frank Sherman's untimely death in 1996. Nonetheless, much of this information is available at a site maintained by Hayward Alker—

<http://www.usc.edu/dept/ancntr/Paris-in-LA/Database/sherfacs.html>—and is in the public domain.

### **Other Episode-Oriented Event Data Sets**

Several other data collections available from the ICPSR such as *The World Handbook* (Taylor and Hudson 1972) contain some limited amounts of event data. Another example is the PRINCE Project data set (Coplin, O'Leary, and Shapiro, n.d.). This data set was originally collected in conjunction with a computer simulation project and contains a small set of event data dealing with political issue positions for the period 1 January 1972 to 30 June 1972. Other data sets have been collected for the study of a specific crisis. For example Lebovic (1993) coded events during the period prior to the 1991 Gulf War (2 August 1990 to 16 January 1991 in order to analyze the impact of foreign policy "momentum" in that crisis. The International Political Interactions project (Moore and Davis 1998;

<http://garnet.acns.fsu.edu/~whmoore/ipi/ipi.html>) provides conflict event information on selected dyads for 1979-92. The International Crisis Behavior data set (Brecher and Wilkenfeld 1997; <http://www.colorado.edu/IBS/GAD/spacetime/data/ICB.html>), while not strictly event data, is another example of a data set focusing on the characteristics of conflict episodes from 1918 to 1994;

## **1.4. Sources of news reports**

One of the perennial problems in event data analysis has been the choice of which news reports to code. Because most event data sets have been coded in the United States, they have

tended to use reports written in English. Because human coding is very labor-intensive, those projects have also tended to favor news sources that are readily available and indexed. For example, WEIS was coded from the *New York Times*; CREON from *Deadline Data on World Affairs*, which abstracts 46 international sources (Hermann et al 1973:18).

There has always been concern that news sources could have substantial regional biases. An early study by Doran, Pendley, and Antunes (1973) showed dramatic differences between *New York Times* coverage of violence in Central America and the levels of violence reported by Central American sources. Because of this, some data sets have used regional news sources. For example, the BCOW codebook (Leng, 1987) lists dozens of periodical and historical sources and Azar (1980: 146) states that COPDAB is based on "events reported in over 70 sources."

During the 1990s, most event data projects (whether human or machine coded) shifted to the Reuters newswire source. At present, Reuters is archived from 15 April 1979 to 10 June 1997 in the NEXIS data service, and material subsequent to that date can be obtained from the Reuters Business Briefing service. Reuters issues about a thousand stories per day, and therefore provides far greater coverage than the *Times* or any other single regional source. While Reuters has some regional biases—it devotes greater coverage to areas that are of interest to institutions that can afford its services but which are not covered by other media—the fact that it is global rather than based in a single city or country reduces these substantially.<sup>7</sup>

Reuters and other news services such as *Agence France Presse* (AFP) and the BBC World Service have been criticized as providing a "white European male" perspective. Based on our unsystematic encounters with newswire personnel in the Levant, we find this criticism itself to be a white male perspective: Most of Reuters correspondents ("stringers") are local to the areas they are reporting from, and quite a few of them (as well as much of the Reuters editorial staff)

<sup>7</sup> The status of the availability of Reuters has been in flux over the past three or four years. For a period of time after dropping NEXIS, Reuters was only available directly from Reuters Business Briefing, and only for a ten year period. In mid-1999, it joined with Dow Jones Interactive to create a service that initially promised to provide the full Reuters archives, possibly back as far as the early 1970s. By early 2000, however, that combined service had yet to materialize, though the Web-based Dow Jones service has itself emerged as an attractive alternative to both NEXIS and Reuters. Given the current wave of media mergers, acquisitions, and spin-offs, anything might happen in the future.

are not male. Most stringers have had a disproportionate exposure to Western education and norms—they are elites, not peasants—but they typically grew up with the local culture and language.

The distinction here is important: the global news agencies are *based* in the capitals of the hegemonic powers, but the reporting itself (and some of the editing) comes from a very decentralized network of observers who are usually native to the areas from which they are reporting. In contrast, the hegemonic "papers of record," such as *The New York Time* rely much more heavily on their own nationals to provide coverage. At best these individuals experience a long learning curve, as the autobiographical accounts of correspondents such as Thomas Friedman and Harrison Salisbury attest; at worst they are Graham Greene wannabees without a long-term commitment to a region whose understanding of the local politics has invariably been filtered through the alcoholic haze of too many expense-account dinners at five-star hotels. The differences in the quality of reporting can be profound.)

#### 1.4.1. Regional versus Global Sources

Data services such as NEXIS and Dow Jones Interactive carry the text of literally hundreds of regional news sources. As these sources became easily available, a number of researchers (including ourselves) thought that the regional sources could be used to fill in the gaps of the reporting of global news sources. provide details as necessary. We assumed that services such as Reuters and AFP would contain a selective *subset* of the events reported in the regional sources, and anything really important in a regional source also would be reported by the global source.

Alas, it isn't so, at least for Reuters. Reuters and the regional sources are supplementary, rather than complementary. A number of different studies in different regions of the world have shown that Reuters reports events that are not reported in the regional sources (and vice versa).

These studies include

Europe	Gerner et al 1994, Huxtable and Pevehouse 1997
Middle East	Gerner et al 1994
Africa	Huxtable and Pevehouse 1997, Moore 1997

## Southeast Asia      Howell and Barnes 1993

The same is true when global sources are compared: For example in his dissertation research on West Africa, Huxtable (1997) assumed that the English-language sources Reuters and BBC would focus on Anglophone states such as Nigeria, Ghana and Sierra Leone, while the French-language AFP would focus on Francophone states such as Senegal, Niger and Côte d'Ivoire. This did not prove to be the case: Reuters would sometimes pick up major events in Francophone states that were missed by AFP, and AFP would sometimes provide better coverage of Anglophone states. In some cases, it was almost possible to reconstruct the travel itineraries of individual Reuters, BBC and AFP reporters as they worked their way through West African capitals, producing a flurry of temporary detail on areas that would receive no coverage for another year.

The fact that event data is coded from a finite number of sources has been criticized by Alker (1988), among others, as privileging some interpretations of history over others. While this may be true, it is no more or less the case than the situation facing traditional studies of political behavior. Short of descending into a post-modernist quagmire where nothing can be assumed, concluded or explained, any political analysis must assume that certain events, conditions, motivations and coalitions occurred, and others did not. The traditional method of composing accounts of political activities using a variety of documentary and autobiographical sources is one way of doing this; the processes by which text is selected for event coding is another. Each method is subject to selection bias and varying interpretation.

In retrospect, the event data community under-estimated the sheer volume of "events" that occur in the world. At 1000 stories per day, Reuters initially looked like a major improvement over the ever-diminishing international coverage of the *New York Times*. But a bit of reflection will show that even 1000 stories is only a tiny fraction of all of the political events that occur in the world on any given day. Any news source is going to nonrandomly sample only a small number of these events, so the question is whether that sample is useful for a specific analytical task.

## 1.5. Example: KEDS Data for the Arab-Israeli Conflict

This section discusses our major validity test of the KEDS system, published as Schrodts and Gerner (1994). It is not the only validity test we have done but it is the most systematic. The test focuses on the Levant for the period 1982-92, and compares the KEDS-generated data with the general textual record of "events on the ground." (In Chapter 2, we will also compare this machine-coded data with the closest comparable human-coded data set, Tomlinson's (1993) extension of the original WEIS set.)

To create this data set, we downloaded the first sentences ("leads") from Reuters News Service stories available from the NEXIS data service. NEXIS is searched using keywords that can be arranged into Boolean statements. To create this data set, the search command was

HEADLINE (ISRAEL! OR JORDAN! OR EGYPT! OR LEBAN! OR SYRIA! OR PLO OR PALEST!).

The "!" is a wild card character that matches any word beginning with the preceding letters; "PALEST!" picks up "Palestinian," "Palestinians," and "Palestine."<sup>8</sup> We examined only those dyads in which both the source and the target of the event were among the seven actors of interest. A total of 23,127 events are included in this eleven year data set; daily events reports were aggregated to the monthly level prior to analysis. Israeli actions toward Palestinians, Israeli actions toward Lebanon, and Palestinian actions toward Israel account for the greatest number of events. The fewest events are recorded for Lebanese actions toward Jordan, Lebanese actions toward Egypt, and Jordanian actions toward Lebanon.<sup>9</sup>

In order to obtain a general sense of these regional interactions, we began by examining the conflict-cooperation patterns of each directed dyad during the eleven year period. (A directed

<sup>8</sup> In later phases of the KEDS project, we further refined this search statement to eliminate sports stories, historical chronologies, monthly news summaries and other information that we did not want to code.

<sup>9</sup> The small number of reported events involving Lebanon can probably be accounted for by two factors. First, the situation in Lebanon was tremendously unstable during the 1982-92 period. Therefore, it is likely that few routine interactions occurred; instead the data consist largely of military events and diplomatic discussions about Lebanon. Second, the instability probably decreased *reporting* of "ordinary" events: the few media representatives in the country focused their attention on the domestic and international conflicts.

dyad refers to the actions of X toward Y and is represented as X → Y.) We were particularly interested in reciprocal directed dyads (X's actions toward Y and Y's actions toward X) as well as any directed dyad with an usually high or usually low net cooperation score. Net cooperation was calculated by weighting each WEIS according to the Goldstein (1992) scale and totaling the events for each month. This produces a single numerical score on a conflict-cooperation dimension similar to that used in COPDAB; negative scores indicate conflict and positive scores indicate cooperation..

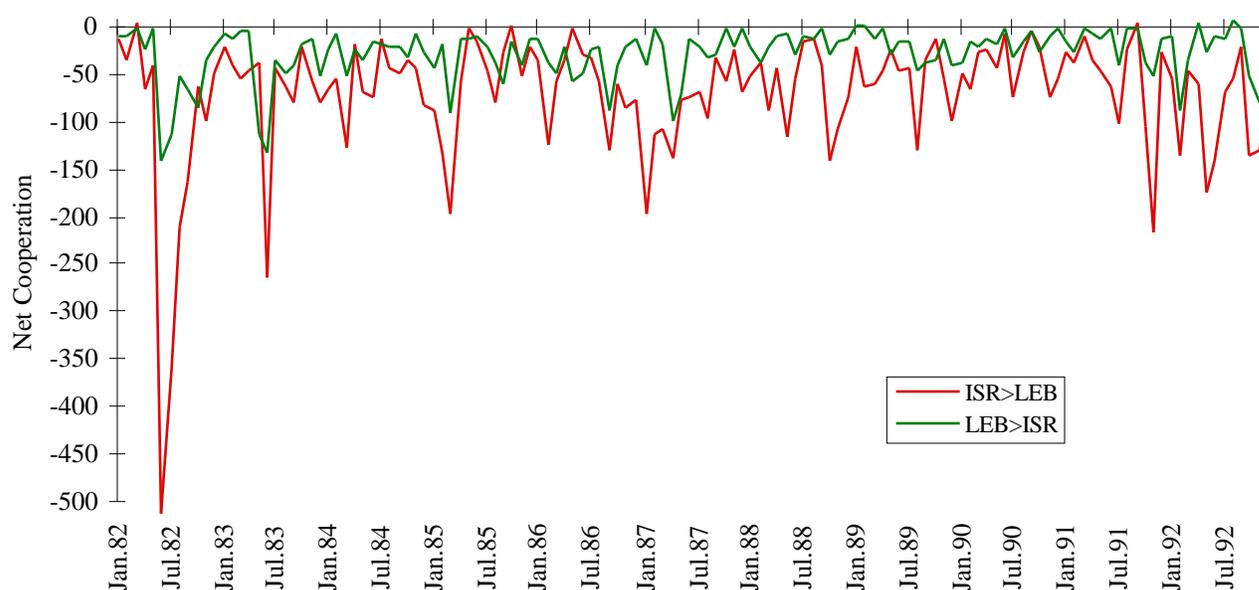
The average conflict or cooperation scores shown by the KEDS data across the entire period are consistent with narrative accounts describing the specific relationships. Between 1982 and 1992, most analysts describe interactions among the United States, Israel, and Egypt as relatively harmonious, albeit with some tensions. In contrast, interactions among Israel, Lebanon, Syria, and the Palestinians were quite strained. These differences show up clearly in the net cooperation measures. For example, Israel → Palestinians has the lowest average net cooperation score of any directed dyad over the eleven year period. Israel → Palestinians are roughly twice as conflictual as those of any other directed dyad examined. Other highly conflictual directed dyads (in order from most conflictual) include Israel → Lebanon, Palestinians → Israel, Lebanon → Israel, Palestinians → Lebanon, Lebanon → Palestinians, Israel → Syria and Syria → Lebanon.

Fifty-five percent of the directed dyads have a mean positive net cooperation score. The variation in the extent of cooperative actions among these 23 directed dyads is not great: most have net cooperation scores that are only slightly positive. The most cooperative directed dyad is USA → Israel. Other relatively cooperative directed dyads include Israel → USA, USA → Egypt, Israel → Egypt, USA → Jordan, and Egypt → USA.

We also looked at monthly net cooperation scores for key reciprocal directed dyads to determine whether the data were accurately reflecting major events in the region such as the 1982 Israeli invasion of Lebanon, the Syrian military presence in Lebanon, the onset and evolution of the *intifada*, and the regional peace talks that began in Madrid in October 1991. We briefly summarize three dyads here; additional cases are discussed in Gerner (1993).

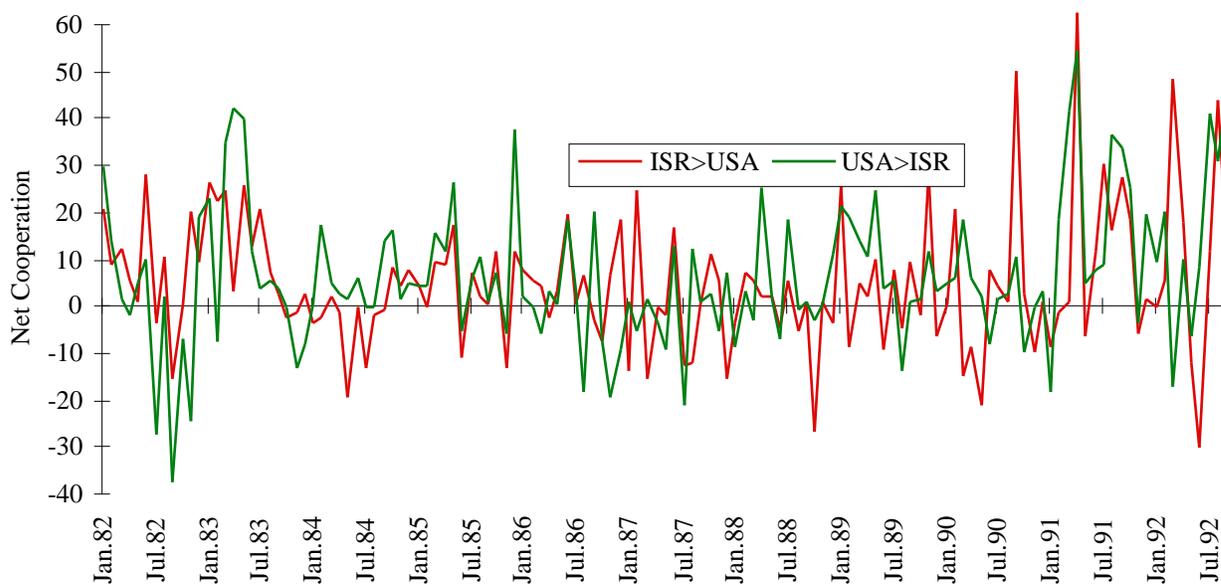
### 1.5.1. Face Validity

As Figure 1.5 illustrates, Israeli-Lebanese relations have been characterized almost entirely by hostility throughout the eleven-year period. The most severe conflict occurred during the 1982 war in Lebanon that began in June and continued through mid-August; this raised the Israel-Lebanon hostility score by a factor of two. Conflict peaks again in mid-1983, before Israel had withdrawn its troops from the Bekaa Valley or the Chouf Hills in southern Lebanon; in early 1985, after Israel had decided to pull out of most of southern Lebanon but before the evacuation had been completed; at the beginning of 1987, as Palestine Liberation Organization (PLO) fighters moved back into Lebanon; and late in 1991, when Israel launched a series of artillery attacks against Iranian-backed Hizballah forces and Lebanese villages. There is no overall improvement in the relationship in the fifteen months following the Madrid talks; this is consistent with the increased Israeli-Lebanese military hostilities that accompanied the early negotiations.

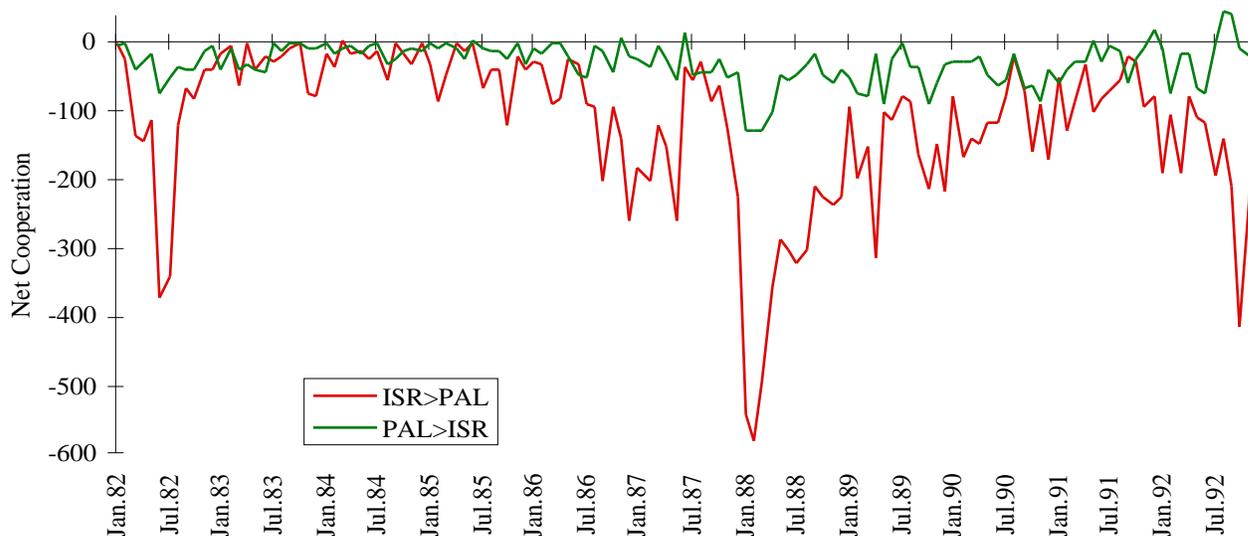


**Figure 1.5: Goldstein series for Israel–Lebanon, 1982-1992**

United States actions toward Israel show the greatest amount of cooperative behavior of any directed dyad in the data set (see Figure 1.6). This can be attributed in part to the consistently high level of consultation between the two countries. Thirty-five percent of reported USA-Israel events and 36 percent of Israel-USA events fall into the WEIS "consult" category, with peaks in 1982-83 and again in 1991-92. The negative net cooperation of USA-Israel during 1982 is the result of an unusually large number of U.S. actions in response to Israel's invasion of Lebanon that fall into the WEIS "accuse" and "reject" categories. In this same period, Israel initiated a number of cooperative consultations, which accounts for the generally positive Israel-USA pattern. During the *intifada* we see the same phenomenon: United States consultations with Israel help moderate the impact of U.S. accusations and rejections of Israeli proposals. Although Israeli-U.S. interactions occasionally move into the net conflict area, this pattern is far less strong than it would have been without the positive consultative activities.



**Figure 1.6. Goldstein series for Israel-United States, 1982-1992**



**Figure 1.7. Goldstein series for Israel-Palestinians, 1982-1992**

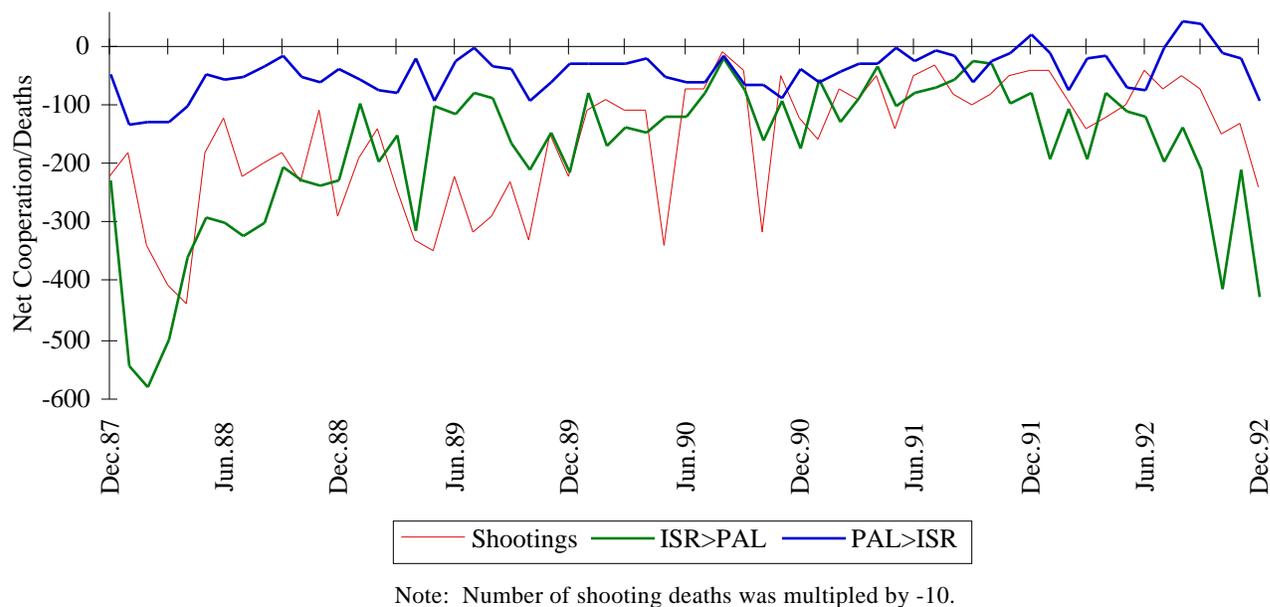
Finally, Israeli and Palestinian net cooperation over time is particularly interesting (see Figure 1.7). The net cooperative measure picks up a number of critical shifts in the overall attitude of each actor toward the other. For instance, the 1982 Israeli invasion of Lebanon is distinctly marked by a dramatic increase in conflictual Israeli actions toward Palestinians (both in Lebanon and in the West Bank and Gaza) preceding the invasion and the sharp drop-off in such events once Israeli troops had withdrawn from much of Lebanon later the same year. Net conflictual actions by Palestinians toward Israel are also higher than average during 1982, although the intensity is much less dramatic than that seen in Israeli actions toward the Palestinians. Israeli conflictual events begin to increase again in 1986, after Yitzhak Shamir replaced Shimon Peres as prime minister in the 1984 Labor-Likud Unity government. The data correctly show the dramatic increase in conflictual actions by both actors, particularly Israel, at the beginning of the *intifada*. The *intifada* continues to affect the pattern of Israeli and Palestinian actions throughout the next several years. The decrease in conflictual actions by each actor matches the decline in

the intensity of the *intifada*; the renewal of strongly conflictual Israeli actions toward Palestinians during the second half of 1992 is also recorded.

The asymmetry in the reported events from Israel to Palestinians versus Palestinians to Israel is probably partly due to the political asymmetry of the situation and partly due to difference in reporting. As the occupying power in the West Bank, Gaza and southern Lebanon, Israel was more likely to initiate action than the Palestinians. For example, during the 1982-92 period, Israeli police frequently broke up Palestinian demonstrations, but there were no instances of Palestinian police breaking up Israeli demonstrations (or, for that matter, Palestinian police). But in other instances, particularly those involving Israeli activity in Lebanon and during the *intifada*, some of the asymmetry may be the result of reporting styles. Many Reuters reports were based on Israeli sources, and consequently the lead sentence would tend to emphasize the Israeli action—"Israeli troops raided Palestinian guerilla bases..."—without necessarily reporting the reciprocal behavior that would occur in a military clash.

An additional check on the face validity of these event data is the relationship between the number of Palestinians shot and killed by Israeli soldiers and settlers during the *intifada*, as recorded by a source independent of Reuters, and net cooperative Palestinian and Israeli actions. While the record of fatalities is a more specific type of behavior than the aggregate net cooperation measure, one would expect the two series to covary. Figure 1.8 illustrates the relatively close relationship between these phenomena. Palestinian deaths by shooting are strongly and negatively correlated with Israeli net cooperative actions toward Palestinians ( $r = -0.51$ ; significant at the .01 level). In other words, months in which Israeli actions toward Palestinians are less cooperative according to the Goldstein-scaled measure tend to be the months in which a higher number of Palestinians are shot to death by Israeli forces. Shooting deaths are also negatively correlated ( $r = -0.47$ ; significant at the .01 level) with Palestinian net cooperative actions toward Israel.<sup>10</sup>

<sup>10</sup> The data on Palestinian shooting deaths by Israeli occupation forces comes from the Jerusalem-based Palestine Human Rights Information Center. Deaths that occur in the context of a military operation or when there is no clear human rights violation are excluded from these figures, as are deaths due to beating, tear gas inhalation, or



**Figure 1.8: Palestine-Israel Net Cooperation and Palestinians Shot to Death by Israeli Forces during the Intifada**

In short, this quick examination of some of the directed dyads indicates that the KEDS Arab-Israeli conflict data set accurately reports key events in the region during the eleven year period. There were no unpleasant surprises. Both cooperative and conflictual patterns show up where expected and there are no major unexplained clusters of events..

## 1.6. Comparison of Lead and Full Story Coding

Most of our KEDS data sets are coded using only the first sentence—the "lead"—of a newswire report. Following standard journalistic practice, the lead sentence usually summarizes the story that follows, and commonly has a relatively simple declarative structure, as illustrated

other non-bullet causes. The level of association between Palestinian deaths and the actions of the two directed dyads is actually *stronger* if beating and tear gas inhalation deaths are included. Gerner (1990, 1991) discusses these data in greater detail.

earlier in Table 1.1. This approach contrasts with earlier human-coded projects, which coded entire stories.

This emphasis on leads—which typically contain less than 10% of the text of the story—would appear to ignore a great deal of useful information, and this has been one of the common criticisms of the KEDS approach. In practice, however, very little seems to be lost, at least when Reuters is being coded.

In a typical Reuters report, the body of the story adds very few events beyond those in the lead, and coding the full story substantially complicates automated coding methods. The body of the story will often repeat the event reported in the lead several different times, provide background information on events that occurred earlier in time, and provide extensive direct quotations that are very difficult to code correctly. When a single Reuters story contains multiple codeable events—for example an outbreak of violence accompanied by condemnations and offers of mediation—these almost always generate multiple stories, each with an appropriate lead. In the Reuters text-stream, in fact, one is far more likely to encounter the problem of the same event being reported in multiple leads than the problem of a significant event not being reported in any lead.

These characteristics of Reuters are, however, more pronounced in an areas such as the Levant, which is reported very intensely. When our colleague Phillip Huxtable was working on a data set for West Africa (Huxtable 1997)—a region that does not receive intense coverage in the international press—he noted that Reuters would sometimes append a series of event reports to the end of a story that were unrelated, except in their regional focus, to the primary focus of the story. The underlying editorial model appeared to be "Hey, if you're sufficiently interested in this out-of-the-way region to read this far, you'll probably enjoy this other stuff as well." Huxtable concluded that it might be possible to increase the density of events by full-story coding.

Huxtable's observation led us to run some specific tests comparing lead-sentence and full-story coding on a data set dealing with the Persian Gulf region, which receives more sporadic

coverage than the Levant. After downloading the full stories from NEXIS, we first filtered the text to remove all sentences appeared to be direct quotes.<sup>11</sup> Lengthy direct quotations tend to be very difficult to code because spoken language is less systematic than journalistic language, and since the main point that the speaker is making is usually summarized by the Reuters text outside of the quotations, little is lost by removing quotes. The remaining sentences were coded using the same dictionaries and complexity filter used to code the lead sentences.

The full-story coding (and downloading) was substantially more time-consuming than lead-sentence coding. Coding the lead sentences required about 2-hours on an 80 Mhz Macintosh 7100; coding the full stories required a full 24-hours.<sup>12</sup> The full-story coding generated 264,421 events, as opposed to the 48,721 events generated from the leads.

Table 1.7 shows the correlation ( $r$ ) between the monthly Goldstein series generated from the lead-sentence and full-story coding for 30 directed dyads. With the exception of dyads involving the United Arab Emirates (UAE), the correlations of the two series are quite high, usually above 0.75. This suggests that in most statistical studies involving linear models, similar results will be obtained with either approach. More generally, lead-sentence coding is probably quite adequate for exploratory work, given the much greater investment of time required to download complete stories.

---

**Table 1.7. Correlation of Goldstein-score time series generated with leads and full stories**

---

Source	Target					
	IRN	IRQ	SAU	USA	KUW	UAE

<sup>11</sup> The filter skipped sentences containing a double-quote character (ASCII 34) that were either preceded by a period (.) or followed by a comma (,). This eliminated all correctly formatted sentences in Reuters that quote a speaker, while retaining sentences that contain short phrases placed inside quotes such as:

Palestinian diplomacy has ended Lebanon's bloody "camps war" but analysts say it is likely to prompt a confrontation between Israel and Amal.

<sup>12</sup> These times are *substantially* longer than those required on a contemporary personal computer, which would have a much higher clock speed.

IRN	---	.90	.73	.85	.76	.47
IRQ	.94	---	.98	.98	.93	.70
SAU	.77	.74	---	.64	.52*	.31
USA	.89	.96	.75	---	.88	.52
KUW	.68	.94	.43	.55	---	.35
UAE	.38	.38	.35	.65	.46	---

N=119

\* Excluding Feb-91, which contains a number of incorrect codes associated with the Second Gulf War. With Feb-91 included, the correlation is 0.17

The lower correlations associated with the UAE are consistent with Huxtable's observations about West Africa: states perceived as peripheral by the international news media are more likely to be discussed only in the body of a story and not in the lead. For example, the SAU UAE dyad contains 200 reported events in the full-story series, but only 6 (!) in the lead sentences. The contrast between full-story and lead-sentence coding in IRN UAE series is less dramatic, but still has 126 events in the full series versus 21 in the leads.

Table 1.8 shows the dyadic reciprocity correlation ( $X \ Y \times Y \ X$ ; see Dixon 1996 and Goldstein and Freeman 1990) for the lead and full-story sequences. These correlations show a very clear pattern, with the full story reciprocity being higher in all but two cases. Once again, the correlations for the minor actors are substantially lower than those for major actors; in some cases they are not even statistically significant. Some of the reciprocity in the full stories may be artificial because the full story is more likely to present "the other side" and thus generate a "reciprocal" event that would not have been present in the absence of the Reuters reporter. However, the full story is also more likely to present secondary events that occurred but which did not by themselves justify a separate story and thus a lead. Furthermore, the fact that these series are aggregated by month should reduce the likelihood that the observed reciprocity is simply an artifact of Reuters editorial guidelines.

---

**Table 1.8. Dyadic reciprocity in Goldstein-score monthly aggregations generated with leads and full stories**

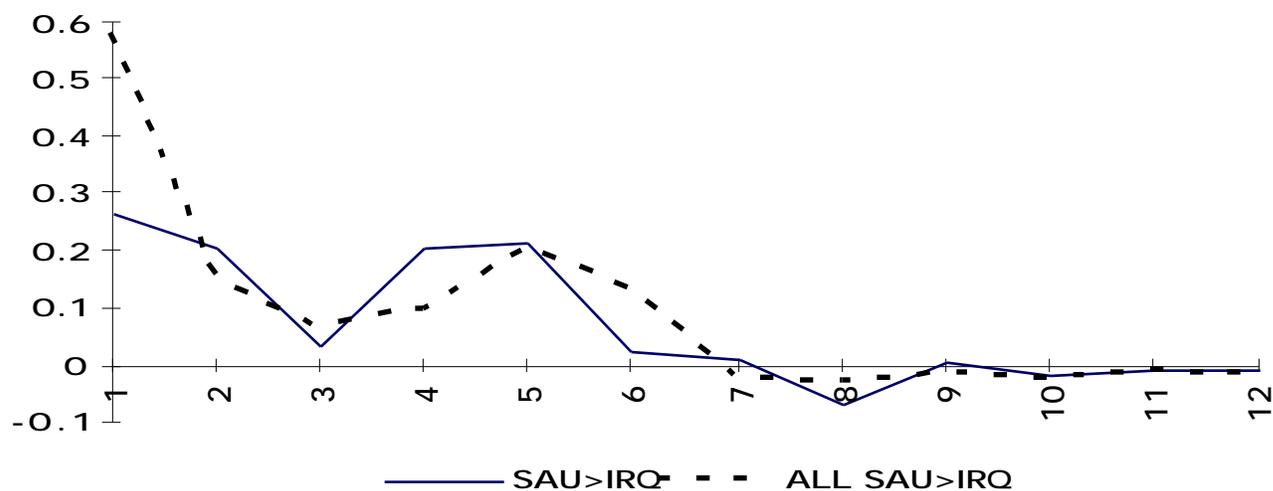
---

<u>Dyad</u>	<u>Full</u>	<u>Leads</u>
IRN-IRQ	.95	.84
IRN-SAU	.85	.61
IRN-USA	.80	.71
IRN-KUW	.71	.32
IRN-UAE	.25	.06
IRQ-SAU	.96	.78
IRQ-USA	.97	.92
IRQ-KUW	.84	.77
IRQ-UAE	.00	.02
SAU-USA	.73	.53
SAU-KUW	.63	.35
SAU-UAE	.76	.93
USA-KUW	.27	-.02
USA-UAE	.22	.27
KUW-UAE	.86	.64

---

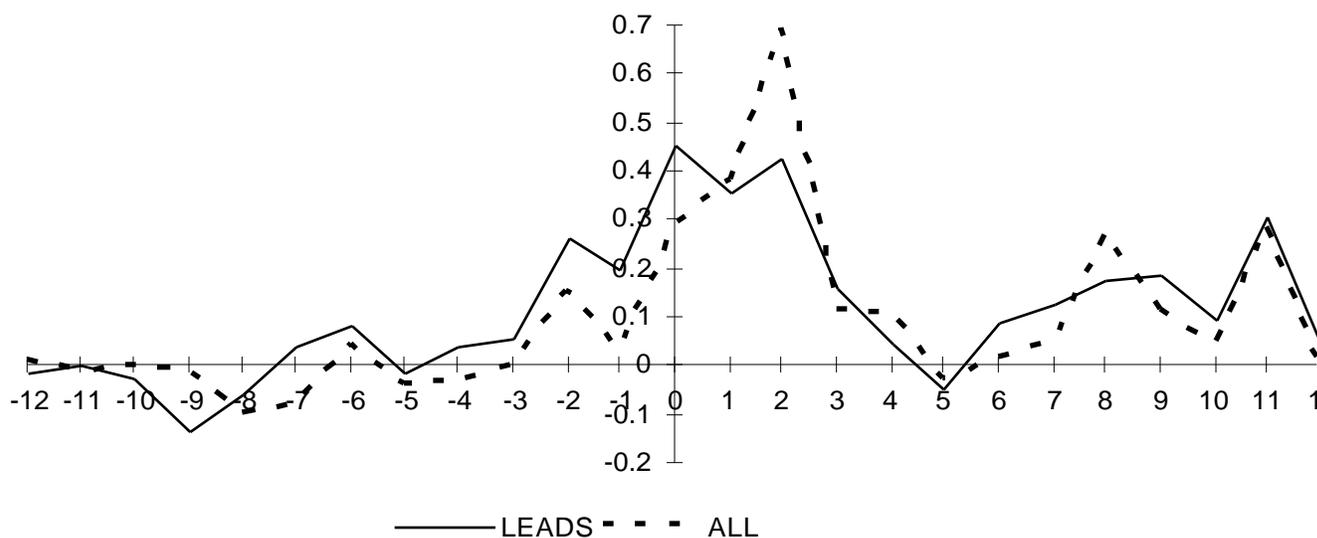
Finally, the similarities between the full-story and lead-sentence series can also be found in more complex measures of the series. Figures 1.9 and 1.10 give examples of the autocorrelation functions and cross-correlation functions of some of these series. The shapes of the curves are similar, although the strongest correlations are found in the full-story data.

### Autocorrelation for SAU ->IRQ



**Figure 1.9. Autocorrelation function of Goldstein series for Saudi Arabia - Iraq for full-story and lead-sentence events**

### Cross-correlation for IRN->SAU x USA->IRN



**Figure 1.10. Cross-correlation function of Goldstein series for Iran - Saudi Arabia with USA - Iran for full-story and lead-sentence events**

In summary, this analysis of the Gulf finds that for major actors there is generally a high correlation between the Goldstein series generated only with the lead sentences and those generated with full stories. Given that full-story coding involves substantially more downloading time, lead-sentence coding is probably sufficient for most exploratory work. However, full-story coding provides dramatically more information on minor actors, which is consistent with Huxtable's conclusions about West Africa.<sup>13</sup> Information is "out there" in the wire service reports, even if one has to dig for it, and minor actors are sometimes dramatically under-represented unless the entire story is coded.

Curiously, the marginal frequencies in human-coded data sets such as WEIS and COPDAB generally look more like the frequencies of lead-sentence coding than the frequencies of full-story coding, even if the coders were supposed to be working with the entire story. For example, COPDAB contains about 350,000 international events for the period 1948-78. Our full-story data set on the Gulf therefore records, on an annual basis, about the same number of events as COPDAB records *for the entire world*, despite the fact that COPDAB was generated by full-story coding from multiple regional sources.

In some early unpublished experiments that we did comparing KEDS automated coding to that of graduate student coders, we found that the human coders were more likely to miss secondary events reported in a story, as well as some of the combinations created by multiple-actor meetings. If the human coder, rather than meticulously coding every single sentence, relies (explicitly or cognitively) on a summary of the story, the secondary events involving minor actors will be missed.

<sup>13</sup> We also found this pattern in a data set we coded for Central Asia, 1989-99. The correlation between lead-story and full-story monthly scores aggregated using the Goldstein scale was very high for Afghanistan, which was covered continuously by the news media during the period. The correlations were substantially lower for minor actors such as Kyrgistan and Uzbekistan.

## 1.7. Conclusion

There are two fundamental reasons for using event data in political analysis. First, politics does not have the convenient numerical measures such as location, momentum, and temperature found in physics, or variables such as price, interest rates, and GNP found in economics. Political activity instead consists largely of discrete actions and communications directed from one actor to another over time. McClelland's (1970) original observations on the potential utility of event data as a method of addressing this problem still hold.

Second, human analysts have a limited ability to absorb vast quantities of largely redundant material. The text of NEXIS news wire leads covering only Israeli-Palestinian interactions for 1989 runs to some 300 pages. The full articles would fill perhaps 2000 pages; we suspect that few researchers would read all of these. The task becomes even more formidable if one is dealing with a long time series such as the Cold War: just what were the U.S.A. and U.S.S.R. doing on 16 August 1955? While most human analysts can memorize the day-to-day details of a short time period such as the Cuban Missile Crisis, or the major events of a long period such as the Cold War, we are skeptical about the human ability to memorize, much less analyze, day-to-day details for a long time period.

Event data fill that gap. The text of the journalistic sources provide memory and a variety of statistical and other computational methods can provide analysis. Between the text and analysis, one needs something similar in content to event data.

*Science* magazine once surveyed how new techniques in the physical and biological sciences sometimes revolutionized not just the methodologies, but also the theories, within their fields:

Not everybody appreciates the importance of technique. Many scientists, in fact, are "theory snobs" who dismiss technique as a kind of blue-collar suburb of science. . . . [But there is,] clearly, enormous transforming power in techniques. In the absence of an essential technique, a researcher or a field flounders, developing elegant theories that cannot be decisively accepted or rejected—no matter how many intriguing circumstantial observations are available. But with a key technique in hand, the individual and field move ahead at almost terrifying speed, finding the right conditions to test one hypothesis after another.

Conversely, new techniques often uncover new phenomena that demand new theories to explain them. (Hall 1992: 345)

The research in international relations, and much of comparative politics, is arguably theory rich and data poor. Too many theories are chasing too few facts, and for large sectors of those communities, “research tools” still consist of CNN, the *New York Times*, a copy of Thucydides and a snifter of brandy. At the same time, the interactions in international system are becoming more complex with the end of the Cold War and the need is greater than ever to be able to systematically study alternative theoretical explanations for that behavior.

One of the favorite parables employed by evangelical preachers is that of a sailing ship becalmed for weeks in the Atlantic, its crew slowly dying of thirst. Sighting a passing vessel,<sup>14</sup> the beleaguered crew appeals frantically for water. The other ship replies, "Throw down your buckets; you are surrounded by fresh water!": they are resting in the outflow of the mighty Amazon River.

The quantitative international relations community has often felt becalmed with respect to data. We have no American National Election Study, no U.S. Census or National Institutes of Justice data, and only so many ways one can analyze the World Handbook, Correlates of War, WEIS and COPDAB. But in fact, we are sitting amid a river of political data—both event-oriented and contextual—flowing past us every day from journalistic sources. Those sources are increasingly machine-readable, and if we can find a means of tapping them using the natural language capabilities of contemporary computers, we will find ourselves awash in data.

<sup>14</sup> As with many parables of evangelical preachers, the movement of this second vessel under windless conditions is not explained...