Validity Assessment of a Machine-Coded Event Data Set for the Middle East, 1982-1992

Philip A. Schrodt and Deborah J. Gerner
Department of Political Science
University of Kansas
504 Blake Hall
Lawrence, KS 66045

email: Schrodt@ukanvm.cc.ukans.edu
Gerner @ukanvm.cc.ukans.edu

September 1993

ABSTRACT

This paper is a study of the validity of a machine-coded event data series for six Middle Eastern actors and the United States. The series is based on Reuters newswire story leads coded into the WEIS categories. The face validity of the data is assessed by examining the monthly net cooperation scores based on Goldstein's (1992) scale in comparison to narrative accounts of the interactions between the actors; the event data series clearly shows the major patterns of political interaction. The machine-coded data are also compared to a human-coded WEIS data set based on the *The New York Times* and *Los Angeles Times*. Almost all dyads show a statistically significant correlation between the number of events reported by the two series, as well as the number of cooperative events. About half of the dyads show significant correlation in net cooperation and the number of conflictual events; many of these differences appear to be due to the higher density of events in Reuters. Finally, the machine-coded and WEIS data sets are used in two statistical time series studies and are shown to produce generally comparable results.

Introduction

This paper is an assessment of the validity of an event data time series produced from Reuters newswire leads and computer coding. The advantage of machine coding is that it can inexpensively produce event data: the source text needs only to be downloaded from a data service and processed by the machine-coding program, so data are available for analysis almost immediately. In contrast, existing methods of event data coding involve the training and maintenance of a large number of human coders, a process that is time-consuming, expensive, and has resulted in event data collections being more than a decade out of date. The open question about machine coding is whether the differences in technique—computer coding of Reuters lead sentences versus the traditional human coding of entire newspaper articles—will produce comparable sets of data. This study will consider both the face validity of the machine-coded data and comparisons of the machine-coded data set with data coded using traditional methods covering the same actors and time period.

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, grants, rewards, protests, demands, threats, and military engagements). The most commonly used event data sets for international relations research are Azar's (1982) Conflict and Peace Data Bank (COPDAB) and McClelland's (1976) World Event Interaction Survey (WEIS). Both data sets attempt to code all publicly-reported interactions by all states and some non-state actors. COPDAB includes the period between 1948 and 1978 and is available from the Inter-University Consortium for Political and Social Research (ICPSR); the public-domain WEIS set at the ICPSR covers 1966 to 1978.[1]   Both COPDAB and WEIS depend on human coding of events.

During the past three years, our group at the University of Kansas has been working on a project to code events from machine-readable sources automatically. We have used a machine coding program—the Kansas Event Data System (KEDS)—to create several dense data sets covering the foreign policy interactions of a limited number of countries between 1982 and 1992 and are using these data sets as a test bed to explore the possibilities of machine-coding generally.[2]   We have experimented with a variety of texts, including specialized regional sources (see Gerner, Schrodt, Francisco, and Weddle 1994) but our greatest success has been with Reuters News Service wire leads obtained from the NEXIS data service. The lead is usually a simple declarative sentence that

---

[1] John Davies at the University of Maryland is supervising a new project—the Global Event Data System (GEDS)—that is coding Reuters stories using an extension of the COPDAB framework; this data set begins in 1990 (Davies and McDaniel 1993). The WEIS data set has been extended until 1992 (see Tomlinson 1993) but at present is not available from the ICPSR.

[2] The KEDS program is described in greater detail in Schrodt, Davis, and Weddle (1993) and Gerner, et al. (1994); it can be used for machine-assisted coding as well as fully automated coding. The program runs on any Macintosh computer with at least 2 Mb memory and is designed to be used by individuals with modest computer experience. A copy of the program and its manual (*Word* format) are available from Schrodt.

summarizes the article: "The United Arab Emirates welcomed a resumption of formal diplomatic ties between Egypt and Syria after a 12-year rift." The machine coding relies primarily on a set of word patterns to identify political actors and countries; for example, it knows that "United Arab Emirates," "Egypt," and "Syria" are countries and that "welcomed" corresponds to the WEIS category 041, "Praise."

We have been coding using the WEIS categories; however, in principle these methods should also work with the COPDAB scheme. Much of our work has focused on the Middle East. This region 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, USA-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. In this paper, we focus on foreign policy interactions of six Middle Eastern actors—Egypt, Israel, Jordan, Lebanon, the Palestinians, and Syria—plus the United States during an eleven year period.

The research reported here has two principal goals. First, we want to test the viability of event data coded from Reuters leads using KEDS by examining their face validity in a region and time period with which we are quite familiar. If these data are able to identify accurately patterns of interaction consistent with the reports of area experts and our own research, this will increase our confidence in the automatic coding of machine-readable news reports as a tool for creating event data.

Second, we wish to compare the time series generated by KEDS with a human-coded WEIS data set covering the same dyads for the same period. We address three questions in this comparison. First, what is the general level of correlation between the two data sets when they are aggregated on a monthly basis? Since the machine-coded data is based on Reuters leads while the human-coded WEIS is based on the *New York Times* and *Los Angeles Times* articles, dissimilarities will reflect the different news sources as well as the distinctions between human and machine coding. Second, we examine several dyads over time to see how the two series compare. For example, do the values tend to increase and decrease in step? Does one series seem to pick up certain types of behavior more frequently than the other? Third, we replicate some of the statistical analyses of KEDS data that we did in earlier papers (Gerner 1993; Schrodt 1993) using human-coded WEIS data to determine whether the two data sets produce similar results.

Event Data and Machine Coding

Event data are one of the most common types of information used in quantitative international relations research[3]. At their most basic level, event data involve identifying the political actor who

---

[3] The advantages and disadvantages of event data are discussed by Azar, Brody and McClelland (1972), Burgess and Lawton (1972), Daly and Andriole (1980), Doran, Pendley and Antunes (1973), Azar and Ben-Dak (1975), Peterson

initiated an event (the source), the type of political action involved (the event) and the actor to whom the action was directed (the target). A date is also assigned to the event. The WEIS coding scheme uses about 60 nominal categories to classify events; COPDAB uses 16 ordinal categories on a conflict-cooperation scale. When a large number of events are considered in aggregate—for example, the COPDAB data set reports about 400,000 events—this can serve to measure the general relations between specific states. While event data necessarily simplify international interactions, "by taking proper account of the measurement properties of even rather noisy data, perfectly respectable statistical work can be carried out, sometimes with surprisingly good fits and statistical errors" (Achen 1987:6).

In the past, event data have been coded from newspaper sources. WEIS relied primarily on *The New York Times* while COPDAB used a variety of regional and international sources. Coding was essentially a process of content analysis: human coders, usually students, would read the newspaper articles, identify whether the article contained any political actions corresponding to events, and then code for the appropriate source, event, and target. While this human coding was able to produce useful data sets—and prior to the introduction of high-capacity microcomputers, human-coding was the only technique possible—it has several disadvantages. Foremost among these is cost: the human coding of tens of thousands of events is a very slow process. After government funding for the collection of event data stopped in the late 1970s, event data sets were impossible to maintain. Cost also reduced flexibility in the development of event data sets. Since it was impractical to code an entire data set more than once, both WEIS and COPDAB use a very general set of event codes that are designed to work with multiple theoretical questions. These coding schemes work relatively well for military and diplomatic behavior, but they often provide insufficient detail on issues such as economic interactions, environmental concerns, the role of sub-state actors, and other aspects of international relations that were not emphasized when the codes were first developed in the late 1960s.

A second problem with human coding is reproducibility. While considerable efforts were initially taken to increase inter-coder reliability in assigning codes, this reliability was difficult to maintain across a continually-changing population of student coders or, in the case of WEIS, across several different institutions over a period of two decades.[4] The coding manuals have not always been preserved, and coder training involved a number of issues not considered explicitly in the manuals.

A third difficulty with human coding involves the influence of preconceptions, culture, and other factors on the individual coders. While two individuals might be taught to code in an identical fashion when dealing with abstract descriptions of political events, psychological and perceptual

---

(1975), Munton (1978), Howell (1983), Vincent (1983), McClelland (1983), Goldstein and Freeman (1990), and Merritt, Muncaster, and Zinnes (1993). Marlin-Bennet and Roberts (1993) provide a recent discussion of event data from a research perspective.

[4] Tomlinson (1993) gives an institutional history of the WEIS project. The collection of COPDAB was done entirely under the supervision of Edward Azar until his death in 1991.

biases are likely to affect the coding of actual events such as the Vietnam War or the Arab-Israeli conflict. For example, if a coder has a preconception that Israel is allied with the United States and antagonistic towards Syria, the coder will tend to miss situations of conflict between Israel and the US and situations of cooperation between Israel and Syria (see Laurance 1990). Because these preconceptions differ between coders and vary with specific political contexts, they are difficult to control.

Machine coding gets around a number of these problems. Machine coding is far less expensive than human coding and, when a contemporaneous machine-readable source such as Reuters is available, a data set can be kept current without difficulty. Machine coding uses dictionaries of actor and verb patterns to determine coding, so it is also relatively easy to experiment with alternative or specialized coding schemes by changing these dictionaries. For example, the Protocol for the Assessment of Direct Action (PANDA) project of the Center for Nonviolent Sanctions at Harvard University is using KEDS to assist in coding a specialized data set dealing with nonviolent sanctions, a type of activity for which WEIS and COPDAB provide little differentiation. This project uses the WEIS dictionaries we developed as a foundation but has supplemented these with the additional codes and vocabulary required to create their specialized data set.

Unlike simple forms of content analysis that involve only counting words, the KEDS program uses some linguistic knowledge in identifying the role of words. For example, while the source of an event is usually the first actor in the sentence and the target is the second, this pattern is reversed when the sentence is in passive voice.[5] KEDS recognizes this formulation and correctly identifies the source and target. Similarly, KEDS recognizes compound noun phrases. The phrase "Jordan asked Egypt and Syria" generates two events with Jordan as the source; one event has Egypt as the target and the other has Syria as the target. While KEDS cannot handle very complicated sentences involving a large number of actors and complex causal relationships—"The United States dismissed Israel's apparent rejection of an Egyptian plan for talks with the Palestinians as 'parliamentary maneuvering' and said the door was not closed to peace"—it can deal with about 90% of the sentences found in Reuters leads.

At present, KEDS's ability to assign the same code as a single human coder is 80% to 85%. Most of the residual coding disagreement that is not due to complex sentence structures comes from ambiguities in the WEIS coding categories themselves. This level of reliability between coder and machine is typical of that found between human coders in event data projects: Burgess and Lawton (1972:58) show a mean intercoder reliability of 82% for eight projects where that statistic is reported.

---

[5] For example, "The resumption of formal diplomatic ties between Egypt and Syria was welcomed by the United Arab Emirates."

Weber (1990:17) divides content analytic reliability into three components: the ability of a single coder to assign the same code to a text when coding it more than once (stability); the ability of two coders to assign the same code consistently (reproducibility); and the ability of a group of coders to conform to a standard (accuracy). Once a set of patterns is established, the stability of machine coding is 100% since the machine is executing a fixed algorithm. In our experience, the reproducibility of machine coding is comparable to that of humans, and a machine is obviously not influenced by the context of an event or by intrinsic political or cultural biases.[6] Furthermore, the machine is not subject to coding errors due to fatigue or boredom, and, once a coding vocabulary has been developed, the machine does not require retraining. This leaves the issue of accuracy, the focus of this paper.

Face Validity of the KEDS Data for the Arab-Israeli Conflict

To create the 1982-1992 Arab-Israeli conflict data set, we downloaded text from the Reuters News Service using the NEXIS on-line data base of Mead Data Central.[7] Of the various news sources available on NEXIS, Reuters provides the most thorough and consistent coverage of the Middle East. In this study, 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.[8] 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.[9]

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

---

[6] The coding dictionaries may reflect biases, but these will be explicit and can be examined by another researcher; these dictionaries are also applied consistently to all actors and in all contexts.

[7] 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."

[8] The coded data are available from the authors in ASCII format; we anticipate depositing a data set covering 1982 through 1993 with the ICPSR in 1994.

[9] 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 1982-1992 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 present focused their attention on the domestic and international conflicts.

WEIS event as described in Goldstein (1992) and totaling the events for each month. This produces a single numerical score on a conflict-cooperation dimension similar to that used in COPDAB.

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 and four times as conflictual as Palestinians→Israel. 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 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).

As Figure 1 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 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
Israeli-Lebanese Net Cooperation, 1982-1992
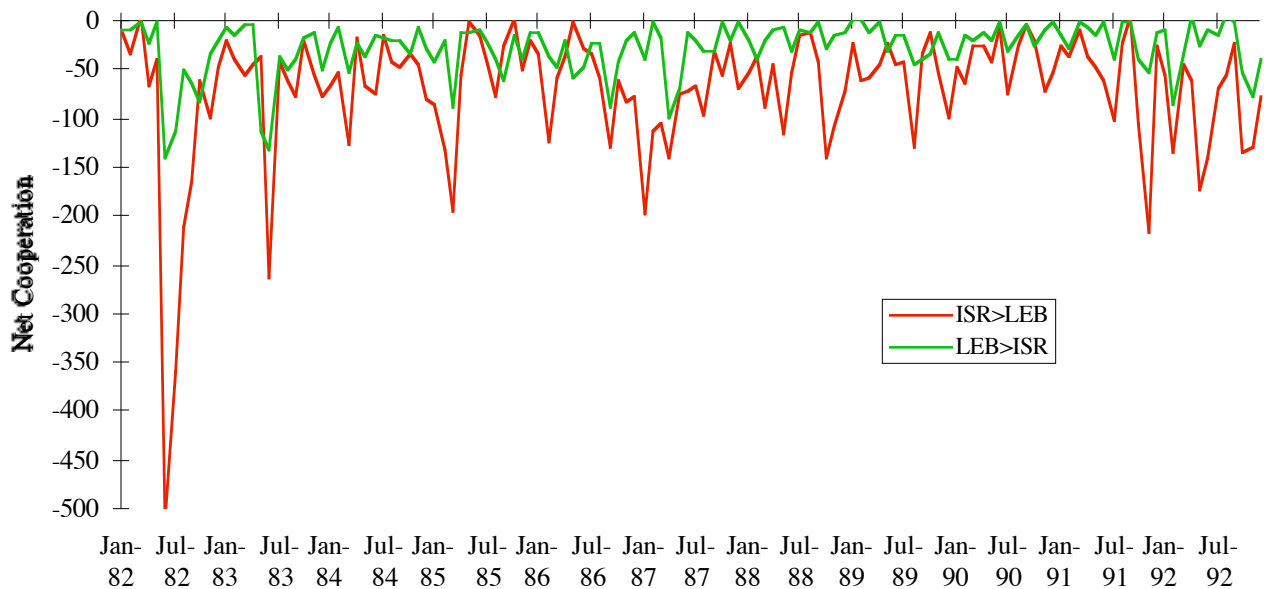

United States actions toward Israel show the greatest amount of cooperative behavior of any directed dyad in the data set (see Figure 2). This can be attributed in part to the consistently high level of consultation between the two countries. Thirty-five percent of reported USA→Israel events and 36 percent of Israel→USA events fall into the WEIS "consult" category, with peaks in 1982-1983 and again in 1991-1992. 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.

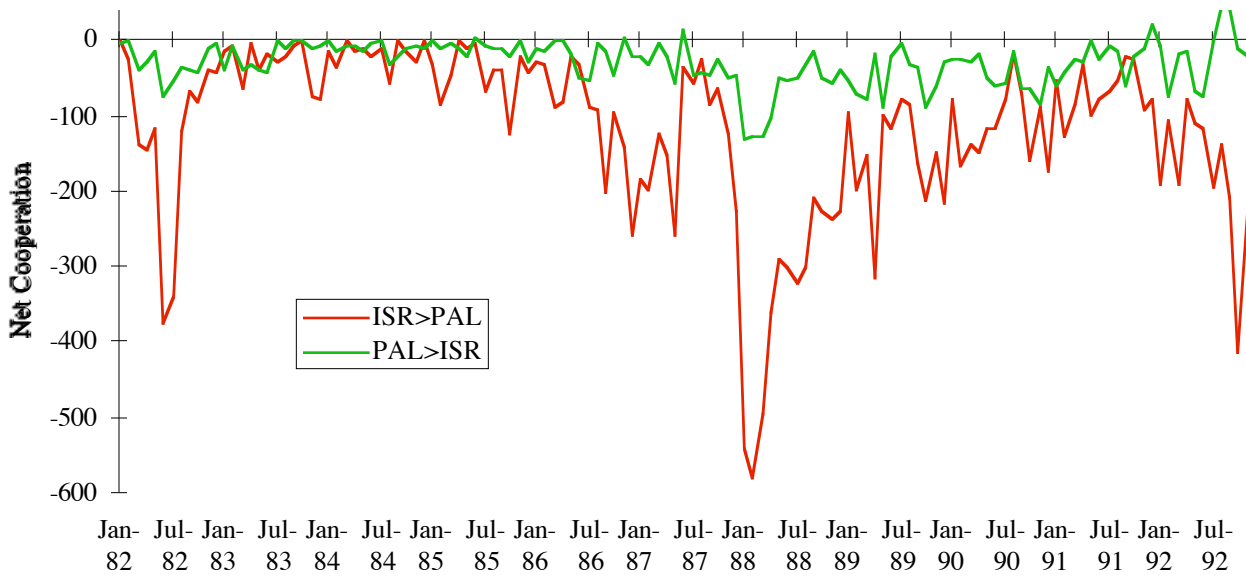--

FIGURE 2
Israeli-United States Net Cooperation, 1982-1992
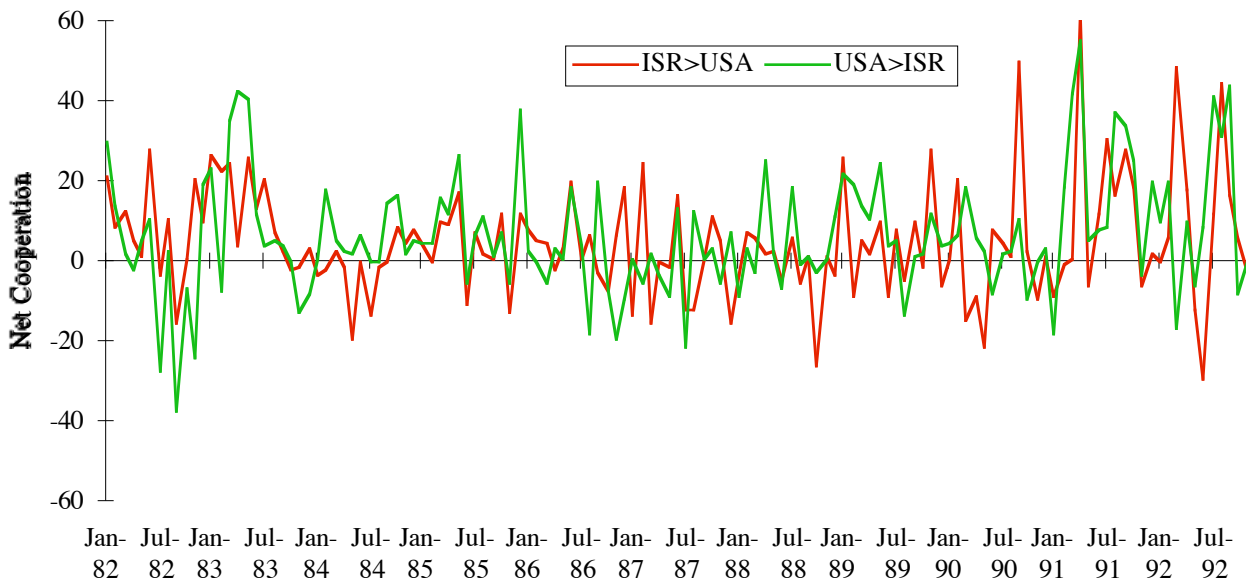
FIGURE 3
**Israeli-Palestinian Net Cooperation, 1982-1992**

Finally, Israeli and Palestinian net cooperation over time is particularly interesting (see Figure 3). 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 Occupied Territories) 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. One puzzle is the apparent drop in Israeli hostility just before the outbreak of the *intifada* in December 1987. 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.

Although the net cooperation measure is valuable in assessing the overall tone of events directed from source to target, it can mask important individual events that are inconsistent with the general pattern in a given month. Thus, it is often useful to look at the distribution of events in specific WEIS event categories. Here we present a brief example with reference to Palestinian-Israeli interactions. Two of the most common types of events between these two actors are "consult" and "force." Figures 4 and 5 illustrate how the extent of these actions has changed over time. As in the net cooperation measures, both the violence and the negotiations associated with the 1982 War in Lebanon show up, as does the beginning of the *intifada*. The dramatic increase in "consult" events in 1991 and 1992 reflects the ongoing regional peace negotiations. The jump in Israeli force actions in 1992 is also consistent with the narrative account for that period, when the new Israeli Labor government escalated activities against Palestinians (particularly Islamic fundamentalists) in the Occupied Territories.

_____

FIGURE 4
**Palestinian "Consult" and "Force" Actions Toward Israel, 1982-1992**
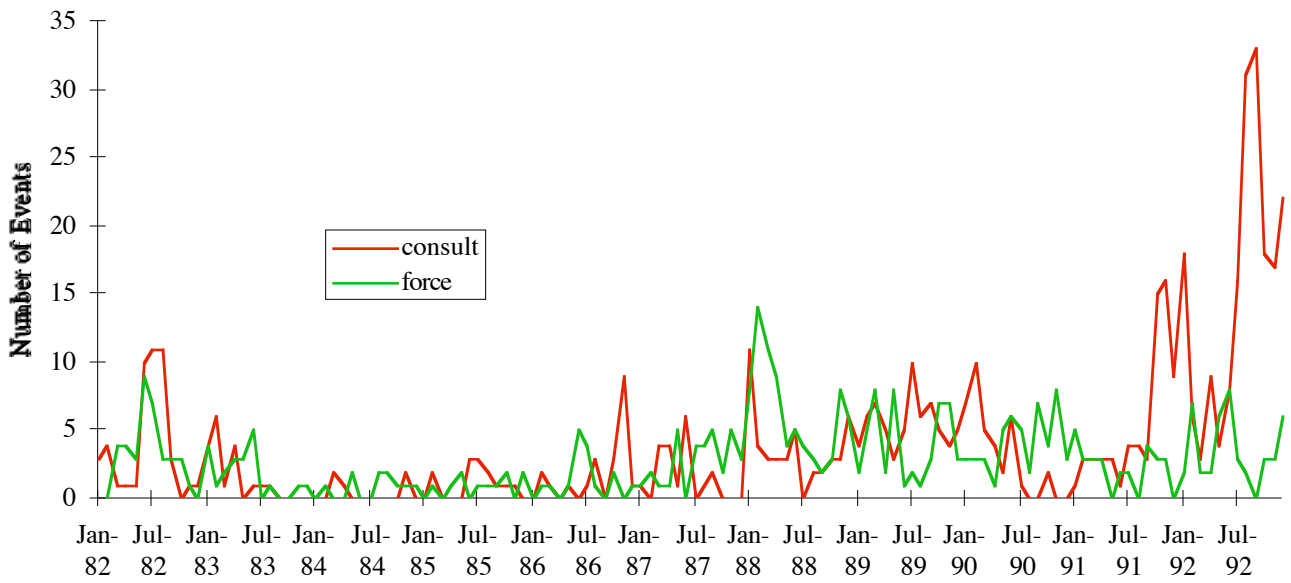
FIGURE 5
Israeli "Consult" and "Force" Actions Towards Palestinians, 1982-1992
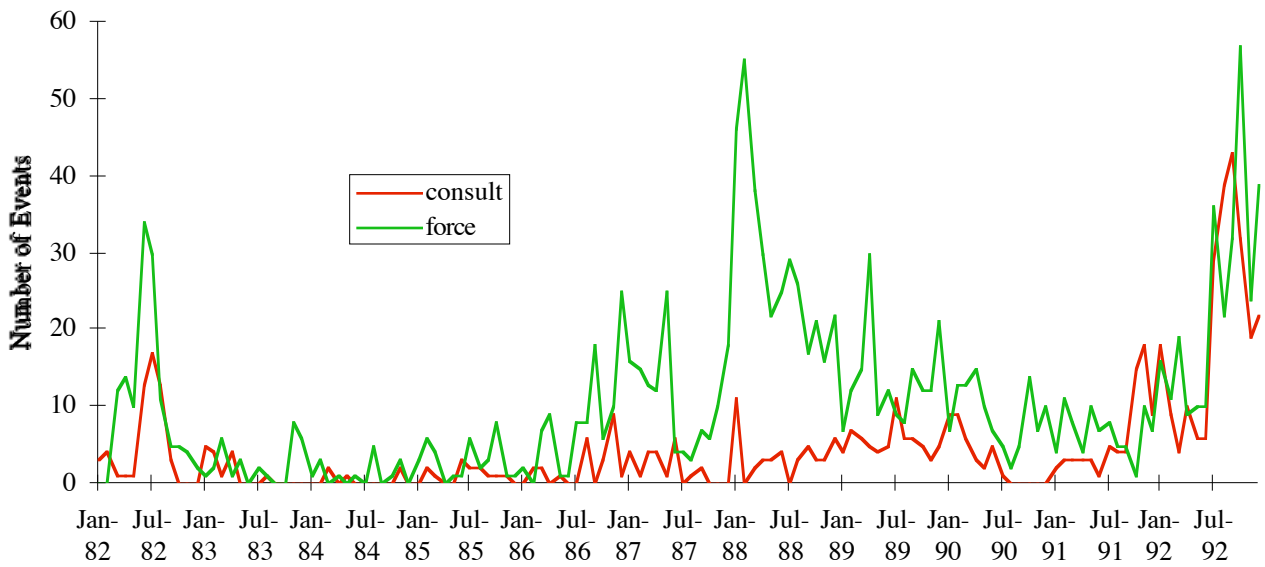
Another check on the face validity of these event data is the relationship between the number of Palestinians shot and killed by Israeli soldiers and settlers during the *intifada,* as recorded by a source independent of Reuters, and net cooperative Palestinian and Israeli actions. While this is a more specific type of behavior than the aggregate net cooperation measure, one would expect the two series to covary. Figure 6 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 overall tend to be those 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 cooperative actions toward Israel.[10]

_____

FIGURE 6
Palestinian-Israeli Net Cooperation and
Palestinians Shot to Death by Israeli Forces during the Intifada
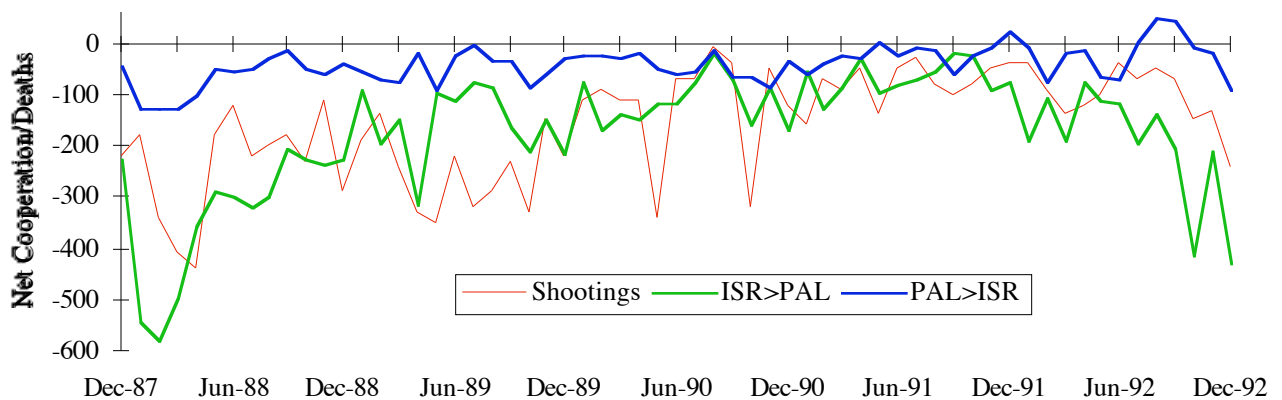
Note: Number of shooting deaths was multiplied by -10.

_____

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 except for the decline in Israeli net conflict toward the

---

[10] 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 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.

Palestinians just before the *intifada*. On the basis of face validity, the potential for using KEDS-coded Reuters wire feeds to measure basic patterns of conflict and cooperation in the Middle East looks very promising.

Comparison with the WEIS Data Set

While the KEDS data have considerable face validity, an equally important issue is whether they are comparable to the existing human-coded WEIS data.[11] In order to test this, we obtained human-coded data from the WEIS data set maintained by Rodney Tomlinson at the U.S. Naval Academy (Tomlinson 1993). To compare the two data sets, the WEIS data were aggregated to monthly intervals and then compared with the monthly aggregations in KEDS.[12] Four variables were correlated for each dyad: the total number of events, the net cooperation measured by the Goldstein scale, the number of cooperative events (WEIS codes 01 through 10, not including comments) and the number of conflictual events (WEIS codes 11 through 22). Table 1 summarizes these correlations; correlations by dyad and the number of events in KEDS and WEIS for each dyad are reported in Table 2 in Schrodt and Gerner (1993:16).

TABLE 1

Percentage of Dyads Showing Significant Correlations between

KEDS and WEIS Measures

Variable	All dyads	Dyads with WEIS>120 events
Number of events	93%	95%
Net Cooperation	52%	58%
Cooperative events	95%	100%
Conflictual events	55%	84%

----

[11] In this section, "WEIS" refers to the data set maintained by Tomlinson and "KEDS" refers to our data set; as noted earlier, the KEDS data are coded using the WEIS coding scheme. The WEIS data goes only to 1991 so we did not include 1992 in the comparison. The WEIS data have been collected by Charles McClelland and several of his students. The collection of the data for 1982-1986 was supervised by Richard Beal and Frederick Roth at the National Security Council in the White House using *The New York Times* as a source; the data for 1987-1989 were coded by McClelland from the *Los Angeles Times*; and the data for 1989-1991 were coded under the supervision of Tomlinson using *The New York Times*.

[12] The WEIS set does not directly code for the Palestinians as an actor, but codes for an "ARAB COM" ("Arab community") that in the context of the dyads we are studying usually refers to Palestinians. This was converted to the PAL code in the comparison.

Three general patterns are evident in from the comparison. First, there are statistically significant correlations—usually at the 0.001 level—between the total number of events in the two data sets for almost all of the dyads. This is also true for the number of cooperative events reported by WEIS and KEDS. The dyads where significant correlation is not found are usually those where WEIS reports only a small number of events. Because WEIS is based on *The New York Times* and *Los Angeles Times*, while KEDS is based on Reuters, KEDS has approximately three times as many reported events as WEIS and consequently often shows monthly variations in behavior where WEIS reports only zeros. There are substantially fewer significant correlations for the net cooperation value and for the number of conflictual events—roughly 55% of these are significant at the 0.05 level—although if one looks only at the cases where WEIS records an average of at least one event per month (i.e., a total of at least 120 events for the period), the percentage of significant correlations increases to 84% for the conflictual events.

While the correlation between the net cooperation values appears low, the correlation between WEIS and KEDS is substantially higher than Vincent (1983) found in a comparison of WEIS and COPDAB. Vincent correlated the total annual dyadic weighted conflict scores reported in WEIS and COPDAB for 128 countries between 1966 and 1978. He found correlations ranging from a high of 0.92 in 1969 to a low of 0.14 in 1972. Vincent's unit of analysis was the country and the annual dyadic behavior for each actor was summed across all of its dyadic partners; Vincent presumably used actor-year totals since a test using all 16, 256 dyads would be dominated by (0,0) cases.

Table 2 summarizes Vincent's findings and also shows the results of a similar comparison between the KEDS and WEIS data sets for the Middle East using both the actor-year (N=7) and the dyad-year (N=42). The average correlation between KEDS and WEIS is substantially higher than that reported between COPDAB and WEIS; the variance between years is also less. In the WEIS-KEDS comparison, only 1983 shows an anomalously low correlation; this is in all likelihood due to events in Lebanon and may involve variations in how actors were identified as well as differences in coding the events.[13] Vincent's average COPDAB-WEIS actor-year correlation is 0.66; the average KEDS-WEIS dyad-year correlation is 0.80 and the average actor-year correlation is 0.92, although the latter figure is undoubtedly somewhat inflated by the small sample size. The COPDAB-WEIS average is affected by the unusually low scores in 1972 and 1975, but even if these years are eliminated, the average is only 0.74, somewhat less than the KEDS-WEIS comparison.

[13] During 1983 a number of sub-state actors in Lebanon contested the presence of Israeli, U.S., Syrian, and other foreign troops. KEDS codes some of the more persistent of these groups—for example Hizballah and Amal—as distinct actors; WEIS does not.

_____

TABLE 2

Comparisons of Event Data Sets by Actor-Year

|	KEDS-WEIS		COPDAB-WEIS	
year	by dyad	by actor	year	by actor
1982	0.74	0.95	1966	0.80
1983	0.59	0.77	1967	0.83
1984	0.76	0.90	1968	0.82
1985	0.84	0.90	1969	0.92
1986	0.80	0.92	1970	0.51
1987	0.91	0.98	1971	0.47
1988	0.96	0.99	1972	0.14
1989	0.93	0.95	1973	0.55
1990	0.69	0.90	1974	0.68
1991	0.80	0.94	1975	0.36
			1976	0.76
			1977	0.87
			1978	0.93
Average	0.80	0.92		0.66
N	42	7		128

_____

We recognize that these tests may not be directly comparable. KEDS and WEIS differ only in their sources, while WEIS and COPDAB use quite different coding schemes and additional variance was introduced in converting them into a single system. Furthermore, COPDAB, unlike WEIS, depends heavily on regional news sources. The comparison does, however, indicate that the variation between machine and human WEIS-coded event data sets is substantially less than the differences researchers already face in dealing with WEIS and COPDAB.

Figures 7 and 8 compare the time-series for the net cooperation reported by the KEDS and WEIS data sets for the Israel→Lebanon and Israel→Palestinian dyads. The obvious general difference between the two data sets is the greater variance in the KEDS series, a consequence of KEDS having more events than WEIS. In a number of months, particularly those after 1989, the KEDS series show activity where WEIS does not.[14]

_____

[14] Schrodt and Gerner (1993:18-20) also compare the KEDS and WEIS time series for the Israel→Syria and Egypt→USA dyads. As in the dyads examined here, most differences are due to the greater density of KEDS; in

_____

FIGURE 7
Israeli->Lebanese Net Cooperation, 1982-1991
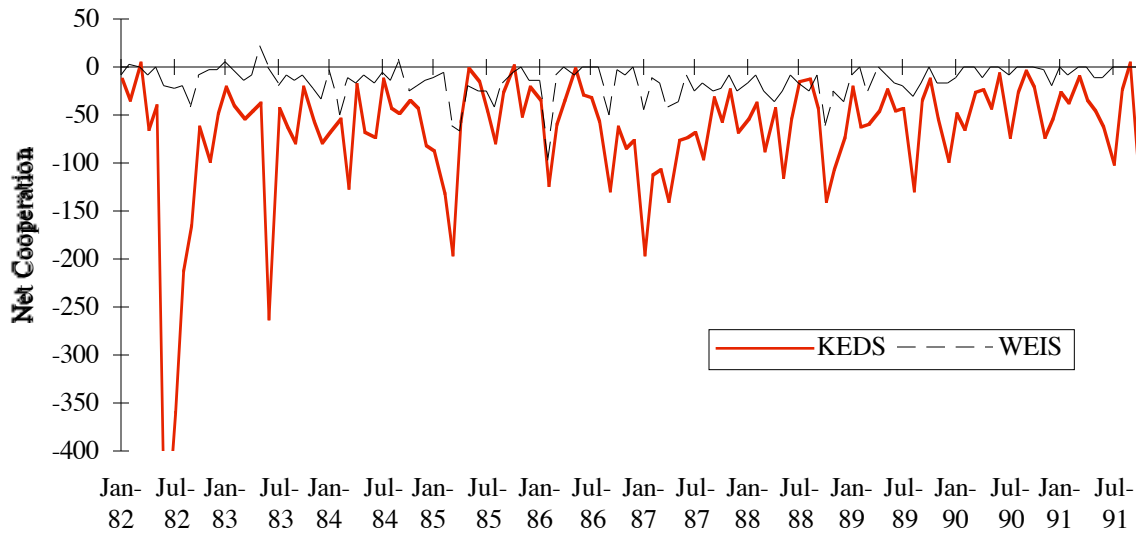
particular WEIS records almost no events for Israel→Syria after 1987. The only point where there is strong

disagreement when both series report events is a six-month period in 1987 where the two Egypt→USA series diverge due to KEDS reporting a series of diplomatic meetings while WEIS reports only disagreements resulting from those meetings.

FIGURE 8
Israel->Palestinian Net Cooperation, 1982-1991

_____
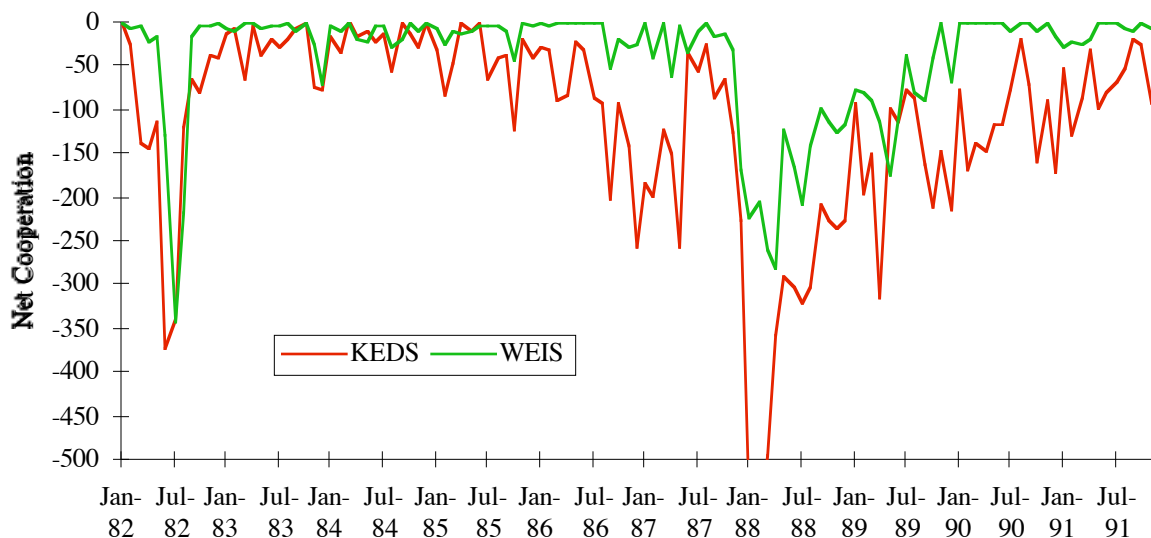
In Figure 7, Israel→Lebanon, both data sets report a significant amount of activity throughout
the entire period. This series also illustrates a major difference in coding rules. The KEDS series shows
a substantial increase in conflictual activity coinciding with Israel's invasion of Lebanon in the summer
of 1982 while WEIS shows almost no activity during the same period. The reason is that WEIS codes
the use of force during the invasion—for example, the bombing of West Beirut—as being directed
against "ARAB COM" (Arab community, which we recoded as "Palestinian"), and reserves the
"Lebanon" designation for actions such as protests and consultations specifically involving the Lebanese
government. KEDS, in contrast, codes "Beirut" as "Lebanon" and only codes Palestinians as a target
when they or the PLO are explicitly mentioned. This difference goes beyond the question of human
versus machine coding and relates to more general event coding issues that often do not have
unambiguous solutions.[15] Except for the invasion of Lebanon, the two series covary fairly well, despite

---

[15] We deliberately included Lebanon as one of the test cases for KEDS because we knew that the various political
groups in that country would prove challenging to code. They have. The WEIS coding is consistent with Israel's
view of the invasion, since Israel's explicit target was the PLO and Palestinians in Lebanon generally. However, the
large number of non-Palestinian Lebanese living in West Beirut in 1982 perceived that they, rather than the
Palestinians alone, were under attack by Israel, and some of those Lebanese, notably the Shi'a Moslems,
dramatically altered their behavior towards Israel as a result. The 1982 war in Lebanon provides a good
illustration of a situation for which the ability of a machine coding system to inexpensively produce two different
versions of the time series, one reflecting each interpretation of the invasion, could be used to an advantage.

the differences in the magnitude of activity; when the WEIS data set shows an increase in activity, this usually is associated with an increase in activity in the KEDS data.

Finally, Figure 8 shows the Israel→Palestinian interactions. This is the most dense of the dyadic series in both the KEDS and WEIS data sets. The two sources track each other quite closely in the major events of the time period such as the invasion of Lebanon and the onset and evolution of the *intifada*. However, the WEIS series shows the *intifada* essentially ending in January 1990, whereas KEDS reflects a more general decline that is consistent with the narrative record as well as with statistics on the number of shooting deaths. We have a sense that *The New York Times* coverage of the *intifada* decreased substantially after the autumn of 1989 when *The Times* shifted resources to covering the political changes in Eastern Europe; the WEIS series is consistent with that interpretation. The KEDS series also shows a distinct two-year increase in conflict prior to the *intifada* ; WEIS shows only a modest increase in conflict in the year preceding the *intifada*.

Overall, the KEDS and WEIS net cooperation time series show similar patterns. In most cases, the exceptions are found in months where KEDS/Reuters records activity while WEIS/*Times* does not; those months probably account for much of the unexplained variance in the correlations between the two series. A scattergram of the KEDS versus WEIS series for the USA→Israel dyad, a case with low correlation but a relatively high event density in both series, shows considerably more variance in the KEDS values than in the WEIS values (Schrodt and Gerner 1993:21). The cluster of points in the scattergram is arrayed around the WEIS=0 line; this in turn leads to the low overall correlation between the two series.

Secondary Analysis Using KEDS and WEIS

Event data are commonly used in statistical analyses rather than simply for descriptive purposes. In many such studies, minor disagreements in two time series will be unimportant, particularly if those differences are randomly distributed. We explored this issue by using the WEIS data set to replicate two of a number of KEDS-based statistical tests reported in other papers we have done recently: Gerner's (1993) assessment of cross-correlation in Middle Eastern foreign policy behavior and Schrodt 's (1993) analysis of the periodograms of behavior in conflictual and cooperative dyads. This section will not discuss the underlying theory or rationale for these tests but simply look at the issue of whether we would have come to different conclusions had we used the WEIS data set rather than the KEDS data.

Gerner (1993) used cross-correlation—the net cooperation of one dyad at time t correlated with the net cooperation of another dyad at time t ± k—to study the possibility of contagion in foreign policy interactions. In most of the dyad-pairs, the KEDS data shows no significant cross-correlation. This

was also found to be the case in spot checks of the WEIS data.[16]   However, in the KEDS data, there is some short-term cross-correlation (correlations at small lead and lag times) in dyad-pairs such as Israel↔Lebaneon. Using the WEIS data, we checked several of the symmetrical pairs discussed by Gerner, and found the same general pattern, although it was usually weaker in WEIS than KEDS (Schrodt and Gerner 1993:34). For example, the KEDS data shows significant cross-correlation in the Israel↔Jordan dyad pair for 0, 1, 2, 3 and 4 month leads, whereas WEIS shows this only at the 2-month lead. This is probably due to the difference in the number of events in the two series: WEIS records a total of 123 events for this dyad-pair while KEDS records 450 events.

Gerner found strongest pattern of cross-correlation in the Israel↔Palestinian dyad-pair. Figure 9 shows the cross-correlations produced by the KEDS and WEIS data sets for 1982-1991; the 0.05 significance level is around r=0.20. The short-term correlation in WEIS is higher than that in KEDS, but otherwise the lagged correlations (Palestinian→Israel $_{t-k}$ correlated with Israel→Palestinian $_t$) are very similar; both curves show significant correlations to -9 months. The leading correlations—Palestinian→Israel $_{t+k}$ correlated with Israel→Palestinian $_t$—illustrate a quite different pattern: while both curves show an initial drop to statistical insignificance at +3 months, the KEDS data picks up a strongly significant annual periodicity that is missing from the WEIS data. There are political reasons to assume that this seasonality is real: major Palestinian demonstrations likely to result in violent interactions with Israeli forces often occur on the anniversaries of important political events, while the weather discourages large demonstrations at other times of the year. If there is true seasonality, the KEDS data show a statistical feature of the political behavior that is not evident from the WEIS data.[17]

_____

[16] This result was surprising given the high level of overall correlation in the symmetric directed dyads as well as the large number of pairs of directed dyads whose patterns of net cooperation appear to be linked. Gerner (1993:16) notes that this absence of correlation may be due to the use of monthly aggregations: states typically respond quickly to each other's actions, so a month is a very long time in diplomatic terms.

[17] Annual seasonality would be centered on 12 months rather than the observed 11 months; the fact that the correlation is strongest at 11 months *may* be due to the use of the lunar calender to determine the timing of Jewish and Moslem holy days, particularly Ramadan, although this might just be a statistical artifact.

FIGURE 9
Cross-correlation of Israeli<->Palestinian Net Cooperation, 1982-1991
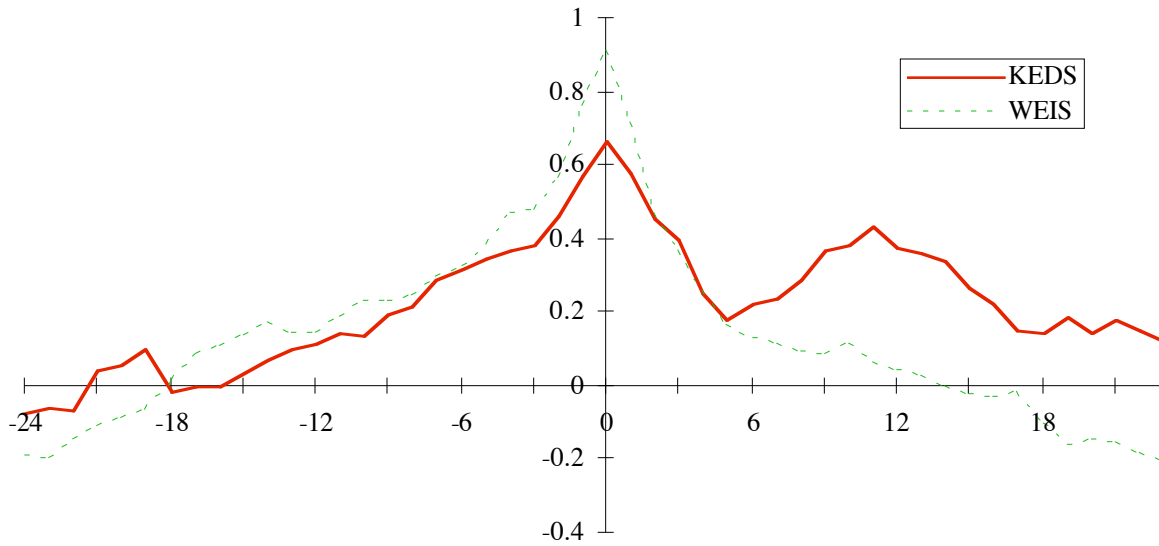
FIGURE 10
Cross-correlation of Israeli<->Syrian Net Cooperation, 1982-1991
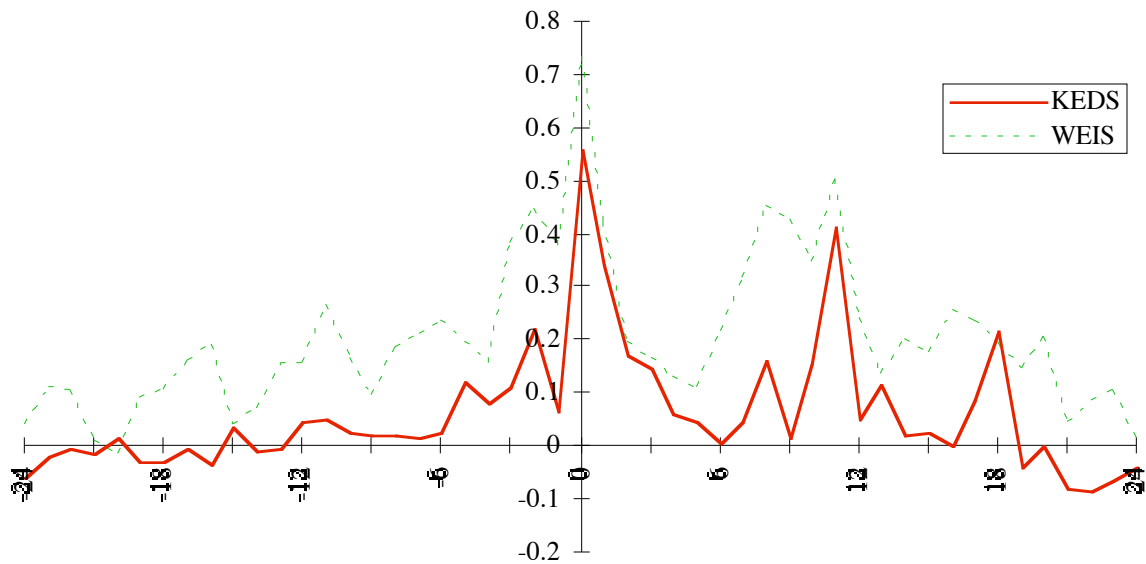
The one exception to the tendency of KEDS to exhibit stronger correlations than WEIS is found in the Israel↔Syria interactions; this is shown in Figure 10. As in the Israel↔Palestinian case, the general shapes of the two curves are similar, but the WEIS correlations are consistently stronger than those found in KEDS. Furthermore, WEIS shows a pattern of statistically significant leading correlations in the period +6 months to +11 months where KEDS finds that correlation only at +11 months. The KEDS correlation at +11 months, taken in isolation, would appear to be spurious, whereas the WEIS results show this to be part of a more general pattern. The number of events reported by WEIS for Israel↔Syria is relatively high; in these two dyads WEIS reports 58% as many events as KEDS, whereas in the data set as a whole WEIS reports only 36% as many events. Consequently when the event density of the two data sets is similar, the human-coded WEIS may be picking up statistical patterns not found in KEDS.

A similar pattern of cross-correlation is found in the two data sets for the *number* of cooperative events rather than the net cooperation values (Schrodt and Gerner 1993:24). For example, in the Israel↔USA dyad-pair the two data sets are in agreement on whether significant correlations exist in 42 of the 49 months for which the correlation was computed; more generally, the pattern of cross-correlation is very similar. The WEIS data produce stronger correlations at the long lags and leads than the KEDS data, although these values are not statistically significant.

Overall, the cross-correlations computed using the WEIS and KEDS data sets are similar; there are no obvious biases favoring one set of data over the other except for the higher event density in KEDS. In most cases, an analysis done with one data set would reach much the same conclusions as the other, although there will still be some idiosyncratic differences such as those found in the cross-correlations at long leads in the Israel↔Palestinian and Israel↔Syria dyad-pairs. Even these are not consistent, however: KEDS has stronger leading correlations for the Israel↔Palestinian dyad-pair and WEIS has stronger leading correlations for the Israel↔Syria dyad-pair.

We conducted one additional comparison of the two data sets using a complicated time series method, the periodogram produced from a fast Fourier transform (FFT) that is used to estimate the spectrum of a time series (see Gottman 1981). Schrodt (1993) used the KEDS data set to test for the presence of a "1/f spectrum" in selected dyadic interactions; the 1/f pattern is predicted for a system in a state of "self-organized criticality" (see Bak and Chen 1991). Schrodt's analysis found that the Israel→Palestinian interactions were characterized by a 1/f spectrum. The USA→Egypt spectrum is almost flat and dyads showing both cooperation and conflict (USA→Israel and Syria→Lebanon) are intermediate between these two patterns.

Figure 11 repeats this analysis with the KEDS and WEIS data sets; the periodograms have been smoothed by summing the 64 frequency bins produced by the FFT into 8 broader frequency bands.[18] The resulting periodograms are almost identical for the two data sets, particularly the distinction between the strongly 1/f distribution of Israel→Palestinian interactions as contrasted to the relatively flat spectra of the other two cases. The USA→Israel periodogram produced from the KEDS data has a distinctive hump in the middle frequencies that is absent in the periodogram produced from the WEIS data, but in both data sets the USA→Israel dyad shows more power in the middle and lower frequencies compared to the USA→Egypt dyad. As with the cross-correlation analysis, studies using the two data sets would produce much the same results, although there would be some idiosyncratic differences.

FIGURE 11

Periodograms using KEDS and WEIS Data
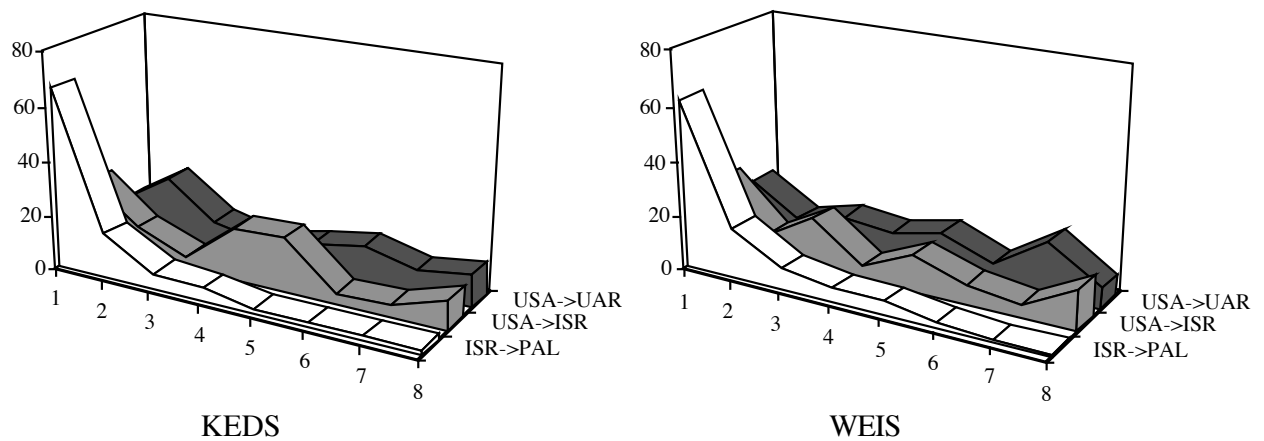
In the absence of a specific theoretical question and research design, it is impossible to predict whether the disagreements in the KEDS and WEIS data sets will produce differences in research findings, but the analyses done here generally indicate that variations will tend to be minor. For the most part, the two data sets mirror each other closely in their statistical characteristics and the differences that exist are largely due to the greater event density in the KEDS/Reuters data set.

---

[18] Since the 10-year series has 120 points and the FFT requires a series that is a power of 2, a short series of 4 zero values was added to the beginning and end of the series to produce a set of 128 points. This addition should have little effect on the estimated spectrum, and, as expected, the periodograms in Figure 11 are essentially the same shape as those of Figure 7 in Schrodt (1993), which was done with the 128 points available in a 1982-1992 series.

Summary and Future Research

This research has attempted to answer two principal questions. First, we wanted to assess whether KEDS-coded event data based on Reuters newswire leads covering the Arab-Israeli conflict for 1982-1992 accurately reflect the patterns of conflict and cooperation found in narrative sources. Second, we were interested in comparing the KEDS/Reuters data set with a *New York Times*-based, human-coded WEIS data set covering the same period.

It is clear from the results reported here that the KEDS/Reuters data set is able to identify the major trends in international conflict and cooperation in the region. Both the overall levels of net cooperation and specific patterns over time are generally consistent with the narrative record for the dyads examined. Major events—the 1982 war in Lebanon, the *intifada*, the peace negotiations that began in October 1991—show up clearly in the data. Given the skepticism with which many scholars regard event data in general—let alone machine-coded event data—the face validity of these data is no small accomplishment. The next step is to investigate whether the KEDS-coded data can also identify more subtle details of foreign policy interactions by looking closely at the specific WEIS event categories over time.

Although the event data set appears to be an accurate reflection of most foreign policy interactions in the region, the low amount of data for several of the dyads—particularly those involving Lebanon (other than with Israel or Syria) and Jordan—is a source of some concern. It is difficult to determine whether this is due to the lack of activity in these dyads or due to a bias in the Reuters reports. A systematic comparison of Reuters News Service daily event reports with other event data sources such as regional chronologies might indicate whether Reuters is systematically missing certain types of dyadic interactions reported in other sources.

The disagreements between the human-coded WEIS/*Times* data set and machine-coded KEDS/Reuters appear to depend more on the different sources than on the issue of human versus machine coding, but several conclusions can be made. First, despite the fact that KEDS was using only lead sentences, which tend to be syntactically simpler than sentences elsewhere in an article, KEDS still found three times as many events as the human-coded *Times* articles. Second, the two techniques clearly produce similar results in terms of gross characteristics such as the number of events and the number of cooperative events. There is less agreement in the net cooperation values, although in many cases this seems to be due to Reuters reporting events that were not reported in *The New York Times* or *Los Angeles Times*. To the extent that one can compare the WEIS-KEDS agreement with the WEIS-COPDAB agreement reported in Vincent (1983), one finds greater agreement between the human- and machine-coded series than between the two human-coded series. Finally, the differences between the WEIS and KEDS sets in two types of time series analysis—cross-correlation and spectral analysis—are quite small and idiosyncratic rather than systematic.

KEDS still makes mistakes that a human coder would not make. Some of these can be eliminated through the development of more thorough coding dictionaries but others will crop up later. Generally the sentences that are too syntactically complex for KEDS to code can be identified by the occurrence of an usually large number of actor and verb patterns; it would be simple to modify the KEDS program either to not code such sentences or, in machine-assisted coding, to allow a human to code them. Since complex sentences are only about 10% of the cases in Reuters, this would still provide substantial economies when compared to all-human coding.

The Reuters news wire is becoming widely available to academic researchers through remote data bases such as NEXIS. The Reuters leads seem to provide greater consistency over time than *The New York Times* and continue to show activity in dyads even when the *The Times* has lost interest because of competing stories. In most of the cases we examined, small changes in reported activity in *The Times* correspond to much greater changes in Reuters, and, in most cases Reuters anticipates the changes in activity recorded in *The Times* by two or three months. This last characteristic is particularly important in dealing with forecasting. Because *The New York Times* devotes considerable attention to the Middle East, this ratio of Reuters to *Times* events may be even higher for other parts of the world.

There are several possibilities for increasing event density even further. First, one could try coding each sentence in a Reuters article, rather than only the lead sentence, rejecting any sentences that do not contain verb patterns corresponding to events and sentences that are too complex to handle automatically. This would result in considerable redundancy in some event categories—for instance, if one actor is protesting the actions of another, that protest will usually be phrased in multiple ways throughout the article—but it would pick up some additional events such as consultations that are not always reported in the lead sentence. Second, we have found that regional sources, which can be made machine-readable with optical character recognition, tend to pick up different events than Reuters (see Gerner et al. 1994); the use of such sources would increase the event density. The question of whether increased density alone provides greater validity depends on the specific theoretical question being addressed, but with machine-coding this can at least be done without a significant additional investment of labor.

In summary, automated event data coding using machine-readable sources appears to be a practical alternative to the traditional human coding of newspaper articles. The machine-coded event series show the expected patterns of political interaction and correlate well with comparable human-coded event data. Machine-coding provides the possibility of inexpensively producing up-to-date event data sets, which should encourage the further development of event data analysis as an international relations research technique.

References

Achen, Christopher. 1987. "Statistical Models for Event Data: A Review of Errors-in-Variables Theory." Presented at the DDIR Conference on Event Data, Ohio State University, Columbus, 29-30 May.

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

Azar, Edward E., and Joseph Ben-Dak. 1975. *Theory and Practice of Events Research*. New York: Gordon and Breach.

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

Bak, Per, and Kan Chen. 1991. "Self-Organized Criticality." *Scientific American* (January):46-53.

Burgess, Philip M., and Raymond W. Lawton. 1972. *Indicators of International Behavior: An Assessment of Events Data Research.* Beverly Hills: Sage Publications.

Daly, Judith Ayres, and Stephen J. Andriole. 1980. "The Use of Events/Interaction Research by the Intelligence Community." *Policy Sciences* 12:215-236.

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

Doran, Charles F., Robert E. Pendley, and George E. Antunes. 1973. "A Test of Cross-National Event Reliability." *International Studies Quarterly* 17:175-203.

Gerner, Deborah J. 1990. "Evolution of the Palestinian Uprising." *International Journal of Group Tensions* 20:233-265.

Gerner, Deborah J. 1991. "Palestinians, Israelis, and the *Intifada*: The Third Year and Beyond." *Arab Studies Quarterly* 13:19-60.

Gerner, Deborah J. 1993. "Foreign Policy Interactions in the Middle East, 1982-1992: An Events Data Analysis." Presented at the annual meeting of International Studies Association, Acapulco.

Gerner, Deborah J., Philip A. Schrodt, Ronald A. Francisco, and Judith L. Weddle. 1994. "The Machine Coding of Events from Regional and International Sources." *International Studies Quarterly*: 37 [Note to editors: we don't know the page numbers yet]

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

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

Gottman, John M. 1981. *Time-Series Analysis.* Cambridge: Cambridge University Press.

Howell, Llewellyn D. 1983. "A Comparative Study of the WEIS and COPDAB Data Sets." *International Studies Quarterly* 27:149-159.

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

Marlin-Bennet, Renée and James C. Roberts. 1993. "Using Events Data to Identify International Processes: A New Unit of Analysis for International Relations." *International Studies Notes* 18:1-8.

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

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

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

Munton, Donald. 1978. *Measuring International Behavior: Public Sources, Events and Validity.* Dalhousie University: Centre for Foreign Policy Studies.

Peterson, Sophia. 1975. "Research on research: Events data studies, 1961-1972." In *Sage International Yearbook on Foreign Policy Studies* 3, ed. Patrick J. McGowan. Beverly Hills: Sage.

Schrodt, Philip A. 1993. "Rules and Co-Adaptation in Foreign Policy Behavior." Presented at the annual meeting of the International Studies Association, Acapulco.

Schrodt, Philip A. and Deborah J. Gerner. 1993. "Statistical Patterns in a Dense Event Data Set for the Middle East, 1979-1992." Presented at the annual meeting of the Midwest Political Science Association , Chicago.

Schrodt, Philip A., Shannon G. Davis and Judith L. Weddle. 1993. "KEDS: A Program for the Machine Coding of International Event Data." Manuscript, University of Kansas, Lawrence, Kansas.

Tomlinson, Rodney G. 1993. "World Event/Interaction Survey (WEIS) Coding Manual." Manuscript, United States Naval Academy, Annapolis, MD.

Vincent, Jack E. 1983. "WEIS vs. COPDAB: Correspondence Problems." *International Studies Quarterly* 27:160-169.

Weber, Robert Philip. 1990. *Basic Content Analysis*, 2d ed. Newbury Park, CA: Sage Publications.