Iowa State University

From the SelectedWorks of James McCormick

1972

Making and Measuring the International Event as a Unit of Analysis

Edward E. Azar
Stanley H. Cohen
Thomas O. Jukam
James McCormick

Available at: https://works.bepress.com/james-mccormick/14/

Chapter IV

MAKING AND MEASURING THE INTERNATIONAL EVENT AS A UNIT OF ANALYSIS

Edward E. Azar, Stanley H. Cohen, Thomas O. Jukam and James McCormick

In the last twenty years, the study of international politics has passed through a number of phases. These years have been very productive in terms of "making" data, building theory and developing sophisticated methodological techniques. Although it is not our intent to recite these developments, we shall attempt to locate the thrust of our work in relation to the research concerns of other scholars in the field. In the 1960s we witnessed valuable advances in the quantification of data for the purposes of developing empirical theory about the domestic and external behavior of nations (Singer, 1968; Kaplan, 1968; Pruitt and Snyder, 1969; Rosenau,

---

AUTHORS' NOTE: *This paper is based on the authors' report entitled, "Methodological Developments in the Quantification of Events Data," originally delivered at the Events Data Conference, Michigan State University, East Lansing, during April 15-16, 1970. We wish to thank the Department of Political Science and the Institute for Research in Social Science at the University of North Carolina at Chapel Hill, and the Department of Political Science at Michigan State University, for their support. Our thanks to Mrs. Betty P. Sawyer for her assistance in producing this in manuscript.*

1969). We believe that the research developments reported in this chapter complement and add to these contemporary concerns of quantitative internation politics.

We do not plan to justify or compare the various approaches to theory-building. We are certain, however, that any scholarly activity which seeks to develop better empirical theory about internation behavior will, at some point, require a data base. The testing of theory is a crucial aspect of the scientific process and is impossible without a data base (Singer, 1968: 1-13). Our main concern, therefore, lies with raising questions and solving problems the inductive theorist confronts when he begins the task of data making, data measurement and data analysis.

EVENTS DATA

Structuring the temporal signals which international actors transmit to each other into meaningful units of analysis has been an important problem for scholars and policy makers alike. Students of international relations have been engaged in the systematic gathering and interpreting of these signals as an integral part of their concern for theory-building about international behavior. The difficult research problems associated with these tasks become more tantalizing as the study of international relations assigns to itself the development of adequate quantitative methods for forecasting in such areas as foreign policy outputs or system transformation.

Once we appreciate that the term "signal" embraces a broad range of actions, we begin to see the magnitude of the analytical problems involved. Signals between international actors may vary in *intensity, frequency,* and *type.* Such signals (whether verbal or physical) can be characterized as actions (or interactions), transactions, and perceptions. Some of the research initiatives which have utilized these types of internation signals have resolved a number of these difficulties. For example, Karl Deutsch's (1957, 1966) work on transactional data, Charles McClelland's on event/interactions data (1968) and Robert North's (1963) on perceptual data have been very significant. However, we think that our work with events data demonstrates that more research is required before the methodological problems associated with analysis of behavior between states can be resolved.

International events, in the broadest sense of the term, refers to the overt behavior of national or international political actors toward each other or their environments. Although general interest in the use and analysis of events is an old phenomenon, there has been only marginal

concern with quantitative coding, measuring, analyzing and predicting events in any systematic manner.

In organizing events data for the purpose of quantitative analysis there are two major research strategies:

(1) The *categorization* strategy attempts to classify events into types of events; and

(2) the *scaling of the content of events* strategy measures events along some dimension of violence, conflict, cooperation and so on.

Although various individuals and groups differ in the number and type of categories or in the sophistication and reliability of scaling instruments they use, there is a good deal of overlap among the proponents of each strategy. However, we find it useful to organize events data in such a way that researchers may be able to scale the content of events along any dimension with any scaling instrument that appeals to them. We also feel that such a strategy allows for more flexibility and comparability than the categorization approach. Furthermore, when events are coded in the manner described below (i.e., without a substantial loss of the original content of the reported event), it becomes possible to categorize these events as well as *measure* them.

Before events data can be rigorously integrated into quantitative international politics, several definitional and methodological difficulties must be solved. Among these are:

(a) What is an event?

(b) How does one translate signals into events data?

(c) With what instruments can we measure events data?

THE CONCEPT OF AN EVENT

A concept has meaning only because the researcher means something by it. The conceptual meaning that the researcher gives to a term, like *event,* must not be so distant from the empirical indicators (the operationalization) that it is misunderstood. Our major task, therefore, must be to define *event* both conceptually and operationally. The conceptual definition of an event should indicate the limits of the variable we are utilizing while the operational definition of an event must specify the empirical referent for our conceptualization.

Therefore, we have conceptually defined *event* in the following manner: *Any overt input or output of the type "who does what to or with whom and when" which may have ramifications for the behavior of an international actor or actors.*

Operationally, an *event* is: *Any overt input or output of the type "who does what to or with whom and when" which may have ramifications for the behavior of an international actor or actors and which is recorded at least once in any publicly available source that meets the requirements discussed below.*

Our operationalization of an *event* is not far from that advanced by McClelland for *event/interaction* and transactions.[1] Both definitions consider the notion of "newsworthiness" implicitly or explicitly. We, however, are not required to make a distinction between "events" and "transactions" because of the more parsimonious operationalization.[2]

We maintain that an *event* has five components: actor, target, activity, issue-area, and time. All of these components must exist within the source if an event is to be identified. A *discrete event* exists only when at least one of the above components is different from one event to the other.[3]

The definitions of these components are:

Actor: That political entity which initiates an activity (i.e., who does or says—). An actor can be a nation-state or its spokesman, a regional organization, or an international organization. An actor can also be a political movement which has gained status within the behavior of the international system or any of its subsystems.

Target: That political entity to which an activity is directed (i.e., to *whom* something is done or said). A target can be a nation-state or its spokesman, a regional, or an international organization. A target also can be a political movement which has gained status within the behavior of the international system or any of its subsystems.[4]

Activity: Those actions, reactions, and interactions which are precipitated by clearly identified actor(s) or directed toward clearly identified target(s).

Issue-Area: A complementary category identifying an important component of an event other than actor, target, activity, or time. It defines the parameters of the event by elaborating the intensity, the arena, the intermediate participants, or topic of the event.

Time: The calendar day on which the source reports the event.

CODING EVENTS DATA

The operationalization of an event requires the text to be translated in a simple form: on a specified day, one international entity engages in an activity directed at itself or another international entity, thus constituting the event. This simple dyad of interaction represents the structure of events. In the "real world," the structural form is frequently deviated from in a variety of ways. Unless a standardized format is adopted by the

coders, the data in future analysis will be of limited usefulness and perhaps even require recoding. Furthermore, events must be coded in a format which maximally facilitates data processing. At our Conflict and Peace Data Bank, such a coding format has been developed for the purpose of banking events for about forty international actors from 1945 through 1969.[5] By systematically coding the components of an event, one will best be able to retrieve, measure, and analyze events data as needed.

In addition to coding the components of an event, we have thought it necessary to specify the source and the coder. The source can be a newspaper, journal, index, or any publicly available printed material. Initially, seventeen sources were chosen for operationalizing our definition of an event.[6] We chose these sources after preliminary investigation of various sources relevant to one of the areas we have been studying, namely, the Middle East. Careful examination showed that these sources made possible the collection of the greatest number of events and also covered all of the years under consideration. Another relevant criterion for selecting these sources was that they had compiled their events with *sufficient context for coding* purposes (i.e., the minimum amount of information necessary to facilitate our coding procedures). Some preliminary empirical work to confirm this statement was performed, but it still offers an interesting area for further research. We are satisfied, however, that these sources are most appropriate for making events data for our purpose.

The coders transfer the event from the source to the coded format. The selection and training of coders are crucial elements in the utilization of the coding of events. We selected the various coders in this manner: (a) coders had to have a basic interest in international politics; (b) coders had to become familiar with our work on the Middle East; and (c) coders had to demonstrate satisfactory coding performance.[7] Each coder was specifically trained in our coding format, given a number of trial events, and then examined by the more experienced coders to check for comparability in utilizing the coding format.

One immediate problem in training the coder was the reporting of multiple interactions in our sources. Since our definition of an event (with its various components) required this discreteness, such multiple reporting was disaggregated by the coder. This task was at first quite difficult and required constant checking for our coder to satisfy our definition of an event.

This disaggregation of reporting was, in fact, a simultaneous enterprise with that of discriminating the components of an event. The selection of time, actor, target, and source by our coder was quickly learned and

required little checking.[8] The selection of the activity and issue-area was much more difficult. The activity was limited to the use of one word and an allotted number of spaces. The basic instruction was to use the word that appeared in the source of the particular event; where this was not possible, a synonym was recommended. However, the coders consulted other coders when an extremely difficult passage could not be condensed to a single word.

The issue-area of the event was a problem because it also called for a summation of the text on the part of the coder. The only positive way of insuring consistency with the format was by constant checking of the coder. The coders were also assisted with questions that they raised over the issue-area. Through the use of such procedures—training and checking of the coders—there has been a substantial reduction of differential and inaccurate coding of the issue-area.[9]

The task of translating text into discrete events is not an interpretive procedure in that the coder is *not* required to evaluate the content for latent dimensions or categorize the event into one of several classes that might be ambiguous. However, the coder must be able to discriminate the actor, target, activity, issue-area, and time from a piece of lengthy, and often wordy, material. Our careful analysis of this latter procedure has revealed that coders are easily trained to operate on the text with this format. Moreover, the discrete and disaggregated events that are generated can be subsequently scaled on various dimensions without any significant loss of information or precision. On the contrary, by eliminating the often turgid and distracting context of the event, the scaler can concentrate on the scaling aspect of the task.

SCALING EVENTS DATA ON A VIOLENCE DIMENSION

Almost all of the instruments designed to measure internation behavior attempt (a) to determine some range of internation overt behavior; (b) to divide this range into sets of distinguished "classes" or "groups" of events; and (c) to arrange these classes or groups along some continuum such as from low to high conflict or cooperation. These instruments also utilize various levels of measurement.[10] With such advances on the part of several researchers, we were tempted to inquire: Are these scales similar enough that they can be integrated in some manageable form?

In order to integrate these various instruments into a simpler and more manageable one, we conducted a battery of experiments which produced a 13-point interval scale for measuring internation violence. Our procedures are outlined below.

A search through the above mentioned scaling instruments generated 201 "categories" or "markers." Each category was printed on a separate card. These instruments represented a broad range of types and levels of violence.

Five judges (three seniors, one graduate student, and one faculty member from the political science department at Michigan State University) reviewed the categories and were required to:

(a) rank-order these cards from very high cooperation to very high conflict;
(b) determine how many different classes of categories were characterized by these cards;
(c) describe each class in a few words to show why they were differentiated.

This procedure yielded thirteen different classes of internation activities ranging from class 1 which characterizes very low violent behavior (such as the merger of two nation-states into a new political entity) to class 13, which characterizes very high violent behavior (such as an all-out war between two nation-states). The interjudge reliability scores (rank order correlation among all pairs of judges) for the 201 categories were all above .84. The extreme categories were highly correlated but the middle rankings appeared more problematic. This is an important observation because it led us to further experimentation which we shall discuss later.

Satisfied that there could be "expert" consensus for 13 classes of internation violence, we next decided to find out whether relatively naive judges could satisfactorily rank-order these events, divide them into thirteen classes, and identify the *typical events* in each class. Fifteen MSU students (thirteen men and two women enrolled in an advanced course in international politics) were recruited to participate in this experiment. Each student was given a set of instructions and the deck of 201 cards. They were given a warm-up exercise and then allowed to perform their tasks independently. This experiment yielded the following satisfactory results: (a) the interjudge reliability scores were above .72; (b) collectively, the judges identified 62 distinct events (or 30.8%) as typical of the thirteen classes.

Now that we had 62 events which typify our thirteen categories, we began identifying the *most typical event* in each class. This most typical event in each class would become our marker-point for that class; whereas, the remaining ones would serve the purpose of further identifying each class of events. In order to carry out this task, we recruited 24 MSU students who had not been previously engaged in this experimentation.

They were seniors and graduate students majoring in political science. Each participant was asked to rank-order the 62 cards, divide them into thirteen classes ranging from class 1 which included the least violent events to class 13 which included the most violent events. They were also instructed to identify the *most* typical event in each class.

In each class, the event which received the highest composite typicality score was designated as the "marker-point" for that class. The list of markers was:

(1) Nations A and B merge to form a new nation-state.
(2) Nations A and B establish a regional organization among themselves.
(3) Nation A extends economic aid to Nation B.
(4) Nation A and B conclude a friendship agreement among themselves.
(5) Nation A receives support for its internal or external policies.
(6) Nations A and B communicate regarding issues of mutual concern.
(7) Nation A experiences limited political difficulties.
(8) Nation A makes a protest directed against Nation B.
(9) Nation A increases its military capabilities.
(10) Nation A encounters domestic politico-military violence.
(11) Nation A initiates subversion in Nation B.
(12) Nations A and B engage in limited war activities.
(13) Nation A engages in an all-out war against Nation B.

While the 13-point scale is capable of differentiating events into distinguishable classes, we nevertheless decided to determine the interval widths between these 13 points through the technique of *paired comparisons.* The paired comparison technique involves: (a) the formation of all possible distinct pairs of stimuli—if n is the number of stimuli, then the number of pairs would be $n(n-1)/2$; (b) the selection by a subject of the event in the pair which expresses or implies more violence; (c) the conversion of the proportion matrix derived from averaging the violence judgments for all subjects to interval scale values. This technique is based on a model proposed by Thurstone and formulated in Torgerson (1958: 159-204).

All possible distinct pairs of the 13 markers (totaling 78 pairs) were submitted in a random order to 52 MSU students.[11] Each student was given an explicit set of instructions and a warm-up exercise. The resulting scale values are given in Table IV-1.

Furthermore, we tested the adequacy of the assumption of an interval

TABLE IV-1
PAIRED COMPARISON SCALE VALUES

13-Point Ordering	Paired Comparison Values	13-Point Ordering	Paired Comparison Values
1	.000	8	3.011
2	.405	9	3.247
3	.957	10	3.741
4	1.321	11	4.203
5	1.623	12	4.828
6	2.331	13	5.536
7	2.774		

scale by correlating the scale widths resulting from the paired comparison technique with the results from the rank-ordering task. The Pearson product-moment correlation between the 13-point ordering and the paired-comparison scores was .98.

THIRTEEN-POINT INTERVAL SCALE: A DISCUSSION OF THE THIRTEEN POINTS

(1) Nations A and B merge to form a new nation-state: This category refers to the state of affairs in which the participants give up their former status as independent or sovereign actors and acquire a new international status in the form of a single nation-state. It is the state of affairs in which the newly established decision-making center begins to make decisions binding on all the merged territory.

This category may include such events: A and B establish a union, form a union, unify, join a union, federate, amalgamate, or integrate.

(2) Nations A and B establish a regional organization among themselves: This category refers to events which specifically indicate that a nation-state has allied with other nation-states by establishing or joining a regional (or an international) organization while retaining its formal independence. Examples of such regional or international organizations are common markets (e.g., EEC), defense organizations (NATO, the Egyptian-Syrian Defense Pact), cultural-political organizations (The Arab League), and so on.

It should be emphasized that the mere agreement to ally is not sufficient to scale events in this category. The event must specifically indicate that a nation-state has actually joined such a regional or international organization or alliance.

(3) Nation A extends economic aid to Nation B: This category refers to events which specifically indicate that a nation-state (or any other international actor) has actually given or consented to give economic or military aid to another nation-state (or any other international actor). This aid may include such assistance as participating in joint military maneuvers and extending military hardware and technical advice, but such aid may also be in the form of outright loans, grants or exchanges.

(4) Nations A and B conclude a friendship agreement among themselves: This category refers to events which specifically indicate that two or more nation-states (or other international actors) have actually concluded a cooperative agreement among themselves. Among the events which fall in this category are the establishment of diplomatic relations, the exchange of war prisoners, the acceptance of cease fire proposals, the reduction of trade tariffs, the suspension of economic boycotts, and the conclusion of cultural exchange treaties.

(5) Nation A receives support for its internal or external policies: This category includes events which specifically indicate that a nation-state's domestic or international policies or behaviors have received internal or external endorsement.

Among the events which may be included under this category are votes of confidence by a legislative body, the legitimate transfer of power within a nation-state, and domestic demonstrations favoring a nation's public policies.

(6) Nations A and B communicate regarding issues of mutual concern: This category includes those events which specifically indicate that two or more nation-states (or other international actors) have begun communicating with one another. Among the events which may be included under this category are the following: visits by key decision makers or official visitors' requests for international meetings, commencement of bilateral or multilateral discussions, issuance of joint communiques and invitations for state visitors, and explanations or comments (or the refusal to provide such) on public policies and behavior of nation-states.

(7) Nation A experiences limited internal political difficulties: This category refers to those events which may indicate behaviors expressing negative affect toward a nation-state's public policies, and public actions which limit the mobility, rights, or activities of individuals. Among these

events are the following: legislative censure of a government's policies, resignation or dismissal of a cabinet or key public officials, devaluation of a nation's currency, dismissal of a legislative body by a president or his cabinet, banning of certain political parties, censorship of the mass media, dismissal of key military leaders, strikes, and anti-government demonstrations.

(8) Nation A makes a protest directed against Nation B: This category includes events which indicate an increasing deterioration of relations between nation-states.

In this category are the following: reject, veto, accuse, demand, criticize, blame, condemn, repudiate, denounce, object, complain, warn, threaten, halt negotiations, break diplomatic relations, arrest and sentence citizens of other nations, boycott, place in embargo, and expel personnel.

(9) Nation A increases its military capabilities: This category includes events which specifically demonstrate that a nation-state (or another international actor) is expanding its strength. Among the events which may be included are: mobilization of armed forces; increase in defense budget; initiation of universal military training; acquisition of additional military equipment; and engaging in military maneuvers.

(10) Nation A encounters domestic politico-military violence: This category refers to those events in which a nation-state (or another international actor) experiences severe internal disorders. Among the events which fall in this category are enactment of martial law, riots, coups, and political assassination.

(11) Nation A initiates subversion in Nation B: This category refers to the initiation of support of subversive activities by one nation-state (or another international actor) toward another. Among the events which may fall in this category are sabotage of strategic goods, foreign support of anti-government guerrilla forces, reconnaissance activity and small scale or very limited border clashes.

(12) Nations A and B engage in limited war activities: This category refers to the participation in limited military hostile actions by a nation-state (or other international actors) toward another nation-state. Among the events which may be included in this category are stopping ships at sea, executing prisoners of war, and initiating bombing sorties.

(13) Nation A engages in all-out war against Nation B: This category

refers to a nation-state's declaration or initiation of either all-out conventional or nuclear warfare.

SELECTION OF SCALERS

From our preliminary work in developing the 13-point interval scale, we had discovered that sets of judges of differing educational backgrounds and other characteristics differed in their respective levels of reliability in the paired comparison experiments. In addition, we found that the results of the paired comparison experiments for certain groups differed somewhat from the rankings of the 13 classes of events. That is, by correlating the scale values resulting from the paired comparison experiment with the consensus rankings, substantial differences between groups were eveident. While we had suspected that some of these differences in ordering were attributable to the wording of certain components of the scale, we questioned whether the different composition of the groups was also significantly related to differences in ordering. We suspected that the "best" judges (i.e., those judges who had the highest interjudge level of agreement and who also had the highest correlation with the consensus ranking of the experts) were U.S. male citizens, seniors, and political science majors.

Thus we became interested in considering the extent to which the assignment of scaled values to the events (data) are influenced by variables such as education, sex, and cultural background of the scalers. Since this problem is open to experimentation, we decided to conduct a series of experiments designed to test some specific hypotheses related to this phenomenon. Our prior work with scalers suggested the following hypotheses:

(1) Differences in sex of scalers account for differences in scaling.
(2) Differences in education of scalers account for differences in scaling.
(3) Difference in ethno-cultural experiences account for differences in scaling.

In order to test these hypotheses 80 MSU students were recruited with the characteristics listed in Table IV-2. Each student was given an instruction sheet, a copy of the 13-point interval scale, a set of thirty discrete events and a set of five events as a warm-up.[12] They were instructed to measure on the 13-point interval scale the amount of violence contained or implied in each event. Before scaling the thirty discrete events, they were specifically advised to measure the warm-up set.

TABLE IV-2
CHARACTERISTICS OF THE SCALERS

Group	Students n	Educational Level	Sex	Citizenship or Country of Origin	Major
I	20	Freshman	Male	U.S.	Varied
II	20	Senior	Male	U.S.	Pol. Sci.
III	20	Senior	Female	U.S.	Pol. Sci.
IV	20	Senior and grad	Male	Arab countries	Varied

The test results were tabulated and the interjudge agreement scores were computed among the members of each group. The Pearson product-moment correlation statistic was used. The *average* correlation values for Groups I-IV were .72, .82, .70, and .69, respectively. These results allow us to suggest that U.S. male citizens who are political science majors make better scalers than either U.S. freshman males, U.S. senior females, or others. In the more general sense, we find that educational differences, ethno-cultural differences, and sexual differences account for differences in scaling.

Having found that U.S. male citizens who are political science majors demonstrated a high degree of internal agreement, we decided to compare their scaling results with those of a group of experts who were also employed to scale the same set of thirty events. The group of experts was composed of the author and two other MSU faculty members teaching international relations. The average expert interjudge agreement score was .97. When we compared the scaling values of the senior males with the group of experts we found a high degree of agreement—the average r was .91. This result further supports the conclusion that U.S. male citizens who are majoring in political science are the most reliable judges.

SUFFICIENT ENVIRONMENT FOR SCALING DISCRETE EVENTS

As our experimentation progressed, it occurred to us that a scaler more familiar with the historical events, their meaning, and the environment from which they were derived might perform more reliably on the scaling task than a scaler who is unfamiliar with the events data environment.

In order to answer this question, we recruited twenty U.S. male citizens who were seniors and majoring in political science (these students had not participated in any of the previous experiments). Each student read a 59-page summary of the behavior of the two international actors (Egypt and Syria) from 1952 to 1966 and listened to a lengthy discussion about

the major developments in the Middle East and particularly to a description of the behavior of Egypt and Syria toward each other and other actors in the region. Following this, each student measured on the 13-point interval scale the amount of violence contained or implied in the set of thirty discrete events used earlier by the other four groups. The average interjudge agreement was .87. The average scale values for this group were compared with the values for the group of experts, and yielded a product moment correlation score of .98.

By comparing the scores of this group with those of Group II, where we employed the same controls except for the amount of information about the events and their initiators or receivers, we found no significant statistical difference. Thus we concluded that scalers who are unfamiliar with the environment of the events they are scaling can perform as reliably as those scalers who are familiar with the historical events, their meaning and environment.

SUFFICIENT CONTEXT FOR SCALING DISCRETE EVENTS

The discrete set of events in the previous experiments had been presented in their full textual form. The gathering, storing, and scaling of events in their full contextual form introduced problems for the establishment of systematic banking and retrieving. Therefore, we decided to test whether a judge can scale events which have been coded according to our format with the same accuracy (reliability) as he would the event in full textual form.

A group of twenty students with the same characteristics as Group II were recruited and instructed to scale the set of thirty events coded according to the format above. Each student was given the same set of instructions that we gave the members of Group II. The average interjudge agreement for this group was .87. The correlation between the average scale values for this group with the average scale values for the group of experts was extremely high: $r = .99$.

Thus, in comparing this result with the reliability and external validity of Group II (i.e., two groups with the same controls except for that of coded versus textual events), we found no statistically significant differences. We concluded that if events are coded according to our format there is no disruption in the scaling task. On the contrary, it appears that our coding format facilitates the scaling task with the same or better reliability than scaling events in their textual form. This should be obvious since our coding scheme attempts to disaggregate the multiple events into discrete events and identify the components of an event; hence, the scaler

TABLE IV-3
SUMMARY OF EMPIRICAL STUDIES OF THE SELECTION OF JUDGES AND ENVIRONMENT-SUFFICIENT CONTEXT QUESTIONS

Judge Characteristics	Training	Event Format	Interjudge Agreement	Agreement with Experts
Freshman, male, U.S. citizen varied major	None	Text	.72	.93
Senior, male, U.S. citizen, political science major	None	Text	.82	.91
Senior, female, U.S. citizen, political science major	None	Text	.70	.96
Senior or graduate, male, a citizen of an Arab country, varied major	None	Text	.69	.96
Senior, male, U.S. citizen, political science major	Short course in the politics of the region	Text	.87	.98
Senior, male, U.S. citizen, political science major	None	Coded	.87	.99

can focus solely on the scaling task. Unlike other scales for the measuring of internation behavior, coding and scaling are procedurally distinct and separate steps.

The major findings of this extensive series of empirical studies on the selection of judges and the environment-sufficient context questions are summarized in Table IV-3.

A FINAL COMMENT

This article has defined an event, devised a systematic coding scheme for these events, and developed a scaling instrument to measure the level of violence in these events. The authors wish to again stress that parametric studies of the coding and scaling of events are fundamental to, and necessary for, the adequate utilization of events data for the development of theories of hypothesis testing in international relations.

By measuring the content of events in terms of some dimension (such as violence) with te use of an adequate measuring instrument (such as the 13-point interval scale), the overt pattern of violent behavior of an international actor can be profiled. In addition, if we assume that past behavior is a partial predictor of future behavior, then by profiling the intensity of past events we can attempt to forecast the intensity of future events. Furthermore, if we choose to determine what type of events is

likely to register that intensity, then we can refer to the 13-point interval scale to identify those types. This is possible because our scale is constructed in such a way that each of the thirteen points defines a set of discrete internation behaviors. For example, if we forecast the intensity of events initiated by Actor A at time t to be at level 8, then we can tentatively deduce that the behavior of Actor A at time t is likely to be characterized by demands, threats, boycotts, and so forth.

In sum, we are convinced that rigorous coding and scaling of international events offers a potentially rich data base for studying international behavior. Indeed, an international events data bank emerges as a useful "laboratory" for the development of quantitative models of international relations.

NOTES

1. McClelland defines event/interaction as "single action items of a non-routine, extraordinary, or newsworthy character that in some clear sense are directed across a national boundary and have, in most instances, a specific foreign target." He distinguished these *event/interaction* from *transactions* which are defined "as items of action that have at some point in time become so numerous, so commonplace, and so normal to their situation that they are accounted for conventionally in an aggregated form" (see McClelland and Hoggard, 1969: 713).

2. We assume, however, that changes in the volume and capacity of these more routinized events could become newsworthy, and therefore may be reported by one of our sources.

3. Therefore, if all the components of two events are the same, the events will be called *duplicate events.*

4. The problem of identifying the *target* in an event can be a difficult one. Although most public events tend to indicate their direct or specific targets in a satisfactory and explicit manner, other events do not. Furthermore, we have come across certain events which required us to make certain decisions as to who is the "real" target. In order to decrease the coding errors, we decided to relieve the coders of such important decisions, and instructed them to code the events "as reported." In other words, only *targets specified* in the publicly reported events would be coded as such. However since we are interested in being very close to the "real" world in our banking activity, we decided to go one step further and check every event as to its "real target." Thus if we find an event which clearly tells those of us who are specialists in the region (where data is being gathered), that the target in such an event is obviously more than one political actor, then we proceed to double code that event so as to accommodate our "direct" target and our "implied" target. Fortunately there are only a few events of this kind.

5. Consider, for example, the coding of the following text: September 1, 1955: An agreement was signed between Syria and Lebanon ending the Bank of Syria and Lebanon's concession to issue currency in Syria and transferring this right to the government.

The above text was disaggregated into two discrete events:

Data	*Actor*	*Target*	*Activity*	*Issue-area*
9/1/55	Syria	Lebanon	Agree	Ending currency concession, transferring right to the government of Syria.
9/1/55	Lebanon	Syria	Agree	Ending currency concession, transferring right to the government of Syria.

6. The sources used were: *New York Times Index, Middle East Journal, New Times–A Weekly Journal of World Affairs, Swiss Review of World Affairs, Middle Eastern Affairs, Deadline Data, Facts on File, World Almanac and Book of Facts, Keesings Contemporary Archives, London Times Index, Annual Register of World Events, Asian Recorder, Al-Ahram (Egypt), Jerusalem Post (Israel), Al-Nahhar (Lebanon), Arab Political Documents, Arab States and Arab League, China Mainland Review, Hsinau News Agency Release, The China Quarterly, Current Scene, China Report, Mizan, Mideast Mirror, Washington Post, Christian Science Monitor, Asian Almanac, Far Eastern Survey, Far Eastern Economic Review, Chronology of Events.*

7. The intercoder agreement among five coders for the actor, target, data, and activity components of each event was very high ($r \geqslant .82$).

8. One problem regarding the designation of actor and target occurred over an event such as the signing of an agreement by two nation-states. In such a case, it was difficult to determine which nation-state was the actor and which nation-state was the target. We therefore coded the event twice with each nation-state once as actor and once as target.

9. We contend that this "free-floating" approach to the coding of the category is more useful (if somewhat less economic) than the McClelland or Corson approach of using differing sets (or categories) of specified activities. This latter approach can make for the "forcing" of data. Our approach is to separate the coding and scaling tasks. Our scaling of the data will utilize all the components–not the activity alone. This will be discussed at greater length below.

10. Measurement of internation events was pioneered by Charles McClelland and his associates of the WEIS project, the groups at Oak Ridge Laboratory, Stanford University, and Northwestern University, and by a number of individual students of international politics (for examples of individual contributions, see Rummel, 1963, 1966, 1968; Moses et al., 1967; Richman, 1967; Corson, 1969).

11. This group consisted solely of U.S. male citizens who were political science majors, since our research findings had indicated that male U.S. citizens who are political science seniors make very good judges.

12. The thirty events were randomly selected from the raw data about Egypt and Syria during the period from 1958-1966. These events were left in their textual form (i.e., not coded in the manner we have described). However, each event was reworded in order to appear as one discrete event in which either Syria or Egypt were actors or targets.

REFERENCES

AZAR, E. E. (1969) "Quantification of events for the analysis of conflict reduction." Presented at the Seventh Research Conference, Peace Research Society (International), Ann Arbor, November 11-12.

--- S. H. COHEN, T. O. JUKAM, and J. M. McCORMICK (1970) "Methodological developments in the quantification of events data." Presented at the Michigan State University Events Data Conference, April 15-16.

--- (1972) "The problem of source coverage in the use of events data." International Studies Q. 16 (December).

BURROWES, R. with B. SPECTOR (1970) "Conflict and cooperation within and among nations: enumerative profiles of Syria, Jordan, and the United Arab Republic, January 1965 to May 1967." Presented at the Annual Meeting of the International Studies Association, Pittsburgh, April 2-4.

--- (1969) "The strength and direction of relationships between domestic and external conflict and cooperation: Syria, 1961-1967." Presented at the Annual Meeting of the American Political Science Association, New York, September 2-6.

CORSON, W. H. (1969) "Measuring conflict and cooperation intensity in international relations." Harvard University. (mimeo)

DEUTSCH, K. W. (1966) Nationalism and Social Communication. Cambridge: M.I.T. Press.

--- S. A. BURRELL, R. A. KANN, M. LEE, Jr., M. LICHTERMAN, R. E. LINDGREN, F. L. LOEVENHEIM, and R. W. VON WAGENEN (1957) Political Community and the North Atlantic Area. Princeton, N.J.: Princeton Univ. Press.

HOGGARD, G. D. (1969) "World event interaction survey codebook." University of Southern California. (mimeo)

--- (1968) "Conflict patterns in the interaction among nations." University of Southern California. (mimeo)

HOLSTI, O. R., R. C. NORTH and R. A. BRODY (1969) "The management of international crisis: affect and action in American-Soviet relations," pp. 62-79 in D. G. Pruitt and R. C. Snyder (eds.) Theory and Research on the Causes of War. Englewood Cliffs, N.J.: Prentice-Hall.

KAPLAN, M. A. [ed.] (1968) New Approaches to International Relations. New York: St. Martin's.

McCLELLAND, C. A. (1968) "The access to Berlin: the quantity and variety of events, 1948-1963," pp. 159-186 in J. D. Singer (ed.) Quantitative International Politics. New York: Free Press.

--- (1967) "The world event interaction survey: a research project on theory and measurement of international interaction and transaction." University of Southern California. (mimeo)

--- and G. D. HOGGARD (1969) "Conflict patterns in the interactions among nations," pp. 711-724 in J. N. Rosenau (ed.) International Politics and Foreign Policy. New York: Free Press.

MOSES, L. E., R. A. BRODY, O. R. HOLSTI, J. B. KADANE, and J. S. MILSTEIN (1967) "Scaling data on inter-nation action." Science 156 (May 26): 1054-1059.

NORTH, R. C. (1968) "The behavior of nation-states: problems of conflict and integration," pp. 303-356 in M. A. Kaplan (ed.) New Approaches to International Relations. New York: St. Martin's.

--- O. R. HOLSTI, G. ZANINOVICH and D. A. ZINNES (1963) Content Analysis: A Handbook with Application for the Study of International Crisis. Evanston, Ill.: Northwestern Univ. Press.

PRUITT, D. and R. C. SNYDER [eds.] (1969) Theory and Research on the Causes of War. New York: Free Press.

RICHMAN, A. (1967) "A scale of events along the conflict-cooperation continuum." University of Pennsylvania Research Monograph Series.

ROSENAU, J. N. [ed.] (1969) International Politics and Foreign Policy. New York: Free Press.

RUMMEL, R. J. (1968) "The dimensionality of nations project: research report 10." University of Hawaii. (mimeo)

--- (1966) "A foreign conflict behavior code sheet." World Politics 2 (January): 283-286.

--- (1963) "Dimensions of conflict behavior within and between nations" General Systems Yearbook 8: 1-50.

SINGER, J. D. [ed.] (1968) Quantitative International Politics. New York: Free Press.

TORGERSON, W. S. (1958) Theory and Methods of Scaling. New York: John Wiley.

YOUNG, R. A. and W. R. MARTIN (1968) "A review of six international event/interaction category and scaling methods." University of Southern California. (mimeo)