# Chapter Seven
# Conclusion

## 7.1. Event data in the contemporary environment

Event data analysis is now well into its fourth decade as a research technique for political analysis. In some respects the field has changed dramatically, notably in our ability to generate data using automated techniques. In other ways the original agendas of the 1960s have survived more or less intact, for example in the continued appeal of the problem of early warning. And finally there are a few characteristics of the 1960's agenda that were abandoned in the 1970s and 1980s only to be rediscovered in the 1990s: the emphasis on discrete sequences rather than continuous time series measures, and the development of new coding schemes to supplement or replace the original WEIS and COPDAB.

Throughout this volume, we have emphasized four themes that have emerged from our work with the KEDS-generated data and other data sets such as BCOW. This section will review those themes.

### 7.1.1. Machine coded event data produce credible results

If a machine-readable source of information is available for the events one intends to code, machine-coding definitely appears to be the preferred approach unless one is coding only a very small volume of reports. Consequently event data, once one of the most expensive and least flexible of the data-generation methods available to researchers, is in many circumstances one of the least expensive and most flexible. Dictionaries have been developed for a variety of coding schemes, including WEIS, BCOW and PANDA, and in all likelihood additional dictionaries and specialized coding systems will become available in the near future as various projects currently underway are completed and published.

In contrast to the earlier human-coded data sets, which locked researchers into maintaining a fixed set of categories despite changes in the theoretical questions being addressed and the political circumstances being coded, automated coding allows for experimentation with a variety of coding schemes, as well as allowing a data set to be maintained in nearly real time. None of our experiments have indicated any obvious statistical problems with machine-coded data; to the contrary, the machine-coded data tend to produce credible results when analyzed with a variety of different techniques.

This is not to say that machine coding techniques are flawless, nor that machine coding is always appropriate. For example, most of the BCOW data set would be difficult to code by machine, both because the historical sources used to generate that data are neither machine-readable nor in the simple declarative format of wire service reports, and because of the complexity of the BCOW coding system. Machine coding is better at coding actual events rather than making subtle distinctions between positive and negative comments, and the extent to which machine coding can accurately code the context of an event is still unclear.[1] As we discussed in Chapters 1 and 2, the parsing methods used in KEDS are relatively simple and we would hope that a new generation of event coding programs could correct some of the weaknesses of the sparse parsing approach. But with a few exceptions, we would expect that most future work will focus on refining automated techniques rather than reverting to a reliance on the coding skills of undergraduates.

### 7.1.2. Event data can be analyzed with a variety of techniques

There is no single method for analyzing event sequences, and since the 1960s a variety of approaches have been used, ranging from simple contingency table analysis to advanced econometric techniques to the clustering and discrete sequence methods we have employed in this volume. Each method has strengths and weaknesses, and they encompass a wide variety of

[1] PANDA codes several contextual "issues" but we are not aware of any studies that have analyzed these systematically.

underlying assumptions. We are not completely satisfied with any single technique, and we would hope that experimentation would continue to this arena. One size will not fit all problems.

Probably the single most important question in the development of analytical techniques is whether or not one can use scaling and aggregation to convert discrete events into a time series of interval-level variables. If this can be done, then a panoply of sophisticated, familiar, and pre-programmed econometric techniques can be used. In Chapter 4 we employed scaling and generated very satisfactory results, and a number of other projects (for example Goldstein and Freeman 1990; Goldstein and Pevehouse 1997) have also used scaled time series to good effect.

However, events are fundamentally discrete, not continuous, and any scaling and aggregation methods inserts a set of arbitrary parameters—the scale and the interval of aggregation—between the data and the statistical model. Hence the appeal of discrete sequence methods. The disadvantage of these methods is that they are not nearly as well understood as the techniques for studying interval-level time series—by anyone, to say nothing of the political science community. The Cowles Commission for discrete sequences has not occurred, though work in genetic and linguistic sequences has provided some guidance. Discrete sequence techniques are appealing, but in developing them a researcher is pretty much on his or her own, at least for the moment.

The final thing that should be kept in mind is that the appropriate analytical method may depend substantially on the problem being studied. For example, scaling methods may work better on problems characterized by a relatively clear conflict-cooperation continuum than on diplomatic and economic problems characterized by multiple bargaining and signaling strategies. Multi-actor systems may require different approaches than systems dominated by two actors. Our sense is that considerably more messy experimentation will be required to resolve this issue.

### 7.1.3. Event data work fairly well for classifying modes of political behavior

In Chapter 5 we made a strong case for the importance of analogical reasoning in foreign policy decision-making, and argued for the development of computer techniques that could simulate this pattern recognition in sets of event data. Because of the differences between the

structure—and capacity—of human memory and computer memory, we can only approximate this process, and most of the distinctions that we have tried to model were simple, but the results were certainly promising.

We do not expect that computer models will replace human analysts in the study of complex political activities. In addition to an associative memory, the human analyst brings to this task a crucial attribute that the computer will never have: she is human, and politics is a human activity! A machine, no matter how sophisticated, cannot get into this loop.

But machine-generated analogies can serve many other functions. The literal character of a computer model can alert an analyst to similarities or differences in a situation that he may not have seen due to cognitive biases and preconceptions. If the analogies produced by a model and a human analyst differ, this can often be a sign that some important variables are not being adequately considered by one or the other.

While a computer model is at a disadvantage at generating obscure analogies—for example comparing a contemporary political situation to Nuur al-Din's consolidation of control in Damascus in 1150—it has the advantage of not forgetting anything that it has learned. Human analysts tend to satisfice when seeking analogies; three or four good examples are usually sufficient to make the point. A computer model, in contrast, will continue to slog away through the entire data set, and possibly bring out a few analogies everyone wishes that they could forget.

Finally, computer models may be very useful in doing background monitoring, an otherwise very tedious task. While no event data systems (at least in the academic community) currently generate true real-time data, doing this is completely feasible and involves nothing more than integrating a assortment of existing programs that are presently discrete. With an integrated coding and analysis system in place, access to an appropriate set of codeable news reports, and the identification of an appropriate set of analogies that an analyst is concerned about, computer

monitoring could be used for a variety of background tasks that currently depend on human analogical reasoning.[2]

### 7.1.4. Event data can be used to generate early warning indicators for at least some conflicts

From the beginning, one of the primary motivations between the collection and analysis of event data has been the development of early warning indicators. Many of the early efforts at this were not very successful, but they were also very primitive by contemporary standards. The general topic of early warning will be discussed in detail in section 6.3, but in this section we will address the specifics of our studies of the Middle East, 1979-1997.

Based on our work on the Middle East, we think there is reason of some optimism that contemporary methods can improve considerably on the earlier efforts. The machine-coding of event data eliminates two problems that may have compromised the early warning models that used human-coded data: the hindsight bias of the coders (which would tend to exaggerate the early warning signal in sequences where the outcome was already known) and temporal stability (where different coders, rather than changes in activity, can introduce distinct patterns in the data). The computer power that can be thrown against this problem has also increased exponentially over the past three decades, and the most effective methods we have developed would have been almost impossible to implement when event data research began.

[2] When assessing the potential of this application, it is important to remember that news reports, cable traffic or whatever have already had an analogical filter applied by the reporters and editors generating those reports. A computer model does not need to be able to make inferences from a totally amorphous set of reports—if there is an armed uprising in some provincial villages, this will be reported as an "armed uprising" (or some equivalent phrase), and that report will look a lot different that a report of a bunch of people shooting off firearms to celebrate the new year. This extensive "prefiltering" makes the analogy problem considerably easier than it might first appear.

That information environment may, in fact, allow non-government efforts to match, or possibly even surpass, those of governments. Philip Zimmerman—author of the widely used *Pretty Good Privacy* encryption software—observes that in the field of cryptography:

> [the expertise] in academia and the private sector has grown and matured to reach parity with military expertise in the field. ... the algorithim [of the U.S. National Security Agency's secret "Clipper" chip] was found to be less conservative, with a smaller margin of safety, than the best designs from academia. It appears that cryptography—like the Internet itself—has stepped from the dark shadows of the military into the bright sunshine of the free market (Zimmerman 1998:115)

Many of the same factors that have driven the private-sector revolution in cryptography—most notably the widespread availability of computing power and the global sharing of information outside the control of national intelligence agencies—apply equally to the field of early warning.

We have focused our early warning efforts on the Middle East because we are very familiar with that situation. Whether or not these results can be generalized to other parts of the world remains an open question. On the one hand, the politics of the region—particularly the Israel-Lebanon-Palestine nexus—are very complex and involve a variety of state and non-state actors. In this respect the Middle East is more typical of post-Cold War early warning problems than the earlier event data studies of superpower relations.

However, the Middle East is atypical in two respects. First, it is closely monitored by the international media, and one cannot expect to find this same level of detail about Central Asia or Africa. Second, as we discussed in Chapter 6, the conflicts are highly institutionalized, and many of the key decision-makers have been in place for thirty or more years. Both of these characteristics may make the Middle East a good testbed for the *development* of early warning models, but those models might be more difficult to implement in other regions (particularly with news wire sources).

## 7.2. Technological Trends

The techniques described in this book are dependent at a very fundamental level on the availability of computing power and large amounts of data. Designing an analytical technique that requires computing power beyond that feasible with current technology is quite straightforward and these limits tend to be encountered relatively quickly when dealing with a complex phenomenon such as political behavior. Fortunately, the availability of computing power, data and software has improved dramatically over the past decade, and further major improvements are close at hand.

### 7.2.1. Computers

Since the early 1960s, computing power has always been available to an elite, although the identity of that elite was, often as not, determined by the United States Department of Defense. Prior to the 1980s, the rest of us had to depend on the campus computer facilities whose effective computing power would have been stressed running a low-level word processor. This restricted what could be done with these machines, as did the limited choice of operating environments and the lack of storage space for large data sets.

At the turn of the century, this situation had changed substantially: Table 7.1 compares the capability of an advanced personal computer that might be purchased by a average computer-oriented social science researcher in 1980 and 2000.[3] In fifteen years, speed has increased by

[3] A typical social scientist doing research in 1980, of course, would have used a shared mainframe rather than a personal computer, but these rationed mainframe resources were woefully inadequate for computationally-intensive modeling.

Superior computing environments were available in the AI labs funded by the Defense Advanced Research Projects Agency (DARPA), but those labs produced no international relations research of note and are thus largely irrelevant to the social science enterprise. Until recently DARPA has been responsible for much of the development of computer infrastructure in the United States but, constrained by a Congressional mandate from working in areas of public policy, DARPA sponsored very little international relations research with the exception of its experiment in event data analysis.

more than two orders of magnitude, random-access memory by three orders, and disk capacity by more than four orders, while the price/salary ratio has *dropped* by half an order of magnitude.

**Table 7.1. Comparison of Hardware**

| | **Apple II** | **Macintosh G4** |
|---|---|---|
| Year purchased | 1980 | 2000 |
| Processor speed | 1 Mhz | 450 Mhz |
| Data bandwidth | 8-bit | 32-bit |
| Random access memory | 64 Kb | 128 Mb |
| Disk storage | 320 Kb floppy disk | 12 Gb hard disk |
| Display | black and white | color |
| Time required to compile 1000-line Pascalprogram | 2 minutes | less than 1 second |
| Cost as percentage of assistant professor's salary | $5000/$15000 = 33% | $2000/$40000 = 5% |

Inexpensive desk-top computers now surpass in speed and memory most of the computers available to national intelligence agencies until the mid-1980s. This is further leveraged by interfaces and programming environments that enable an individual to manage far larger projects than would have been possible earlier. These machines can process text-based electronic communications of news organizations, IGOs, and NGOs directly without the delays caused by labor-intensive human event coding. The growth in personal computing power shows no signs of stopping, and within the next ten years the power of a desktop machine will probably surpass that now associated with supercomputers.

Whether this massive quantity of information can be effectively *analyzed* is another issue—this is the crux of the early warning challenge—but a researcher working with public domain sources in the late 1990s has access to dramatically more real-time information and data processing capability than she or he would have had available even a decade earlier.Increased speed, memory and storage will not, by themselves, solve all of the problems of modeling

sequences of political behavior: one cannot bludgeon all problems into submission with faster hardware.[4] Given that researchers have finite amounts of time, experiments that were impractical when a program required ten hours to run become practical when that same program requires two minutes. Because sequence analysis methods are information intensive, the increased disk and memory capacity is also very important.

### 7.2.2. Data Sources

Improvements in communication and computer technologies have changed dramatically the quantity and timeliness of the information available for use in event data analysis and early warning. Material relevant to political early warning is available from the commercial efforts of Reuters, *Agence France Press*, and other news agencies, and from the equally vast, if more specialized, networks of intergovernmental and nongovernmental organization (IGO and NGO) fieldworkers such as the UN Department of Humanitarian Affairs "ReliefWeb" site (King 1996; http://www.reliefweb.int). The Internet and data providers such as NEXIS provide a quantity of real-time information far exceeding that available to the Central Intelligence Agency (CIA) and KGB during most of the Cold War period.[5] While most of the machine-coding efforts to date have focused on newswire reports, texts dealing with treaties, parliamentary debates, NGO position papers and administrative rules are also readily available.

The question remains as to whether open sources are sufficient to predict international events. Whaley notes:

> The international and major independent news media are, in a practical sense, intelligence services, and they are so used by most persons regardless of their station in the polity or their access to conventional intelligence sources. Indeed, the "prestige" newspapers are

[4] Although one can bludgeon *some* problems into submission: In the computer chess competitions, the machines that attempted to play chess using human-like information processing—e.g. pattern recognition—were eventually outpaced by specialized chess-playing hardware that used simple search techniques and extremely fast parallel processing. Bad for theory, but whoever had the most CPU cycles won.

[5] This is particularly true when one focuses on strategic political intelligence to the exclusion of the tactical military intelligence provided by satellite imagery and the monitoring of electronic communication.

> known to draw readers from a far higher proportion of senior government officials than of the literate adult citizens in general...
>
> International news media resemble professional intelligence services in both function and structure. They ferret out, collect, collate, evaluate, analyze, summarize and report vast quantities of information, and they do so with an organization comprising correspondents in the field with their local networks and stringers. (Whaley 1973, 238-239)

Political and technological changes since Whaley made this statement—the end of the Cold War, advances in telecommunications—have further enhanced these capabilities. International business elites use the news media to obtain the same information that foreign policy decision-makers use. News agencies such as Reuters know this and their economic well-being depends on their ability to provide appropriate information. Whether this information is in fact sufficient to predict international events is an empirical issue, but the incentives to provide it are there.

Imagine that academic researchers had access to a machine-readable copy of Reuters with the same ease that they can access human-readable copies of *The New York Times* (the traditional source of event data). Any decent research library has copies of *The Times* on microfilm. If we want students to examine events prior to the outbreak of the 1967 Middle East War, we just send them to the library and they read the relevant account at no cost. Human-coded event data can take advantage of this accessibility, but human coding is extremely slow and expensive

If Reuters were similarly accessible, we would simply download the appropriate texts from the library via a high-speed communications link, reformat them, and code them. Under those circumstances, there would be very little reason for standardized event data sets: Because machine coding is completely reproducible, archiving dictionaries and search commands would be equivalent to archiving the original data, just as a research project that has constructed a scale from the American National Election Survey data reports only the questions used in the scale, not the transformed data.

This hypothetical situation would also lead to much greater experimentation with event data coding schemes, a development that we believe is long overdue. Most research designs that employ event data focus on only a small subset of political behaviors among a relatively small set

of states. When dictionaries of verb phrases are available, it is relatively easy to develop a new coding scheme because it is much easier to assign a code to a known verb phrase than it is to anticipate that the verb phrase will occur. As we noted earlier, *.actors* dictionaries need to be supplemented with the names of regional actors, and internal political activity is frequently rather idiosyncratic.

That is our ideal scenario: customized, regionally-specific, machine-coded event data. Unfortunately, the current situation is far from this ideal. Reuters is not available cost-free from the library, but only through relatively scarce NEXIS subscriptions and through a very slow downloading process. For the past nine months, current reports from Reuters have not been available from NEXIS at all. From the perspective of machine coding, the books are still locked up in the monastery libraries, chained to the desks, and copied by hand onto expensive parchment.

This situation is presumably temporary. Archives of most news periodicals (although not Reuters) are already available electronically through services such as NEXIS and, increasingly, on the World Wide Web.[6] Once in electronic form, the cost of moving those archives to a high-density archival medium such as DVD is minimal, and someone will presumably find a way to profitably sell these archives to libraries, just as microfilm is currently sold.[7]

---

[6] The availability of news reports on the web suggests another future model: automated web "spiders" that would continually search web-based sources of news reports, summarize and rewrite these using a standard vocabulary and syntax, eliminate redundant reports, and archive the results for future analysis. If this were done as a cooperative venture, much as large scale social science surveys are currently done, it would provide the missing link of a large corpus of reports on political activity unencumbered by copyright. All of the required components for such a system exist (or could be readily developed), and there is probably sufficient Web coverage of the industrialized states (but not the Third World) to provide data at least as good as what we currently get from Reuters.

[7] The economics of the NEXIS archives also make very little sense in the contemporary computing environment. The NEXIS system of storing files on a central server was economically rational when storage and processing power were sufficiently expensive to justify the time, bandwidth, and hardware dedicated to a NEXIS connection. But with the dramatic decrease in the cost of storage, it makes little sense to devote hours of low-speed connection time to transferring data that would fit on a $1 CD-ROM. It also makes little sense to process

It would be very nice if this occurred with Reuters, but we are not holding our collective breath—Reuters has become notorious in the world of information services for resisting innovation. More likely we will find that by the year 2010, a more up-to-date company such as CNN, AFP, Bloomberg or even Microsoft will fill this niche, and provide full-text sources going back to 1990 or so. But until that point is reached, machine-coding will be constrained by the availability of source texts, so please make your coded data sets available on the Web!

The appropriateness of any event data analysis ultimately depends on the theoretical model to which those events will be applied. Because of the labor involved in the collection and coding of the early event data sets, much effort was spent in carefully constructing and implementing coding schemes, and training coders, so that high intercoder reliability could be achieved. With the advent of inexpensive machine-readable text sources and machine coding, the cost of generating event data has dropped dramatically and considerably greater attention can be devoted to developing coding schemes that correspond to contemporary theoretical approaches.

### 7.2.3. Software

Computer programming is very labor-intensive and likely to remain so—it is a craft akin to carpentry rather than a repetitive task akin to the manufacture of pins—and this presents one of the major constraints on wider use of novel techniques for sequence analysis. Unlike pins or carpentry, however, the marginal cost of *reproducing* an existing program for use at another research site is arbitrarily close to zero. Consequently, greater availability of software could have a substantial influence on the field, just as the availability of packaged statistical programs such as SPSS, SAS and BMDP greatly accelerated the use of statistical methods during the 1970s.

While methods such as political sequence analysis are likely to always remain idiosyncratic, there are a number of other areas related to event data analysis that may benefit from the

---

those archives centrally on a NEXIS mainframe when a desktop computer could accomplish the task as almost as easily. At some point, these technological changes will (presumably) cause a change in the attractiveness of NEXIS, although apparently NEXIS has anachronistic but exclusive contracts on many of its data sources. Meanwhile, we *love* those educational subscription rates...

development of related commercial software and shareware. Software for natural language processing has begun to reach the commercial market in the past two or three years. Two types of programs may be of particular interest: automated indexing software, and natural language translation (e.g. English/Spanish) software. The indexing software is useful for extracting networks of relationships in a large body of text; the translation software is useful because it must deal with such problems as disambiguation, idioms, and subject-object identification; these same problems are critical in coding events and rules from natural language text. While these programs are still relatively primitive—probably about at the level of word processors in the early days of WordStar—they meet substantial needs in the business sector and are likely to experience continued development in the next decade. Accompanying the software will be the development of linguistic databases such as dictionaries, thesauri and semantic networks that could be used in specialized programs to code more complex data structures than the simple event data we currently code.

As we noted in Chapter 2, KEDS was the first attempt to create a robust system for converting newswire reports into event data. It was never intended to be the final or even the definitive work on the subject, and it contains obvious weaknesses in such areas as disambiguation of parts of speech, assessment of attribution, and correct handling of grammatically complicated sentences. Substantial progress as been made in the natural language processing literature on all of these issues since the basic design of KEDS was laid down in 1991, and we also now have substantially greater knowledge about the strengths and weaknesses of a machine coding system that is dealing with a very large corpus of unedited text that has been generated over a period of decades. We look forward to seeing the next generation of machine coding systems that are built upon this base of knowledge and experience.[8]

[8] But at the present time we *don't* intend to write one...

# 7.3. Early Warning

In recent years the topic of early warning has received renewed attention in the international relations literature (Rupesinghe and Kuroda 1992; Gurr and Harff 1994; Gurr and Harff 1996). With the end of the perceived threat of Communist exploitation of ethnic divisions, the liberal-democratic military powers—the United States, Britain, and France—are less inclined to intervene unilaterally in local or regional disputes. The international community has instead increasingly relied on multilateral responses, including the recycling of Cold War organizations (NATO in the former Yugoslavia and the United Nations generally), *ad hoc* initiatives (Iraq-Kuwait, Rwanda, Bosnia), and the use of existing non-military organizations in a peace-keeping role (ECOWAS in Liberia).

This dependence on multilateral responses enhances the attractiveness of early warning in two ways. First, there is general agreement (Cahill 1996; Crocker and Hampson 1996; Lund 1996; Schmeidl 1997) that a conflict in its early stages can often be contained by either limited force or diplomacy backed with the threat of force or other international sanctions. Second, multilateral actions require substantially longer to orchestrate than did the rapid responses of a superpower or a Cold War alliance. This has led to significant interest by international organizations in early warning (e.g., Boutros-Ghali 1992; Dedring 1994; Alker, Gurr, and Rupesinghe 1995; Mizuno 1995).

Despite all of this attention, early warning remains a difficult problem, whether done with quantitative or qualitative methods. For example, notwithstanding funding in the billions of dollars, access to a wide variety of information sources and a clear focus on a single opponent, Western intelligence agencies failed to anticipate both the timing and characteristics of the collapse of the Warsaw Pact. Early warning is almost nonexistent in low-priority areas such as Somalia, Rwanda, and Sierra Leone. In some of these cases, as we will argue below, early warning may be impossible for theoretical reasons. However, there are other cases where advances in communications and

analytical techniques should make possible the development of indicators that would not have been feasible when quantitative early warning research began thirty years ago.

Statistical early warning is often in the position of the story of the talking dog who can't get the punch line of a story right ("Damn dog never could tell a joke")—if we focus on the punch line, we miss the fact that the dog can talk. Instead, one should take a "bottom-up" approach of solving the cases that look easy first—and relegating these to a low-level monitoring—and classifying some other cases as unsolvable with the existing technology (not necessarily permanently unsolvable—as a sufficiently complex system could anticipate strategic moves as well—but we aren't at that point at the moment.

In the present state of our knowledge, it is not clear whether statistical forecasting—using this or some other method—will be superior to human forecasting. One possibility is that human forecasting will be better for some types of conflicts—for example those that are linked to relatively obvious ideological or economic changes that do not manifest themselves in discrete events—and worse in others, such as situations that can be predicted only by monitoring a fairly diffuse set of indicators consistently over a long period of time.

In an era of modern communications, with its accompanying increase in both the information available to human analysts and the analysts understanding of each other's cultures, human political forecasting has scored some impressive victories, most notably the 40-year balancing of the military standoff of the Cold War. But at the same time, the list of major changes over the past twenty years that were not anticipated in sufficient time for the international community to respond is significant:

- stability of the Islamic Republic of Iran
- Soviet invasion of Afghanistan and the subsequent failure of that effort
- collapse of the Warsaw Pact
- economic disintegration of the USSR (but the *pattern* of dissolution was anticipated correctly)
- civil war in Yugoslavia

- civil war in the Great Lakes
- structure of political power in Somalia
- Iraq's invasion of Kuwait

and our favorite:

- the Palestinian *intifada*

which was anticipated neither by the Israeli occupation forces nor by the leadership of the PLO. There is plenty of room for improvement here.

We are not suggesting that a single event-based would take care of all of these. However, a set of models, having different theoretical basis and in all likelihood distinctly different coding systems (which are simple to implement with machine coding), but having in common the use of inexpensive, open-source material such as Reuters and analytical methods designed to deal with nominal irregular time series might be able to predict a large class of these behaviors.

Before looking at the issue of the relationship between quantitative and qualitative early warning, a couple of remarks are in order on the relationship between our clustering approach and other forms of quantitative early warning. First, we regard the structural and the dynamic approaches as complementary rather than competitive. Structural methods are particularly good for mid-level warning: telling analysts where to look for potential trouble. Structural methods are also more likely to provide theoretical guidance as to why a system is likely to experience problems, which might provide insights as to the types of actions that could be taken to ameliorate an impending crisis. Structural models are unlikely ever to excel at predicting the exact timing of breakdowns—the variables that they have identified as theoretically important change too slowly—and this is where dynamic models come into play. The relationship of the approaches, then, may be something like the situation illustrated in Figure 2.

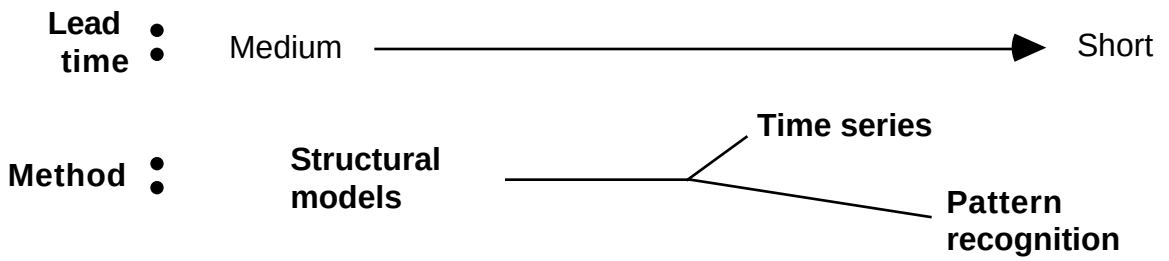


Figure 2. Time horizon for early warning methods.

We suspect that the ideal early warning model would combine elements of both the structural and dynamic approaches: It should be possible to refine dynamic early warning models based on different categories of structural precursors. Presumably the internal breakdown in a Lebanon—which is relatively wealthy and highly differentiated by religion—occurs in a different fashion than a breakdown in Rwanda, which is relatively poor and not differentiated by religion. The reason that such integrated models have not been developed to date is largely one of resources: the political science discipline is still in the process of developing accurate structural models and accurate dynamic models, and at present no researcher has been able to assemble data sets sufficiently large to study both simultaneously. As the research on both types of models identifies more focused sets of variables and techniques, it should be practical to combine the approaches.

Andre Goodfriend[9] echoes this concern in arguing that the UN system has a greater need for decision-support system that focus on assessing the likely effectiveness of various policy options, rather than early warning per se. He quotes a 1995 UN Joint Inspection report (JIU/REP/95/13-A/50/853) that is quite pessimistic in its assessment of early warning:

> In the Inspectors' view, independent analyses/early warnings, no matter how useful they are to serve their own purposes, do not facilitate concerted and effective preventive actions of the United Nations system on the basis of a common understanding of the problems.

[9]Posting to the EWNET-L list, 10 June 1998.

The problems of *decision-support*—evaluating the general future consequences of a specific sets of current actions—and *early warning*—evaluating the likelihood of a specific future event based on a general set of current circumstances—are obviously related but not identical. Both involve prediction over time, and both types of systems probably require similar sets of information. Decision-support requires methods that can deal robustly with counter-factuals: the nature of a decision-support system dictates that it will be projecting the future consequences of policies that are *not* implemented. The policies of interest, however, are known. Early warning systems, in contrast, must monitor a very large number of possible sources of political change, but only those sources that actually exist. The existing warning literature has primarily focused (obviously) on the second problem but might well pay more attention to the first.

Finally, some comments are in order on the relationship between quantitative, statistical methods of forecasting and the traditional qualitative, non-statistical methods. We regard statistical early warning indicators as a supplement to, rather than a replacement for, traditional qualitative early warning methods. Because political behavior is a human activity (in contrast, for example, to weather or earthquakes), human understanding and intuition are likely to be powerful tools in predicting that behavior. Early warning is also an "ill-defined problem" (Moray 1984, 11) within a complex system, where neither the relevant variables nor the relevant processes have been fully, or even adequately, identified. We also face the practical constraint that purely statistically-based warning systems are unlikely to be accepted in the qualitatively-oriented policy community (Laurance 1990).

At the same time, statistically-based forecasting methods fill two gaps that are inherent in human-based qualitative approaches. First, while human intuition is a valuable tool in understanding political behavior, cognitive biases can blind an analyst to a situation that is rapidly changing despite his or her expectations to the contrary. Second, statistically-based methods are capable of consistently monitoring a much larger amount of information than a human analyst can monitor. A system based on computerized analysis of machine-readable

sources can monitor 24-hours-a-day without fatigue, distractions, political pressure or committee meetings.

The qualitative opportunities for acquiring information relevant to early warning has increased dramatically in the past five years with the availability of inexpensive machine-readable commercial news sources and the proliferation of reports available from IGOs and NGOs via the internet. During this same period the challenges have also increased, for example in the potential dissolution of some states in the post-Cold War period and the appalling resurgence of genocidal outbreaks such as those witnessed in Cambodia, Rwanda and Bosnia. Consequently we believe that there is an important role for the development of quantitative indicators. To the extent that an area is adequately monitored by electronically-readable sources, real-time quantitative forecasting using machine-coded event data is quite inexpensive and can easily operate in the background as a supplement to qualitative forecasting.

The appropriateness of any event data analysis ultimately depends on the theoretical model to which those events will be applied. Because of the labor involved in the collection and coding of the early event data sets, much effort was spent in carefully constructing and implementing coding schemes, and training coders, so that high intercoder reliability could be achieved. With the advent of inexpensive machine-readable text sources and machine coding, the cost of generating event data has dropped dramatically and considerably greater attention can be devoted to developing coding schemes that correspond to contemporary theoretical approaches.

## 7.4. Three Grand Challenges for Event Data Analysis

We close this volume with three suggestions for "grand challenge" projects that we believe could be done with existing—or soon to exist—theory and technology and which would significantly advance the field. None of these are easy—and we certainly would not recommend them as dissertation projects—but parts of these might be feasible for a single (tenured) researcher.

### 7.4.1. Identification of changes in policy due to adaptive behavior

In Chapter 6 we argued that foreign policy is adaptive, and that at certain points in time organizations make fundamental changes in their rules that profoundly affect subsequent behavior in the system. In some cases, these changes in policy are explicit and have almost immediate effects, for example radical changes in government or the US renunciation of the gold standard in 1971. In other cases, however, the actual policy change is kept secret for an extended period of time before it is revealed: examples would include Nixon's rapprochement with China, the abandonment of the Brezhnev Doctrine, and the initiation of direct negotiations between Israel and the Palestine Liberation Organization.

Detecting such changes in the censored and noisy environment of international events will unquestionably be a difficult task, particularly when the changes are accompanied by extensive efforts to maintain secrecy. However, precisely because these efforts at secrecy are directed at human analysts, some forms of computational analysis may be more amenable to detecting the relevant changes because the computer is unaffected by human cognitive preconceptions. A computer is a difference detector, the human brain is a similarity detector.

Constructing such a model would involve monitoring two things. First, one would need some measure of the success or appropriateness of the existing policy that would indicate whether a policy change was likely. Second, one would need to monitor the observed behaviors to see whether those behaviors have departed from earlier norms—for example have new, unprecedented event sequences have begun to appear? To date, most early warning methods have focused only on the second part of the problem, and these methods could be significantly enhanced by models that show how routine learning occurs.

More generally, if it were possible to establish the regularity of certain sequenced behaviors, then—under appropriate assumptions about payoffs—it might also be possible to predict a set of likely adaptations to those behaviors. Taken to the limit, this just gives a game theoretic mixed strategy, but most organizations are insufficiently flexible to take it to the limit. This is similar to the approach used in developing influenza vaccines, where the vaccine is based on the most

probable mutation from the previous year's virus. Sometimes they get it right; sometimes they don't. But in order to solve the problem of adaptation, one must first solve the problem of prediction.

Sequence analysis, in other words, would fit in the middle of a hierarchy of decision environments listed in Table 7.2. When an environment is relatively static in terms of payoffs and alternatives, it can be analyzed with the traditional tools of expected utility and game theory. A highly dynamic environment characterized by continually changing strategies that are generated by rational adaptation, in contrast, is essentially (and possibly formally) chaotic—the nearly continuous changes of financial markets may be example of this. But a middle ground exists where payoffs change, and strategies are modified, only gradually because of economic and organizational constraints, and this is where the precedent-based modeling of sequence analysis comes into play. As we have argued throughout this volume, slow-but-rational change characterizes much of the decision environment found in international affairs, and consequently sequence analysis methods are particularly relevant to this domain.

**Table 7.2. Models of decision-making**

| Decision environment | Type of model |
|---|---|
| Static utility maximization | Expected utility; game theory |
| Co-adaptive SOPs | Sequence analysis |
| Rational adaptation | Chaos |

### 7.4.2. Chronology Generators[10]

Political analysts spend a lot of time with textual materials simply figuring out what events happened in the course of a political episode. Newswire sources such as Reuters completely mix

[10] Our thanks to Ed Laurance for suggesting this one.

topics: a story on Israeli-Syrian negotiations may be followed by one on rice production in Indonesia. A "chronology generator" would extract from a global news source only those stories relevant to particular policy (e.g. negotiations concerning Bosnia or U.S. policy regarding Somalia).[11] This project would be a basic technology that would enhance all of the others discussed here, and might also be useful for historians and journalists.

A basic system might involve building a network model of the actors involved in the issue, then identifying keywords associated with the issue. The chronology generator might borrow some of the statistical methods that have been developed for automatic indexing (see Salton 1989, Smith 1990); other elements would require political knowledge, including sequences generated earlier by the system itself. The generator would then pull out a large set of possibly relevant articles from the text base, edit out the duplicates and irrelevant material, do some sort of hypertext linking between the information in the articles (e.g., if an article mentions that a meeting is planned, the system would search to find a subsequent report about the meeting) and then concatenate these into a natural language or event-coded chronology.

A chronology generator would clearly benefit from a knowledge base of "political common sense" (e.g. knowing that an arms transfer requires a contract, payment, delivery; see discussion below). It might also be able to use some of the methods developed in the "qualitative reasoning" literature in artificial intelligence (see Iwasaki 1989). While the existing AI literature deals almost exclusively with reasoning about physical systems, it is concerned with issues of time, contingencies and with the possibility that multiple qualitative rules might be affecting the operation of a system. All of these problems have parallels in the analysis of political behavior.

[11] As noted earlier, extensive work has been done in AI on the use of scripts and story-generators in limited domains, but there is little evidence that the labor-intensive methods used in these systems can be scaled up to handle a large corpus of unedited text (Schank 1991). We already have labor-intensive methods for building chronologies in restricted domains from unedited text—they are called graduate students—and to be useful computational methods must provide some additional advantage.

### 7.4.3. Construction of organizational rules directly from documents and political "common sense"

This is without question the most technically challenging of the suggestions we are making, and in fact it may not be possible with current technology. We know that it is possible to create realistic rule-based models of decision-making organizations that account for over 50% of observed behavior (for example Tanaka 1984, Cimbala 1987, Sylvan, Goel, and Chandrasekran 1990, Taber 1992, Taber and Timpone 1994) and if a number of such models were available and could be made to interact, they would be quite useful in analyzing "what if" scenarios (e.g. if Carter had been re-elected in 1980, would the Cold War have ended in 1983?). However, at present the construction of these models is very labor-intensive because the rules need to be extracted using human coding.

It is clearly possible—although unfortunately also labor-intensive—to create systems for the machine coding of complex event structures in some limited substantive domains. For example Alvarado's OpEd program (1990) "understands" editorials on issues in political economy and can answer questions about them;[12] Kolodner's CYRUS (1984) deals with the official activities of Secretary of State Cyrus Vance; and Pazzani's OCCAM (1988) deals with international economic sanctions. DARPA's MUC project (Lehnert and Sundheim 1991) experimented with a variety of approaches for extracting specific pieces of information on Latin American terrorist incidents from newswire text. The performance of such systems was impressive, although they were dependent on a large amount of domain-specific knowledge and can only work with text dealing directly in that domain. It might be possible to make a more general system by using the experience of earlier cognitive-mapping projects (Axelrod 1976), and rule-extraction might also be simplified with the availability of a very large corpus of redundant text in machine-readable form—for example parliamentary testimony or government-sanctioned speeches and editorials.

[12] The source texts require some editing before processing, so it is not clear that the system would work with unedited newswire material.

For the reasons discussed above, it will still be necessary to add some ad hoc tacit knowledge to these models, although this ad hocery would be no greater than that currently used in the construction of human-coded rules. Supplementing the internal sources of rules with societal sources, particularly those directed to an external audience, might provide a data-driven way to partially get around the problem of tacit knowledge. For example, while Sylvan and Majeski (1985) found that opposition to communism was not mentioned in the internal debates within the State Department in the early 1960s, Voice of America broadcasts made no secret of this position.

In addition to explicit rules, tacit knowledge and precedent, a rule-based system will require a great deal of political "common sense." The CYC project (Lenat and Guha 1990) is based on the premise that once a sufficient amount of basic knowledge and natural language skills are provided to a computer, it will hit a take-off point where it can learn subsequent information in a fashion similar to a human: reading reference materials and integrating this new knowledge with what it already knows.[13]

Whether the CYC project will succeed or fail is still unclear, although the results should be known within a couple of years. If it succeeds, one might envision a similar project—on a much smaller, social science scale—for modeling political knowledge that uses the higher-level information acquisition and representation technologies of CYC, as well as some of its lower-level social knowledge. We have readily available basic books on political behavior, ranging from the simplified histories using in elementary schools to the social studies books of the secondary schools to college level textbooks. It might be possible to derive a fairly elaborate model of general political behavior from these, perhaps one capable of answering political questions at the level of a college sophomore.

---

[13]The CYC project is attempting to construct a set of common sense rules that will approximate the knowledge base of a five-year-old. The estimated size of this data base is 10-million rules (*Economist* 1992,13). In political analysis (as well as CYC), common sense is not limited to physical phenomena but also includes information on social norms, although these will obviously be culturally dependent.

If this proves possible, an interesting experiment would be to try learn the political knowledge taught in different cultures, for example that reflected in the primary school history textbooks used in the United States, Egypt, Brazil, Nigeria, Russia and Japan, to see what differences emerge. At this elementary level, where the most basic formal political sequences are first learned, we suspect one would find a significant cross-cultural divergence that could be systematically modeled.