PATTERNS, RULES AND LEARNING: COMPUTATIONAL MODELS OF INTERNATIONAL BEHAVIOR

(2nd Edition)

Philip A. Schrodt
University of Kansas

Version 2.0
April, 2004

Contact information (April 2004)
Philip A. Schrodt
Department of Political Science
University of Kansas
1541 Lilac Lane
Lawrence, KS 66045
USA

phone: +1-785-864-9024   fax: +1-785-864-5700
email: schrodt@ku.edu
professional web page: http://people.ku.edu/~schrodt
KEDS project web page: http://www.ku.edu/~keds

An electronic version of this manuscript can be downloaded from:
http://www.ku.edu/~keds/books.html

PDF File generated: 3-May-2006

Published by Parus Analytical Systems, Vinland, Kansas, USA

To John, Dorothy, David and Patty
where it all began.

Table of Contents

Acknowledgements

No idea is really taken seriously until it is written into a book
Betty Friedan

It has been an immense pleasure to participate over the past ten years in the development of a research paradigm involving the computational modeling of international politics. This has been, without question, a social endeavor rather than a solitary one, and the list of individuals to whom I owe intellectual debts is long. These people may not agree with my approach or methodologies—in fact many of their key contribution were disagreements—but all contributed in some sense, and some undoubtedly contributed in ways I do not remember but should have cited. In alphabetic order: Christopher Achen, Paul Anderson, Edward Azar, James Bennett, Scott Bennett, Jacob Bercovitch, Daniel Caldwell, Shannon Davis, Christopher Donald, Gavan Duffy, Ronald Francisco, John Freeman, Peter Frey, David Garson, Deborah Gerner, Steve Greffinius, Harold Guetzkow, Reid Hastie, David Heise, Charles Hermann, Margaret Hermann, Paul Johnson, Marita Kaw, Richard Kimber, Edward Laurance, Russell Leng, John Mallery, Benjamin Most, Anthony Nownes, Elinor Ostrom, Helen Purkitt, Frank Sherman, Harvey Starr, Arthur Stinchcombe, Charles Taber, Stuart Thorson, Sigrid Unseld, James Voss, Michael Ward, Mark Warren and Judith Weddle. I should particularly single out the individuals with whom I have shared many conversations over beer and the internet and who have influenced my entire approach to this issue, the "AI/IR Mafia" who, without even trying, always ended up on all the same ISA and APSA panels in the 1980s: Hayward Alker, Robert Boynton, Valerie Hudson, Brian Job, Stephen Majeski, Dwain Mefford, David Sylvan and Donald Sylvan. I would also thank Robert Boynton, Harold Guetzkow, William Mills and an anonymous University of ¥Ω'~£§#@ø&"[1] Press referee for their helpful—and remarkably consistent—criticisms on an earlier draft of the book. As usual, the responsibility for subsequent integration, permutation, recombination and misinterpretation of the ideas of any of these individuals remains my own.

This work uses two data sets provided through the Inter-university Consortium for Political and Social Research: The Behavioral Correlates of War data were originally collected by Russell J. Leng; the Interstate Security Conflicts data were originally collected by Robert L. Butterworth. I also use a draft version of the COPDAB data set provided by Edward Azar. Neither the original collectors nor the ICPSR bear any responsibility for the analyses or interpretations presented here.

The Midwest Consortium for International Security Studies, the NSF-funded Data Development in International Relations project, the NSF-funded Political Methodology summer conferences, and the Austrian Federal Ministry for Science and Research have greatly aided this research through the sponsorship of various conferences; these were facilitated in part by Richard Merritt, Marian Rice, Frank Scioli, Robert Trappl and Dina Zinnes, as well as by many of the individuals named earlier. The research was directly supported at times by National Science Foundation Grants SES89-10738 and SES90-

[1] Text tragically lost in translation—see next section.

25130 and by the University of Kansas General Research Allocation grant 3500-X0-0038.

I owe a great debt to my colleagues in the Department of Political Science at the University of Kansas for providing a supportive and collegial environment over the past seven years. Much of this text—as well as a great deal of material that justifiably did not survive to the final draft—was originally presented to students at classes in the ICPSR summer program and in international relations methodology classes at Kansas: the patience and feedback of these students is gratefully acknowledged.

Philip A. Schrodt
Lawrence, Kansas
October 1995

Preface to the Second Edition

"When I *use a word," Humpty Dumpty said in a rather scornful tone, "it means just what I choose it to mean—neither more nor less."*

Charles Dodgson (Lewis Carroll)
Through the Looking-Glass

Second edition?! This thing didn't even have a *first* edition!

Well, yes and no. The 1995 version of this manuscript has been around for a while and, frankly, I've done okay by it. It has been favorably cited in an *American Political Science Review* article (Beck, King, and Zeng 2000), and my understanding is that parts have been used in graduate classes at various institutions, though presumably not recently. Computer scientists seem to find it more useful than political scientists—make of that what you will—and every few months I get a phone call or email from some consultant about embark on a new contract applying computational methods to the analysis of political behavior who has found the manuscript on the Web and was wondering what had subsequently become of the enterprise (answer: see http://www.ku.edu/~keds).

What has not happened is the appearance of this text between two covers bearing the imprimatur of some august academic press. Wherein lies a sad but not all that interesting tale of academic life in the last decade of the 20th century: I wrote the manuscript over a period of years, it was sent out by a major academic press for comments by two reviewers, I then revised the manuscript to account for those comments, it was sent out again, and the reviewers still didn't like it.

At this point the standard operating procedure would have been to further "shop the manuscript down the food chain" until I found a press who would publish it. But at the time the revision was rejected, I was in the Middle East on a Fulbright grant, living in a war zone, and was dealing with an extended serious illness in my family. I'd had tenure for over a decade, published fifty or so mostly refereed articles, and was having good success securing external funding for my research. "Artificial intelligence in international relations" was clearly past its prime as an academic enterprise (another sad history discussed in the epilogue in Chapter 8) while the automated event data coding techniques of the Kansas Event Data System project were getting serious attention and competing for my limited time.

Most importantly, after all of those revisions, I was simply tired of the whole thing. So I eventually posted the text to the Political Methodology paper server, and went on to other pursuits. I occasionally imagined finding a time when I was free from more pressing obligations and could return to the manuscript to do yet another set of revisions, update the citations, and perhaps even re-do some of the analyses with computers that were 1,000 or more times more powerful than those I'd originally used. But somehow that elusive time "free from more pressing obligations" never occurred.

In 2004, I started work on a new project with Valerie Hudson (Hudson, Schrodt and Witmer 2004; http://www.nkss.org/)—another refugee from the original AI/IR group—and went back to borrow some material from the manuscript. Reading through the text I—as ever, the totally unbiased observer— concluded "Hey, this stuff isn't bad…" But I also concluded that it was so thoroughly a product of the mind-set of the late 1980s and early 1990s that it would be impossible to update, and perhaps—for the purposes of providing a window into that period—it should not be updated.

Two other factors also came into play at this point. The first was a gradual recognition—based on my own work and that of virtually everyone else doing research on any topic—that with the emergence of sophisticated search engines for the World Wide Web, there were major advantages in a manuscript *not* being instantiated as dead-trees-and-ink, as dead-trees-and-ink were quite notoriously inaccessible to Google and its successors.[1]

I had also noted the recent comments of Swiss economist Bruno Frey on the intellectual inefficiency—to phrase the issue more politely than it is characterized by Frey—of the academic publishing enterprise:[2]

> Many authors feel that the refereeing process robbed them of the chance to really contribute what they find important and innovative. An example is given by Brian Arthur, who states that "I put the paper ("Competing Technologies, Increasing Returns, and Lock-In by Historical Events", finally published in the *Economic Journal* 1989) through eight rewrites in this (revision) process; each time it became stiffer, more formal, less informative, and as a result more publishable" (Gans and Shepherd 1994, reprinted in Gans 2000: 35). Sometimes the papers published reflect more the referees' than the author's ideas. Robert Frank (personal communication 14 May 2002) provides a case in point: his paper (Frank 1987) was published in the *American Economic Review* and originally contained two parts. The first part contained what he really wanted to convey to the readers, and the second part was a formal appendix Frank himself did not find of much importance. One of the referees demanded the first part be deleted, and the (according to Frank clearly less interesting) appendix to essentially constitute the paper. Frank agreed, because he knew that this was the price to have the paper accepted by the *AER*. Many such stories can be heard in our profession.
>
> … [E]ven among extremely successful economists, crowned by the Nobel Prize, there are some who harshly criticize the existing journal publication system. Examples are Leontief 1971, Coase 1994 or Buchanan 2000; see more generally Leijonhuvfud 1973 and Cassidy 1996. … See the responses from 140 leading

---

[1] *Preservation* is different issue. Which medium will have the edge across the centuries?: The bulky mass of volumes of acid-free paper, accessible to anyone with a linguistically educated but otherwise unaided eye, or digital patterns preserved in exquisitely minute optical pits and magnetic domains, accessible only to advanced technology but stored at high densities and amenable to inexpensive reproduction and translation across systems and natural languages. I don't know the answer. What I can say—based on a survey of items sitting in my office—is that the text that clearly *will* survive is that which is deeply raised or embossed upon multiple pieces of nearly indestructible plastic, all carrying the identical ritual incantation: **MADE IN CHINA**.

[2] Frey, to deflect the inevitable accusations of "sour grapes", prudently prefaces his criticisms by noting

> I believe I have some experience and competence in this area. I have published more than 250 papers in over 140 refereed journals during the period 1965–2002. Among them are leading economics journals such as *AER, JPE, RES, REcsStats, EJ, JEcLit* and *JEcPersp*., but also in political science (e.g. *APSR*), psychology, law and sociology journals. I have written 16 books, served as one of the two (and later three) managing editors of *Kyklos* since 1970, am a member of the board of editors of 23 journals and over the years have served as referee for numerous journals. (Frey 2002: 6)

> economists about their journal submission experiences and the list of "classic" papers once, and often more than once, rejected (Gans and Shepherd 1994). A well-known example is Akerlof's "Market for Lemons", which was rejected by the *American Economic Review* and the *Review of Economic Studies* as being "trivial", and by the *Journal of Political Economy* for being "too general" before it was accepted by the *Quarterly Journal of Economics*, which was instrumental in him winning the Nobel Prize. (Frey 2002, 15, 22)

My manuscript is clearly not Nobel Prize material, but rejection by an editor and a couple of referees should perhaps not be taken as the final word on its utility.

Finally, the emergence of the PDF page description standard made it possible to convert the document into a readily accessible format, which had not been guaranteed by the Microsoft *Word* files of the earlier manuscript. More generally, a thorough reformatting of the document could well preserve a readable version of the text through another generation or two of software. I was on sabbatical in 2003-2004, and for a couple of months faced the prospect of considerable of free time on airplanes, so this seemed like a worthy investment of time and effort.

Hence the document that you have in front of you in some form. For the most part I have simply cleaned up the 1995 manuscript by standardizing fonts[3] and reformatting where necessary. The new version is physically formatted as a book, with single-spacing, justified text, and asymmetric margins that allow for binding if you are so inclined. It still has not had the attentions of a professional copy editor—which undoubtedly would improve it—and still lacks an index (but then you know how to use the "Search" function in the PDF reader, correct?). But otherwise it looks like a book.

So publishers be damned…

An epigraph—chapter 8—contains an update on where I think the field has gone since the work described in the first edition was completed (and this was, effectively, around 1994 except for some minor additional updating). In a small number of places, I've added explanatory footnotes for material that might otherwise make no sense: for example you young whippersnappers probably don't remember the days when the Apple Computer logo was multi-colored. The remainder of the manuscript is unchanged.

Philip A. Schrodt
Lawrence, Kansas
April 2004

[3] The entire manuscript uses only the monospaced Courier, sans-serif Helvetica, and serif Times New Roman fonts, plus the "Symbol" font from the Macintosh computer. Courier, Helvetica and Times New Roman are among the most widely used—and most easily substituted—fonts at the present time, and are likely to remain so in the future. The "Symbol" font is a bit dodgier but is used sparingly.

Chapter 1
Introduction and Overview

> *If, with a solemn feeling of the importance of things as they really are, we were to admit the irregularities of the actual world into the statement of our problems, we should of consequence have to attend to enormous elaborations of mathematics in the process of solution, whereby our attention would be for a long time distracted away from the actual world.*
>
> *Lewis F. Richardson*

> *Worrying about complications before ruling out the possibility that the answer was simple would have been damned foolishness. Linus Pauling never got anywhere by seeking out messes.*
>
> *James D. Watson*

In 1519 Hernando Cortes landed on the east coast of Mexico to confront a civilization arguably more advanced than his own. Neither Cortes nor any European had prior knowledge of the culture, languages or history of this system, yet within a couple of months, utilizing the techniques of Renaissance statecraft, Cortes assembled a successful military coalition based on Meso-American groups disaffected with the prevailing political and religious leadership in Tenochtitlan. That coalition, while ultimately disastrous for its non-European participants, was in the short run international politics as usual for both the Spanish and the Mexicans.

This volume presents a set of formal approaches for studying the regularities of international political behavior. It is theoretically based in studies of human cognition and organizational information processing; it is methodologically based in computational modeling. The first half of the work argues that political decision-making produces regularities in international behavior that have not been tapped by the existing statistical, dynamic or rational choice approaches to formal modeling. Among the processes generating those regularities are individual pattern-recognition, the organizational use of formal rules, and learning, by individuals and organizations, from history and experience. Those processes can be approximated using computational methods that are specified algorithmically rather than algebraically. The second half of the book illustrates the use of computational models to study a variety of different types of international behavior.

This approach is reflexive: to learn to model a social behavior, model how that behavior is learned. A model of international politics can be based on how international politics itself is understood, and hence organized, by its participants. My theoretical approach is presented in detail in Chapter 3 but its outline is simple: patterns, rules and learning:

ASSUMPTION 1: Individuals understand international events using pattern recognition

Human beings, possessing associative memories, are extraordinarily skillful at pattern recognition. This task they find not only relatively painless, but, to judge from the popularity of crossword puzzles, *Trivial Pursuits* and *Wheel of Fortune*, downright pleasurable. In contrast, humans are resistant to logical deductive reasoning and avoid using it whenever possible. In the absence of pattern recognition, the international system rarely provides sufficient information for the use of deductive reasoning.

To "understand" an international event is to fit it to a pattern one has previously learned through experience, formal education or acculturation. The most common patterns are sequences of events: history. These sequences come tagged with contextual information—for example, the knowledge that the United States will respond differently if attacked by Libya than Libya will respond if attacked by the United States. Patterns are often based on idealized versions of historical events—Munich, Pearl Harbor, Vietnam—and some are completely hypothetical, for example the crisis escalations leading to nuclear war.

Pattern recognition also accounts for the human ability to infer motive and to correct for missing, deceptive and erroneous information in an observed sequence of events. The ability of decision-makers to engage in meaningful behavior depends on such short-term predictions; without this information international politics would consist of random events rather than the highly regularized sequences actually observed.

ASSUMPTION 2: Organizations respond to events using rules.

International behavior is primarily the result of organizational decision-making. Organizations do not possess associative memories and therefore favor deductive reasoning based on *if...then* rules. Compared to the complex patterns accessed by the human brain, rules are simple, easily stored, and easily transmitted. Theses rules involve not only formal procedures, but heuristics, rules of thumb, and shared practices and expectations that allow an organization to link actions with outcomes.

Because of their orientation to rules, organizations are typically much less flexible in their information processing than are the individuals who comprise the organization. Organizations sacrifice flexibility for capacity, but they operate in situations where it is advantageous to do a very large number of relatively simple things. The fact that organizations are large and inflexible contributes to the predictability of international behavior—the United States does not decide daily whether to invade Canada or to praise Fidel Castro—even though the irregular sequence of behavior found in crises capture much of our attention.

ASSUMPTION 3: Foreign policy behavior is affected by continual learning and adaptation.

Information relevant to determining appropriate policies in international politics is complex, scarce, expensive and exceedingly noisy. Foreign policy is an ill-structured problem: optimization techniques that might work in the design of an bridge or the determination of price in a market are usually ineffective. While the international system may be equilibrium-*seeking*, it typically reacts slowly and incrementally, and the equilibria to which it is moving are be based on information years out of date. The international system is very different from a market.

In place of optimization, the international system uses adaptation. Organizations do not make the same major, obvious mistake twice; successes reinforce the patterns and rules that generated them. The behaviors of a complex organization are thus grown rather than designed. Memory is found in the minds of its practitioners and the rules of its organizations. Organizational memory is imperfect and systematically rather than randomly distorted, but it exerts a profound influence over organizational behavior.

From these assumptions, I conclude that the international system, despite its quasi-anarchic character, self-organizes and exhibits regular behavior that can be mimicked, to an extent, by computational models.

Models

The primary motivation behind this work is the development of formal models.[1] These models are a simplification in international behavior, but it is the job of models to simplify. The historian and the journalist can deal with the complications of events in all their detail; the task of the social scientist is to find some means of simplifying that torrent of information. Ashby observes:

> …every model of a real system is in one sense second rate. Nothing can exceed, or even equal, the truth and accuracy of the real system itself. Every model is inferior, a distortion, a lie.
>
> No electronic model of a cat's brain can possibly be as true as that provided by another cat, yet of what *use* is the latter as a model? Its very closeness means that it also presents all the technical features that make the first so difficult [to study]. (quoted in Fiorina 1973,136)

To the extent that a formal model can remain empirically accurate while simplifying, it has utility. Computational modeling dictates a particular set of compromises; correlational statistics, differential equations, expected utility or two-person games would dictate others, as would an emphasis in a traditional theory on power or on economic inequality. The choice of any tool or approach closes some doors while opening others.

My efforts are addressed primarily to a behavioralist audience and proceed from behavioralist premises such as the desirability of unambiguous specification of concepts, utility of formalism, the necessity of replicable empirical tests, and so forth. However, my approach departs from the simple equations, borrowed from the physical sciences that characterize much of formal modeling in political science. There is a vast difference between studying falling rocks and studying politics, for rocks do not alter their behavior to suit prevailing theories. Human decision-makers convinced of the legitimacy of theories of nuclear deterrence will proceed to act in accordance with those theories; nuclear weapons, in contrast, are by all indications completely indifferent to the theories concerning their use. My general philosophical approach is closer to that of Weber than Newton, Marx or Pareto:

> On the other hand Max Weber viewed the notion of a social science which would consist of "a closed system of concepts, in which reality is synthesized in some sort of permanently and universally valid classification, and from which it can again be deduced" as entirely meaningless. … [For Weber,] the subject matter of the social sciences—human action—involves value orientation, memory and learning, which can only yield "soft" regularities, "objective possibilities" and probabilities. (Almond 1988,837)

In a computational model, the Weberian elements of values, memory and learning can be introduced, but at the expense of parsimony. Computational models are far more

---

[1] I will be using the concepts of a model and a formal empirical methodology interchangeably, though political science tends to treat them separately. A regression equation is a model—in fact when fully explored in the context of probability theory it is an extremely complex model—even if we treat it merely as a "technique" to be chunked through using SAS.

complex than differential equations or expected utility calculations, though not much more complex than a large global model or the specification of the maximum likelihood estimators for a large system of equations. The simplicity of a computational model lies in its postulated *processes*; these models also use a great deal of domain-specific information because human decision-makers use a great deal of domain-specific information. However, because computational models are specified in an unambiguous form, the process by which this information is employed can be replicated and studied.

A variety of arguments can be made for the intrinsic value of formal models but, briefly stated, natural language is not a particularly good medium for developing complex, logically consistent arguments when compared to "formal" languages such as mathematics.[2] McCloskey notes

> The economic conversation has heard much eloquent talk, but its most eloquent passages have been mathematical. ... The *American Economic Review* of the early 1930s, by contrast, contained hardly an equation; assumptions were not formalized... The consequences of the primitive machinery for conversation was an inability to speak clearly. Economists could not keep clear, for instance, the difference between a movement of an entire curve and a movement along a curve. Being mathematically innocent, they were unable to talk in curvy metaphors. ...Economists before the reception of mathematics fell headlong ... into confusions that a little mathematics would have cleared up. (McCloskey 1985,3)

Natural language can be very ambiguous and a theory may appear to be successful because it can be warped to fit any situation: "Balance of power" is the obvious example here. Richardson noted:

> Another advantage of a mathematical statement is that it is so definite that it might be definitely wrong... Some verbal statements have not this merit; they are so vague that they could hardly be wrong, and are correspondingly useless. (Newman 1956,1248)

Formalization is as fundamental to the development of human knowledge as writing or music, and attempts to apply formal methods to the study of human behavior date from the Enlightenment in the efforts of Leibniz, Laplace, Condorcet and others. In some studies of human behavior—economics, public health, demographics, election forecasting—these formal approaches have become the dominant approach; in the study of international behavior they have been less successful to date. But the goal is still worth pursuing.

Rationality

For reasons that will be discussed in considerable detail in Chapters 2 and 3, this work de-emphasizes the role of rationality as it is currently understood in the formal modeling literature. There are at least three reasons for this. First, I will argue that pattern recognition is a necessary *prerequisite* for any self-contained "rational" model of political behavior. Unlike a market, the international system rarely provides reliable information about the consequences of ones actions, or even on many of the critical

---

[2] Archibald and Lipsey (1976,1-10) provide a particularly good rendition.

variables reflecting the state of the system; this information must instead be inferred. Any model that presupposes this information—and that is the case with most rational choice analyses of international politics—is as much a model of the modeler as it is a model of the system.

Second, I am concerned about the disjuncture between the assumptions of the rational choice approach and the observed behavior of organizations. This is scarcely a novel criticism—most of these arguments were elucidated three decades ago by Simon, March and others, and the applicability of the rational choice model to individual decision-making has been under empirical attack from cognitive psychologists such as Kahneman and Tversky for the last decade and a half. This debate often rages at a near-theological level, and I make no claims of contributing to its resolution, but it will become quickly apparent that I'm more convinced by the critics of rational choice than by its proponents.

More generally, I am concerned by the failure of many proponents of the rational choice approach to subject their models to large-sample empirical tests. As noted in Chapter 2, this situation has improved in recent years, particularly in the work of Bueno de Mesquita and his co-authors, and is obviously justified when dealing with counterfactuals such as the failure of nuclear deterrence. It is not justified when dealing with models of conventional war initiation, crisis, arms control or balance of power. The relevant data are available or could be collected; the requisite statistical tests are available or could be developed.

Given my de-emphasis on the rational choice approach, two caveats are in order (and will be further elaborated in Chapters 2 and 3). First, my emphasis on the intuitive and the routine in international affairs is not meant to imply that rigorous deductive thinking plays *no* part in foreign policy decision-making. To the contrary, the examination of international affairs in any century will reveal a great deal of cleverness—for good and evil—and the issues of international politics have without question attracted some of the greatest political minds of each era. I am arguing, however, that these plans occur against a *background* of regularities that are, for the most part, determined by intuitive understanding and the routine application of rules. While analyzing the chess-like strategies of international politics is fascinating, we also need to occasionally step back and ask why the chess pieces are constrained to move as they do, and what contributed to the design of the board and the pieces.

Second—as I hope to make clear in Chapter 2—I am not completely rejecting formal models of rationality based on optimization under constraints, and I am certainly not rejecting formal modeling. While the reader glancing at chapters 5 and 6 may find the latter statement ludicrous, suffice it to say that on multiple occasions my efforts at developing an alternative to rational choice have been deemed a "betrayal" (the critic's choice of words, not mine) of the heroic effort to mathematically unify the study of international politics within the framework of microeconomic optimization.[3] Such doctrinarian defensiveness—ontological arguments divorced from empirical validation—disturbs me: As a paradigm within the behavioral sciences, rational choice

---

[3] Judging from Green and Shapiro (1994,Chapter 8), this vintage whine is endemic among defenders of rational choice.

should be willing to tolerate alternative approaches and should be prepared to defend the validity of its assumptions rather than imposing them as an article of faith.

Artificial intelligence

As reflected in the titles of Sylvan and Chan (1984), Cimbala (1987) and Hudson (1991), the style of modeling used here was originally characterized as "artificial intelligence" (AI).[4] The links between the computational modeling efforts in international politics during the 1980s, and the concurrent developments in artificial intelligence at that time are discussed in detail in Schrodt (1988b, 1991b), and during much of this period, the type of modeling discussed in this book was considered by its developers to be AI.

This characterization has changed for two reasons. First, the term "AI" is so broad, and so over-hyped, as to be both ambiguous and disadvantageous as a label. Second, and more importantly, most of the methods discussed in this book—with the exception of rule-based models—came from the fringes of 1980s AI,[5] and in several instances the methods are closer to those of statistics than AI. In the harsh light of the 1990s, most of the core AI technologies that originally appeared to be applicable to the study of international politics—for example scripts, story generators, case-based reasoning and natural language processing—were never demonstrated as leading to working systems that could deal with large amounts of data. In addition, with the exception of a dissertation or two, there is little evidence that computational modeling in international politics has had any influence on AI, in contrast for example to the influence that work on expert systems had for medical diagnosis.

As a consequence of this disjuncture, in this book I've consistently used the term "computational modeling" rather than "AI", even though much of the work was originally labeled by its authors (including myself) as artificial intelligence. The computational modeling designation is now widely used within the international relations community and is appropriate since algorithms generally define the techniques rather than by algebraic or statistical theory, they can be practically implemented only on digital computers, and they were developed with the strengths and weaknesses of the computer in mind.

Methodology

Social science methodology involves the development of tools that enable one to discover regularities that cannot be uncovered without their aid. In international relations, the use of computerized analysis bears the same relation to the traditional approaches of "wisdom", slow journalism and postmodern textual exegesis as a chain saw bears to a flint ax, a bronze scythe or a feather pillow. A chain saw is not always an appropriate tool: it is more trouble than it is worth when clearing small brush, it is

---

4 Russell and Norvig (1995) provide a contemporary survey of this field; Barr, Cohen and Feigenbaum (1982), Winston (1984), and Charnick and McDermott (1985) show the fields as it was in the mid-1980s; Crevier (1993) provides an interesting analysis of the social, intellectual, and troubled history of the field.

5 Neural networks and genetic algorithms are mainstream AI technologies today, but they were definitely considered fringe when the AI/IR literature was developing in the 1980s.

positively disadvantageous when cutting into a soufflé, and unlike a feather pillow, it is dangerous in untrained hands. Knowledge develops through the creation of appropriate tools: as long as one knows when to leave a tool on the shelf, one can never have enough tools.

There are typically two different approaches to tool development. One can take a single tool and refine it: development in depth. Alternatively, one can develop a series of simple tools that attack various and disparate parts of a problem: development in breadth.

Development in depth is most familiar to social scientists, particularly those working with statistical approaches. Most social science statistical modeling relies on variations of a single technique, the "general linear model", that encompasses correlation, multiple regression, factor analysis, discriminant analysis, ANOVA and so forth, and can be used in both cross-sectional and time series research designs. By investing a great deal of effort on this single model we know its quirks and capabilities in considerable detail.

I would suggest, however, that with our limited understanding of international behavior, there is also a place for tool development in breadth. International politics is a complex phenomenon unlikely to reduce to a few simple equations, and the human objects of our study may employ different methods to solve different problems.[6] I am consequently reluctant to identify a single technique that I consider worthy of a discipline-wide research effort in depth: the development efforts should proceed in parallel, not serially.

An historical analogy has been influential in my choice of this strategy. With the hindsight of history, we can see that in 1900 there were two significant projects in the United States working towards heavier-than-air flight. The project with the money and the heavy-duty technology was headed by the patrician Samuel Pierpont Langley, working out of the Smithsonian Institution with substantial government funding. The project that would succeed, of course, was that of bicycle mechanics Wilbur and Orville Wright, working in Dayton, Ohio with canvas and baling wire.

While both projects were based on the theoretical work in aerodynamics established by German researchers such as Otto Lilienthal, their strategies differed dramatically. Langley developed his project *de novo*. He succeeded in constructing a flying prototype of a machine about the size of a modern model airplane, then attempted to scale this up by almost an order of magnitude. The resulting craft was launched from a houseboat in the Potomac, quickly became unstable, crashed and could not be rebuilt because of its cost. End of project.

The Wrights, in contrast, used an incremental strategy based on extensive empirical testing, starting with kites and models and gradually scaling up to a motorized human-carrying craft. This approach had two advantages over that of Langley. First, in their early, small scale experiments the Wrights were able to isolate the causes of failures, and in particular made critical discoveries about control—problems that had killed Lilienthal—at the stage of non-powered flight where the weight and torque of a motor

---

6 The development of models in breadth seems to common in the decision-making literature in international relations: For example Allison's (1971) classic study of the Cuban Missile Crisis developed three different models to explain the same set of events. Jervis (1976), Lebow (1981) and Vertzberger (1990) provide other examples of model development in breadth. The rational choice and statistical traditions in international relations, in contrast, generally develop tools in depth.

were not an issue. Second, damage to their simple systems could be repaired in hours or days, so an erroneous hypothesis only delayed the research rather than terminating it.

The research community developing computational models of international behavior is working with at least two, and probably three, things we know little about: We clearly are using new techniques. There is almost universal agreement that we need new data. And I suspect, in a Kuhnian fashion, that ultimately we will find ourselves asking new types of questions. In such an environment, the Langley strategy of assuming that the right combination of methods and data and questions will come together all at once seems risky.

In 1992, an issue of *Science* surveyed how several new techniques in the physical and biological sciences had revolutionized not just the methodologies, but also the theories, in their fields. The article observes:

> Not everybody appreciates the importance of technique. Many scientists, in fact, are “theory snobs” who dismiss technique as a kind of blue-collar suburb of science. ... [But there is], clearly, enormous transforming power in techniques. In the absence of an essential technique, a researcher or a field flounders, developing elegant theories that cannot be decisively accepted or rejected—no matter how many intriguing circumstantial observations are available. But with a key technique in hand, the individual and field move ahead at almost terrifying speed, finding the right conditions to test one hypothesis after another. Conversely, new techniques often uncover new phenomena that demand new theories to explain them. (Hall 1992,345)

Contemporary international relations research is arguably theory rich and data poor. In contrast to our colleagues in U.S. and European politics, who are provided data from the massive public opinion surveys of the U.S. National Election Study and Euro-Barometers, as well as governmental surveys such as those provided by the U.S. Bureau of the Census, Bureau of Labor Statistics and Department of Justice, we have relatively little data on international interactions. Concurrently, the international system is becoming more complex with the end of the Cold War, and the need to systematically study alternative theoretical explanations for that behavior is greater than ever.

It is not coincidental that the root of “statistics” is the same as that of “state”: for several centuries those in powerful positions—governments and corporations—have known that much can be learned from the study of the regularities of aggregate behavior. In the 20th century that use of information has become much more decentralized. Among the formal concessions to the democratic process made by the Sandinista government of Nicaragua at a conference of Central American presidents in August, 1989 was a promise not to interfere with public opinion polling. A statistical methodology was thus elevated to a fundamental democratic right. Modern social and political activities, from tax policy to health care to advertising, presuppose mathematical regularities. The knowledge gained from these formal studies is imperfect but it is certainly an improvement on the superstition and convenience sampling they displaced.

Validation

> *Mathematical social science is first and foremost social science. If it is bad social science (empirically false), the fact that it is good mathematics (i.e. logically consistent) is of little comfort.*
>
> *Herbert Simon*

Any proposed universal law of human behavior should be demonstrated to apply in at least one case. Formal models can, and should, be tested; models without empirical tests are at best of heuristic value. It is easy to write equations that at least vaguely mirror human behavior; the challenge lies in finding systems that do this beyond the level expected by chance. Eyeballing will not suffice: the human brain is highly susceptible to seeing patterns where none exist and is notoriously poor at statistical reasoning.

This work deals with the substance of international politics in the sense that it is contains explicit assumptions about how decisions are made, about how international behavior is understood by its practitioners, but it deals with specific situations only in the context of empirical demonstrations of various models. It contains no new insights on the Cuban Missile Crisis, the US involvement in Vietnam, or the end of the Cold War. Fundamentally, it is about fishing rather than fish, though any work on fishing will pay some attention to the characteristics of fish.

All of the models I discuss in this volume have been tested against data on actual international behavior. The tests are illustrative rather than definitive as the data sets are adequate rather than perfect: For reasons that will be explored in more detail in Chapters 2 and 3, the theory of data underlying computational modeling differs substantially from that underlying much of the existing behavioralist enterprise, which is based almost entirely on parametric (and usually correlational) statistics. But those data reflect at least *some* aspects of the empirical world and are certainly better than nothing.

There is a chicken-and-egg problem here. One could insist on collecting appropriate data before attempting to test a new theory or method. However, twenty or thirty years of experience in international relations research has shown that collecting data can take, well, twenty or thirty years, as well as being horrendously expensive.[7] By the time the data are collected, both the theories and techniques may have changed. The available data sets on international behavior are adequate to illustrate many of the strengths and weaknesses of computational modeling; demonstrations of the full potential of these methods will have to wait for new and more appropriate data sets.

Prerequisites

Political scientists are the intended audience for this book, and I've attempted to maintain a technical level comparable to the *American Journal of Political Science*. Individuals with training in economics should not encounter difficulty understanding the arguments, nor should quantitatively trained psychologists and sociologists. The mathematical content of the book is very limited—for example there are no proofs—and

---

[7] The Correlates of War and CASCON projects come to mind, though useful research could be done with those data after five to ten years of collection.

the mathematical notation used is comparable to that of a basic probability and statistics text. I assume the reader to be acquainted with contemporary social science statistical and data collection techniques and do not discuss the statistical models—for example logit and discriminant analysis—that I've tested as alternatives to the computational models.

In place of proofs, most of the formal presentation is in the form of algorithms presented in a Pascal-like pseudo-code. These are supported by some discussion but, as with a proof, the discussion alone is insufficient to render the technique completely clear unless one is already acquainted with the basics of algorithms. I have endeavored to keep this level of detail comparable to that found in computer publications aimed at computer professionals who are not academic computer scientists—for example *Byte*—rather than at the level of *Communications of the ACM* or technical journals on artificial intelligence.

I am often asked whether it is possible to understand and utilize computational models without doing computer programming. I suppose the answer is affirmative, but the experience would be akin to moving snow with a garden rake: it can be done as an expedient but involves ten times more work than the same task with the proper tools. The basic techniques of computer programming are easily learned in a few weeks, the subject is taught at virtually every college and university—as well as many secondary schools—and in the 1990s it is unlikely that the effort required to learn computer programming will be totally wasted in the long term.[8]

That being said, shareware and commercial software is available to implement many of the techniques discussed here, including rule-based systems, genetic algorithms, correspondence analysis and neural networks. ID3 is also widely implemented in expert-systems packages, though often not by that name. I have usually added a twist or two to the standard techniques—for example bootstrapping the ID3 algorithm—though the commercial packages contain their own twists that can also be useful. Sequence analysis methods and sophisticated rule-based simulations such as JESSE and POLI, on the other hand, necessarily involve custom programming.

The language used in this research was Pascal, which I prefer for developmental work because it tends to self-document and employs strong type-checking. I will be happy to provide the source code for these programs to interested academics in "as is" condition; the code is generally follows the de facto "Turbo Pascal" (rather than ISO) standard. These are research rather than production programs, so the input formats are often quite idiosyncratic and involve considerable pre-processing of data. The core

---

[8] Mastery of computer programming takes substantial practice beyond learning its fundamentals, but in that respect programming is no different from learning a natural language or a statistical technique. I have often seen graduate students claim they "can't program" while devising extensive files of SPSS or SAS commands that are nothing more than needlessly elaborate programs performing tasks that could be much more easily accomplished using Pascal, C, or FORTRAN.

Many political science graduate curricula anachronistically recommend that students supplement their training in formal methods with one or two courses in basic calculus. While calculus reinforces algebra skills and may be useful in some simple rational choice and dynamic models, it is usually quickly forgotten and a two-semester course is inadequate for analytical work in advanced econometrics or mathematical statistics. If one expects to invest only a limited amount of time in learning formal methods, I suggest that coursework in elementary computer programming and data structures will be far more useful than calculus, particularly when dealing with social science datasets.

algorithms used are sufficiently simple that for most applications, it will be easier just to start over and write a new program.

All of this research was done on personal computers: the earliest work was done on a lowly but expensive 64Kb Apple II running at 1 Mhz; the most recent work was done on a 8Mb Macintosh II supplemented with a 50Mhz coprocessor. Contemporary computer hardware is dramatically less expensive and more powerful than it was when I began this research around 1984—to say nothing of the bad-old-days of mainframes—and many of the projects that took me days or weeks of processing could be done much more quickly today. Consider these efforts as a starting point, and go on from there.

Organization

This book is approximately half theory and half computational techniques. Chapter 2 is a survey of existing mathematical models with the objective of identifying weaknesses in those methods that such be filled by a computational modeling approach. Chapter 2 presupposes some familiarity with academic research in international relations: Those who are primarily interested computational modeling per se can safely skip it; political scientists will probably want to read it.

Chapter 3 provides a theoretical justification for computational modeling using the patterns/rules/learning approach. This chapter relies heavily on literature in psychology and organizational behavior as well as classical international relations, though it uses those literatures only to the extent that they relate to computational modeling rather than providing a thorough literature survey. The development focuses on identifying the foreign policy decision-making processes and international behaviors that can be expected to be appear regular from the standpoint of the assumptions of computational models. To this extent, Chapter 3 simply justifies the choice of a set of models and their independent and dependent variables, but given the novelty of the computational modeling approach, it does so in considerably more detail than is normally found in empirical work.

Chapter 4 covers rule-based models. It is the only chapter without an original empirical test and focuses instead on models developed by other researchers. Rule-based models are the most widely used type of computational model in international politics—more than half of the articles in Cimbala (1987) and about a third of these in Hudson (1991) employ the approach—so their justifications and methodology have been discussed extensively elsewhere. I review some of the basic premises, techniques and problems of rule-based modeling but, having been encouraged to confine this manuscript to a finite length, the discussion of rule-based modeling is not nearly proportional to the importance of the technique in the computational modeling literature as a whole.

Chapter 5 deals with several machine learning methods. The discussion covers ID3, a method of generating classification rules; genetic algorithms, which produce rules using a system of simulated evolution; neural networks, a method based on information processing in biological nervous systems that I argue parallels several important structural characteristics of organizational decision-making; and nearest-neighbor or clustering methods, which overlap with some of the techniques used in statistical modeling. The ID3 and neural network methods are applied to the Butterworth interstate conflict data set; the genetic algorithm is used to predict patterns of international behavior found in the COPDAB event data set and the nearest-neighbor method is illustrated by using correspondence analysis to cluster the rule-governed international systems proposed by Kaplan (1957) and Rosecrance (1963).

Chapter 6 deals with sequence analysis and is the chapter most specific to international relations research. These methods directly model the recognition of event patterns that I argue are central to human understanding of international politics. The Levenshtein metric, a method originally developed for the comparison of sequences of DNA, is combined with a training method similar to that used in neural networks to differentiate between crisis and non-crisis sequences in Leng's (1987) *Behavioral Correlates of War* (BCOW) data set. The BCOW data are also analyzed with an algorithm that finds similar subsequences in multiple crises; nearest-neighbor methods based on these subsequences can then be used to categorize the larger sequences.

Chapter 7 concludes with a consideration of several issues related to computational modeling generally. These include a comparison of the computational modeling approach and classical approaches to the study of international behavior; two key limitations of the method; a defense of the use of induction, and a discussion of the implications that developments in computer technology and data sources might have for future research using computational models.

Chapter 2
Formal Models of International Behavior

[Human and physical events are] equally susceptible to being calculated and all that is necessary to reduce the whole of nature to laws similar to those which Newton discovered with the aid of calculus is to have sufficient number of observations and mathematics that is complex enough.
Condorcet (1743-1794)

To understand a science it is necessary to know its history.
Auguste Comte

A perennial affliction of the behavioral approach to international politics has been the obsession with philosophy of science: in the words of Kenneth Oye (1987), "To avoid studying world politics, study how others study world politics". While in general I am in sympathy with Oye's complaint, I am also sensitive to the admonition "Before you figure out where you're going, figure out where you are." This volume proposes a variety of approaches and methods, which, if taken seriously, would require a substantial investment of time and effort. With that prospect, only the most avid collector of new techniques would fail to ask whether it is really necessary: if it ain't broke, why fix it?

I contend that it's broken—after thirty years of work on formal models of international behavior, we have yet to create models that have an influence comparable to that of public opinion models in domestic political studies or economic models in that field as a whole. Some isolated triumphs exist—most notably the game theoretic analysis of nuclear deterrence, possibly the Richardson arms race model and recently the empirical generalization that democracies don't engage in wars with each other—but after three decades of effort, it is still difficult to convincingly identify a set of politically relevant knowledge obtained primarily through formal or quantitative analysis. This chapter will survey what has been achieved in formal modeling to date, identify some of the problems in those efforts, and briefly introduce computational modeling as an alternative approach.

A Typology of Mathematical Models

Almond (1988, 1990) provides a long-overdue review of the origins of the scientific approach to the study of politics. In an extensively documented historical discussion, Almond notes that popular mythology, particularly prevalent among Marxists and Straussians, credits the development of "scientific" approach to a Sputnik-crazed United States at the [choose one]:

- height of its Cold War imperial dominance of the global order (Marxists);

or

- nadir of the popularization of the previously august academy by the unwashed masses unleashed by the post-WWII GI Bill (Straussians)

In reality, scientific approach had deep European roots dating to the mid-19th century:

> There were two schools of thought in the 19th and early 20th century social science regarding the possibilities for a science of social behavior. The work of Auguste Comte, Karl Marx and Vilfredo Pareto makes no distinction between the social and the "natural" sciences. Both groups of sciences sought uniformities, regularities, laws. … [The] counterposition of a European and an American approach … around the issue of humanist vs. scientific scholarship will simply not bear the light of day (Almond 1988,837-839)

This tradition was transmitted to the United States by the influx of European social scientists fleeing Nazi persecution in the 1930s, and found fertile ground in North America already prepared by Charles Merriam and the "Chicago School".[1] Concurrently, political science was influenced by the axiomatic modeling of neoclassical economics, starting with the work of Léon Walras and his successor Pareto, that sought to develop a mathematical science of economics similar to the highly successful models of 19th century mechanical physics (see Ingrao and Israel 1990).

In addition to the philosophical roots noted by Almond, two technological factors played a role in the rise of the scientific approach. First, the increase in international communications made more data available. A state such as Thailand ceased being the mystical and exotic kingdom of Siam, and became just another data point in the *UN Statistical Yearbook.* Those data were frequently flawed but were decidedly better than nothing: Since the purpose of statistical analysis is sorting signal from noise, imperfect data present a only challenge, not an insurmountable obstacle.

Second, the influence of computers cannot be overrated. Electronic data analysis finally provided social scientists with tools sufficiently powerful to systematically study social behavior. The motion of an object falling in a vacuum can be modeled by a single equation so simple that it can be taught in high school calculus; the study of an electoral system requires a computer. In the words of Lave and March (1975,2), "God has chosen to give the easy problems to the physicists". For example, the inversion of a 30 by 30 matrix is estimated to have taken about 10 weeks of effort for a team of human "computers" using mechanical calculators; this is a fairly typical calculation in an exploratory technique such as stepwise least-squares or factor analysis. Statistical analysis at the scale common today was impossible forty years ago.

The development of the scientific approach in international relations in the post-WWII period was also influenced by the Department of Defense use of researchers from the natural sciences and engineering to study political phenomena. During the 1950s, formal approaches to international conflict were extended by game theory, operations research and simulation work sponsored by the RAND Corporation, the Office of Naval Research and other defense-oriented groups (e.g. Williams 1954, Goldhamer and Speier 1959, Saaty 1968, Schelling 1960).

---

1 In international relations, the influence of the University of Chicago was noticeable , for example through Quincy Wright, Harold Guetzkow, Morton Kaplan and Rashevsky's work in publishing Richardson's opus. Karl Deutsch also notes Wright's influence in formal comparative approaches such as those of Ernst Haas, Amitai Etzioni and himself (Wright, 1964: xvi). Chicago's Hans Morgenthau, now considered the archetype of the classical approach, forcefully contends in the first chapter of *Politics Among Nations* that the realist theory he develops is "scientific".

The serious development of mathematical models of international behavior in the academic community effectively dates from the founding of the *Journal of Conflict Resolution* in 1957 and was ably ushered into the 1960s with the help of researchers such as Anatol Rapoport, Kenneth Boulding, Richard Snyder and Harold Guetzkow, who brought to the discipline substantial skills in experimental design and mathematical modeling. By the mid-1960s, coincident with the upheavals of the "Behavioral Revolution", mathematical modeling secured a place in the study of international relations from which it has yet to be dislodged.

International relations is now one of the most developed subfields of political science in terms of mathematical models (see Cioffi-Revilla 1979; Nicholson 1989; Ward 1985; Luterbacher and Ward 1985; Zinnes and Gillespie 1976) and employs a wide—perhaps excessive—range of mathematical techniques.[2] By the early 1980s, most formal models in IR could be classified into one of three categories:

Statistical models. These are usually correlational models, but occasionally involve the development of original statistical techniques. The statistical models have been accompanied by large-scale data collection efforts dealing with war, international events, and international attributes. A closely related set of models has used stochastic modeling techniques such as the Poisson model, simple models of contagion and diffusion, and Markov chains.

Dynamic models. Differential equation models developed out of the Richardson arms race model tradition and became a substantial, if specialized, literature. Computer simulation developed jointly out of a human-machine simulation tradition (for example the RAND Corporation's war and crisis simulations, and Guetzkow's INS simulation work); and all-computer simulations, particularly "world models" coming out of the Forrester tradition (see Ward 1985).

Rational models. Rational choice models of international behavior primarily involve decision-making under uncertainty using the expected value decision making framework. Game theory models of interdependent choice have been particularly important to the study of international conflict. The mathematical techniques and most of the assumptions usually have been borrowed from economics with few modifications.

Due to their paradigmatic history rather than logical necessity, the underlying assumptions of the models in these categories tend to be almost mutually exclusive. Statistical studies are inductive and usually deal with aggregated data; only recently have they used time series approaches or explicitly tested hypotheses based on assumptions of rationality. Dynamic models, following Richardson, tend to avoid explicit assumptions of rationality; until fairly recently rational models have not to incorporated explicit notions of time. In the last ten years greater efforts have been made to bridge these gaps, but they are still incomplete.

The successes and failures of each of these general types of models will be discussed in turn, with the not-so-hidden agenda of showing that each has some

---

[2] These vary as widely as graph theory (e.g. Saaty 1968), optimal control theory (Gillespie et al 1976), field theory (Rummel 1972) and catastrophe theory (Holt, Job and Markus 1978; Zeeman 1976). Unfortunately, few of these approaches have been applied in multiple domains, which has contributed to the methodological fragmentation of the literature. Such eclecticism is less prominent in rational choice, which has the advantage of receiving most of its outside advice from economics—a relatively mature and stable social science—rather than from borrowing from natural sciences ranging from meteorology to topology.

substantial weaknesses. The discussion assumes that the reader is already familiar with the basic modeling approaches, though citations are given to appropriate reviews of the relevant literature.

Statistical Studies

> *I must confess that I believe quite firmly that an inductive knowledge of a great number of things in the future is becoming a human possibility. So far nothing has been attempted, so far no first-class mind has ever focused itself upon these issues. But suppose the laws of social and political development, for example, were given as many brains, were given as much attention, criticism and discussion as we have given to the laws of chemical composition during the last fifty years -- what might we not expect?*
>
> *H.G. Wells*

The statistical approach to the study of international behavior is sufficiently well known as to require little explanation; for reviews of this literature see Zinnes (1976), Vasquez (1976), Chase-Dunn (1979), Midlarsky (1989), Vasquez and Henehan (1992), Wayman and Diehl (1994) or any issue of *International Studies Quarterly, International Interactions,* or *Journal of Conflict Resolution* for the past twenty years. Statistical studies are a form of model, both in the explicit models found in regression analysis or factor analysis, and in terms of the complex mathematical framework underlying tests of significance and parameter estimation. In virtually all of the statistical literature, these assumptions are taken as givens, but they are assumptions nonetheless, and as susceptible to misspecification as the explicit assumptions regarding the choice of variables and the algebraic form of the model.

In stochastic models, the dependent variable is random and is explicitly studied as such.[3] The predictions of stochastic models are probability distributions rather than individual points, and the mathematical techniques come primarily from probability theory rather than deterministic calculus or linear algebra. While the concept of randomness is found in many of the other models in political science, it is usually considered as "error", "imperfect information" or some other undesirable characteristic, whereas in stochastic modeling randomness is the primary focus. Because stochastic models deal with probability distributions rather than points, they are more difficult to work with analytically and empirically, but there is still a substantial literature dealing with them.

---

[3]Bartholomew (1971) provides a general introduction to the models commonly used in the social sciences; Schrodt (1985b) provides a survey of these models with respect to international relations research; and King (1989a, 1989b) provides a sophisticated integration of the stochastic modeling and statistical approaches.

Successes of the Statistical Approach

It basically works.

The most fundamental success of the statistical approach is the fact that it works at all: Techniques originally developed for the study of phenomena unrelated to international behavior—for example agriculture, medicine, economics, voting— will by and large give reasonably coherent results when applied to international politics. This proposition is considered self-evident now but was quite controversial when first proposed. The basic problems of data collection, validity and reliability that might have doomed the statistical approach have been resolved, even if more subtle problems remain and there is clearly additional work to be done (see King, 1989a; Most and Starr 1989).

After the testing of thousands of hypotheses, some general results have emerged that can qualify as statistical "fact" in the sense that a number of researchers, using a variety of techniques, have come up with the same results (see Vasquez 1976; Midlarsky 1989; Gochman and Sabrosky 1990; Wayman and Diehl 1994). In this category I would place, for example, the Poisson distribution for the occurrence of war, the lack of statistically significant interactions in the Richardson arms race model, the symmetry of dyadic behavior reported in event data, the correlation (though not the direction of causality) between military expenditures and participation in international conflict, and the pacific character of relations between democracies. While future research may provide new explanations for these results, it is doubtful that the empirical findings themselves will be refuted.

Statistical refutation of the simpler versions of the realist approach.

Most of the work on war and alliances has, implicitly, been in the power-oriented "realist" framework and tested hypotheses suggested by that literature; Wayman and Diehl (1994) provide a comprehensive survey of these efforts. To the extent that the hypotheses tested are accurate reflections of that approach, the statistical studies provide little support for it. The realist response, predictably, has been that statistical tests do not capture the subtleties of the theory. If this is true, the statistical studies have at least established that these subtleties verge on the metaphysical. The empirical inconsistencies in classical combinatorial balance of power heuristics linking alliances and war that have been demonstrated by the Correlates of War research are a major accomplishment, albeit one that the researchers probably didn't originally expect.

More generally, statistical studies find that the international system is a lot more complicated than first believed by the quantitative researchers. While this is sometimes seen as an indictment of the scientific approach, it applies at least as strongly to such hoary aphorisms as *Qui desiderat pacem, præparet bellum* that have long been used to justify policy. With a few exceptions, we now know that simple generalizations do not work, and additional variables alone do not provide a quick fix: The gems of wisdom must be mined from hard rock rather than plucked from the surface.

The creation of large data bases.

The empirical efforts of the past thirty years have resulted in the creation of some very useful data bases, and hence many new theories can now be tested without a massive initial investment of effort in data collection. This has been demonstrated recently in the outpouring of studies on the issue of pacific democracies (see surveys in Maoz and

Russett 1993 and Dixon 1994), which were able to quickly employ sophisticated statistical techniques on existing data sets when this issue became highly salient at the end of the Cold War.[4]

I suspect that many of these data sets have been refined to the point where future theories are unlikely to require significantly different operationalizations, though they may require variables that have yet to be collected. For example, it is unlikely that a totally new project collecting data on the Correlates of War variables over the same set of cases would produce data that would yield substantially different results than provided by the original. Event data are likely to be coded differently in the future (Merritt, Muncaster and Zinnes 1993) but the concept of event data now seems firmed fixed in the discipline. Even if existing data sets are not ultimately appropriate for the tests of new theories, they provide a starting point and we now have considerable experience in collecting such data.

Problems of the Statistical Approach

While there have been some notable successes in the statistical study of international behavior, the approach as a whole is beset by a number of very fundamental problems. International behavior is a hostile environment for traditional statistical techniques in many of its key aspects, and this limits the utility of many conventional statistical techniques.

Universe versus sample.

The most obvious problem is the statistical study of international behavior is our inability to sample or experiment when the unit of analysis is the nation.[5] There exists exactly one contemporary international system and the set of nations comprising that system is sufficiently small and inhomogeneous that there is little reason to sample. Experimentation is possible only in the highly artificial setting of simulations, and introducing statistical controls while preserving a reasonable sample size is frequently problematic. Most studies use either the nation-year as the unit of analysis or study relatively rare phenomena such as wars and crises. Consequently, statistical techniques originally developed for inference are essentially used descriptively, and the meanings of "error" and "significance" are different from what they would be in a research design using randomized samples.[6]

---

[4] The "rational deterrence debate" (see Huth and Russett 1990 and Lebow and Stein 1990), on the other hand, provides an example of a complete failure of empirical studies to reach closure—due to disagreements on operationalization and case selection—on an issue of importance to policy.

[5] I will be occasionally use the word "nation" to refer to the sovereign political entities studied in global politics. The word "state" would usually be more accurate but this can be confused with the quite different concept of "state" in systems theory; "nation-state" is both awkward and inaccurate, since most contemporary states are multi-national. Unless the text is referring to cultural characteristics, the word "nation" should be assumed to refer to a political, rather than cultural, unit.

[6] Some leeway exists—for example one can compare a system over time or try to find historically isolated subsystems such as Hawaii—but most such samples are quite uncontrolled compared to those available to a researcher in voting behavior or local politics, not to mention medicine and agriculture.

In addition to their small population of cases, studies of international behavior usually require, for theoretical reasons, a wide variety of variables. Most traditional theory suggests that the determinants of international behavior are multidimensional. Quantitative researchers have made serious efforts to incorporate the criticisms of their earlier studies as being too simplified by developing more complicated models, but encounter problems with a lack of degrees of freedom, particularly when these new variables are used as controls. Traditional theory calls for considering large number of variables even when the number of cases is small; common statistical techniques such as linear regression simply cannot handle such problems.[7]

The issue here is less one of statistical interpretation as one of limiting the effectiveness of the statistical techniques. Inferential statistics are popular because they allow an economy of effort. In predicting an election or measuring an unemployment rate, statistical techniques can forecast the aggregate behavior of 100-million people on the basis of questions asked to a sample of 4,000. This is a remarkable achievement, but many of those techniques will be less effective in the study of international politics because the problems are very different.

Lack of a obvious stochastic structure.

When a statistical test is based on a large, randomly sampled universe of cases, the stochastic structure of the model can usually be approximated after sufficient effort. In many cases, this structure is in the normal (Guassian) family of distributions, or sufficiently close to the normal that the techniques such as Guass-Markov regression are robust. This distribution can often being justified theoretical grounds based in issues such as sampling, the Central Limit Theorem, the multinomial distribution and so forth.

International behavior is problematic in this regard. First, the number of cases is sufficiently small that it is often impossible to empirically ascertain the relevant stochastic distributions with any degree of confidence. Furthermore the complexity of the processes under consideration—and the number of unknown parameters—often makes the distribution difficult to ascertain theoretically. Finally, the available empirical evidence argues for distributions that are anything but normal and independently distributed. Variables measuring international behavior tend to have distributions with high temporal and spatial autocorrelation, very high variance due to outliers and nonrandom missing value problems. In such cases, the departures from normality are so great that even robust techniques such as the Gauss-Markov linear model may behave unpredictably.

---

[7] In their widely discussed critique of political science methodology, King, Keohane and Verba (1994) argue that in such circumstances one should pay attention to what the regression is saying about the research design, and not try to make inferences when the number of variables exhausts the degrees of freedom available from the cases. While this is true in a traditional regression approach, there are other circumstances—for example where the large number of variables is being used to tap underlying behavioral dimensions that are substantially fewer in number than the sample size—this may still be justified if done in a statistically cautious manner. See *American Political Science Review* 89,2 (June 1995) for additional comments.

Policy substitutability

In an influential 1984 article, Most and Starr discussed the problem of "domain specific laws" and "foreign policy substitutability", the fact that nations have a variety of means to implement foreign policy goals, and that "through time and across space, similar factors could plausibly be expected to trigger different foreign policy acts" (Most and Starr 1989,98). This leads to "many-to-one" and "one-to-many" mapping of circumstances to behaviors that run counter to the one-to-one mappings of the "nice laws" found in the natural sciences and presupposed by many statistical studies. An average soybean plant tested under a set of controls can be expected to respond consistently to an application of fertilizer; it is unlikely to grow on one occasion and wilt on another. A nation responding to a hostage incident, in contrast, might well respond with negotiations on one occasion and bombing on another, and the dependent variables commonly employed in statistical studies will not necessarily be able to differentiate these two divergent behaviors.

Substitutability is compounded by the lack of numerical measures of the behaviors of greatest interest to international relations researchers. Economics probably developed as the first highly mathematical social science in part because the variables of interest to economists—inflation, production, employment, interest rates and so forth—can easily be expressed quantitatively. Voting behavior and budget studies share this characteristic in the political realm. International behavior, in contrast, is usually manifested as qualitative *events* or *relationships* rather than numbers. Thus while the Correlates of War project produced a variety of clever numerical measures of warlike behavior, the bottom line is that international relation theory deals largely with existence of war and not its exact size or duration (Blainey 1988). Similarly, traditional theory puts much more emphasis on the membership and avowed purposes of an alliance than on its size. Data on international behavior tends to be nominal or ordinal; it is rarely interval; and it is frequently dyadic. Conventional social science statistical techniques are weak in this area, and in particular most of the techniques borrowed from econometrics assume interval or ratio data.

The issue here is more than the simple distinction between quantitative and qualitative data: it is the difference between *simple* and *complex* data structures. Most statistical methods deal with data encoded either as scalars (single values such as number of military personnel) or vectors (an ordered set of values, such as population, GNP, number of borders and number of war involvements during the previous two decades). Many important political characteristics, in contrast, are best described by complex data structures. For example, the phrase "Bretton Woods economic system" is meaningful in the discourse on international political economy, but cannot be easily described by a simple number or vector. Instead, it encompasses a variety of rules, institutions, distributions of resources, processes for feedback and self-correction that distinguish the Bretton Woods system from the international economic system of the 1930s and the system of the 1980s. As will be discussed extensively in the next chapter, humans are very facile at dealing with complex of data structures—likely more so than in dealing with numbers—and hence these are likely to be important in determining political behavior.

Overall Assessment of the Statistical Approach.

In contrast to thirty years ago, one can now make a convincing case that some statistical order exists in international behavior, and that the problems of data collection and data analysis are not so great as to preclude any application of these techniques. At the same time, such studies clearly stretch the limits of those techniques more than is done by studies of voting behavior or budgets. By and large, international system presents a difficult statistical environment: we can't sample, we can't experiment, we've got too many independent variables, and our models contain stochastic components with very non-normal distributions. In such an environment, results may be obtained, but they are not obtained easily.

Dynamic Models

> *60% in dreadnoughts over Germany as long as she adheres to her present program and two keels for every additional ship laid down by her.*
>
> *Winston Churchill, 1912*[8]

Traditional international relations theory is frequently concerned with the development of events over time, and therefore it is not surprising that many formal models explicitly deal with dynamic processes. The first widely studied formal model of an international process—the Richardson arms race model (Richardson 1960)—was dynamic and the contemporary computer-based simulations followed Richardson in this regard.

Two types of dynamic models are widely used in international relations. Differential equations, starting with the pioneering work of Richardson, have been used extensively to model arms races and some other types of international behavior. The behavior of some differential equations can be well understood using existing analytical results from mathematics, and these equations can be used to construct parsimonious models that capture the general characteristics of the behavior in question.

Large scale computer simulations grew out of the two traditions. Harold Guetzkow's "Inter-nation Simulation" began by using person-machine simulation but eventually involved into all-machine formulations: the GLOBUS project (Bremer 1987) is the contemporary descendent of INS. Jay Forrester's techniques for simulating social systems were applied to the international environment in the early 1970s in the *Limits to Growth* model produced by the Meadows' research group; this effort spawned a plethora of all-machine "global models" in the subsequent decade. At the core of most all-machine simulations is a set of finite-difference equations, but the simulations emphasize complexity and numerical results whereas differential equations emphasize parsimony and analytical results. Because a simulation is freed from the constraint of having to be algebraically tractable, vastly more complicated systems can be modeled and the simulation literature covers a far richer set of international phenomena as a consequence.

---

[8] Churchill (1948,79). This is the Richardson arms race model, minus the economic burden term, as policy. Consistent with theoretical expectations, the policy led to an unstable arms race and war.

Successes of the Dynamic Approach

Richardson arms race model.

Having languished for forty years in obscurity between its initial creation in 1918 to its publication in the first volume of *Journal of Conflict Resolution* in 1957, the Richardson arms race model is now the most widely studied dynamic model in international relations and, with the possible exception of the incremental budgeting model, the whole of political science. It is, for example, the only political science mathematical model one is likely to encounter in general texts on mathematical modeling (e.g. Olinick 1978; Luenberger 1979). Richardson's equations have all the characteristics of a good mathematical model: they are simple, provide common sense results while also revealing some deeper results, and can be incorporated into other models.

Richardson's pioneering work spawned an entire genre of arms race models that flourished in the post-1960 period. For example, Anderton's (1985) bibliography on arms race modeling contains 224 entries and new models are still being proposed, though arms race modeling is not the active research area it was twenty years ago. Differential equations have subsequently been extensively used to model not only arms races but other types of international behavior such as protracted conflict and event interactions.[9]

The Richardson model has been extensively tested, with mixed results. An assortment of empirical problems exist with the model—the estimates of the coefficients are often very strange, the parameters estimating the interaction between pairs of nations are often statistically insignificant due to collinearity problems, and there are numerous measurement and statistical problems involved in working with the model.[10] The problem of empirically verifying the model is further complicated by the presence of nuclear weapons and/or alliances in many of the arms races on which the model is being tested; Richardson's original two-nation model was not intended to deal with either of those complicating factors.

But even if the Richardson model is rarely completely correct, it is also rarely completely wrong. Most arms races show behavior similar to that predicted by the Richardson model—they do not, for example, tend to be chaotically irregular or cyclical—and no models have emerged that provide a consistently better fit to observed arms race behavior. The statistical record of the Richardson model is also mixed in part because it is a known quantity, having been tested on perhaps 50 or more arms races, whereas most alternatives have been seldom tested.

Complex computer models of global systems.

The publication in the early 1970s of *Limits to Growth* led to a proliferation of large-scale computer models of global systems. These have become a standard research

---

[9] Zinnes 1976; Zinnes and Gillespie 1976; and Luterbacher and Ward 1985 provide a variety of examples; Cioffi-Revilla 1979; Anderton 1985; and Hartley and Sandler 1990 provide extensive bibliographies and summaries of this literature.

[10] See discussions by Majeski and Jones 1981; Schrodt and Ward 1981.

technique in international relations, and several dozen major models now exist.[11] This simulation work initially incorporated few political components and thoroughly ignored the extensive earlier international relations simulation work of the 1960s (e.g. Guetzkow's INS and Smoker's IPS) but more recently global models have incorporated significant political and political-economic factors.

These models have in all likelihood had greater impact on policy—and certainly on the popular perceptions—than any other mathematical models of international behavior. Much of the early simulation work was done at the behest of the United States Department of Defense; the initial all-computer work of the 1970s was funded by NGOs such as the Club of Rome and the Ford Foundation that were explicitly interested in using the models for global planning, and by the 1980s a number of simulations had been developed for governments.[12] These simulations were seen as the international equivalent of large-scale macroeconomic models, and were developed in the hope of providing the same sort of guidance in international affairs that macroeconomic models provided in national fiscal planning.

Freed from the necessity of analytical tractability, global simulations incorporate a far more complex set of assumptions than differential equations.[13] Differential equation models of international behavior have seldom gone far from their roots in Richardson, whereas simulations have gone well beyond the original models of INS or *Limits to Growth* in terms of ideological assumptions, relevant variables, and levels of aggregation.

The complexity of simulations and their emphasis on long range forecasting makes them very difficult to test, and as such they do not have a statistical track record comparable to that of the Richardson model. The models are more commonly used as scenario generators to ask "what if?" questions and generate sets of behaviors and interactions, which might not have been anticipated by researchers.

Problems of the Dynamic Approach

Chaos

Until recently, the implicit assumption underlying dynamic models in the physical sciences was the clock-work, Laplacian universe: a closed deterministic system where the knowledge of the parameters and initial values would forever determine the values of the variables. While it was recognized that problems of measurement and parameter

---

[11] Reviews are found in Deutsch *et al.* 1977; Guetzkow and Valdez 1981; Hughes 1984, 1993; Ward 1985; and Bremer 1987.

[12] For example Luterbacher's SIMPEST was developed for Switzerland; Hughes did extensive government consulting with IFS, and the Federal Republic of Germany financed the ten-year GLOBUS effort.

[13] The "coop model" of territorial aggregation, where the future status of territories are determined by the status of adjacent territories (Schelling 1978,chapter 4; Schrodt 1981a, Cusack and Stoll 1990) is an exception to tendency of simulations to be complex: this type of model contains a small number of parameters and the effects of those parameters can therefore be systematically studied and conclusions obtained about the general behavior of the model. Under auspices of the Santa Fe Institute (Langton 1989; Langton et al 1992) there is now considerable research dealing using simulations that employ large numbers of autonomous agents acting under simple rules, though to date few of these have involved political behavior.

estimation would keep this ideal system beyond the grasp of researchers, one could approach it with arbitrarily high precision in a sufficiently well-understood system. In the astrophysical system, of example, the Voyager 2 spacecraft swept past the planet Uranus only 60 seconds off schedule after a journey of some eight years and 3.2-billion kilometers.

The Laplacian ideal was implicitly adopted by most of the dynamic models of international behavior, particularly the simulation research that came out of the engineering and operations research tradition (e.g. *Limits to Growth*). While astrophysical precision was certainly beyond reach, two characteristics of the international system seemed to promise that dynamic models would provide reasonable statistical approximations. First, social systems were assumed to be largely homeostatic, with feedback mechanisms that would limit large deviations in behavior and keep the system near a predictable path[14]. Second, because aggregate behavior was being modeled, the statistical errors could be fairly well understood. Contrary to popular belief, statistical errors do not actually "cancel out"—in coin flipping the sequence HHHHHHHHHH has precisely the same 1/1024 probability as the sequence HTHTHTHTHT—but the characteristics of large samples of random variables are known and tend to exhibit tractable features such as the normal distribution. The dynamic modeling community implicitly assumed that these two conditions would insure a level of behavioral regularity that could be usefully approximated by deterministic systems. This assumption was not confined to models of international behavior but permeated other models of social systems, in particular macroeconomic models.

During the 1970s, two cracks appeared in this elegant edifice, which, by the late 1980s, had expanded into the field now called "chaos theory".[15] In biology, Robert May (1976) observed that the finite-difference form of the well-known logistic equation (the most common model of population growth where resources are constrained by crowding)

$$x_{t+1} = r\, x_t \left(1 - \frac{x_t}{K}\right)$$

exhibited increasingly random-looking behavior for the values of r in the range $3 \leq r < 4$ despite the fact that the equation is smooth and self-correcting for $1 \leq r < 3$. Figure 1

[14] The classical homeostatic model of social behavior is the basic market mechanism in microeconomics. If supply exceeds demand, prices fall, which makes producers less interested in providing an item and consumers more likely to purchase it. If demand exceeds supply, prices rise, causing additional suppliers to enter the market and consumers to be less inclined to buy. This negative feedback causes the behavior to be self-regulating and relatively stable so long as the environment remains constant. The concept of homeostatic systems was important in early dynamic models due to the influence of the "General Systems Theory" approach of Ludwig von Bertalanffy, itself based in Norbert Weiner's modeling of "cybernetic" self-regulating systems. General Systems Theory was seen in the 1950s as providing the mathematical tools that would allow social and biological systems to be modeled with the same rigor and precision that the differential calculus had provided for physical systems; this strongly influenced early behavioralists such as Boulding, Deutsch, Kaplan and Easton (see Dougherty and Pfaltzgraff 1981, chapter 4).

[15] Gleick (1987) provides a good non-technical introduction; Casti (1989) a short mathematical treatment; Devaney (1986) a book-length mathematical treatment; and Kellert (1993) discusses chaos from the perspective of a philosopher of science. Saperstein and Mayer-Kress (1988) applies chaos theory to an arms race problem; Richards (1993) to the issue of power concentration.

shows both the smooth and chaotic behaviors of the model for K=100 and two values of r.

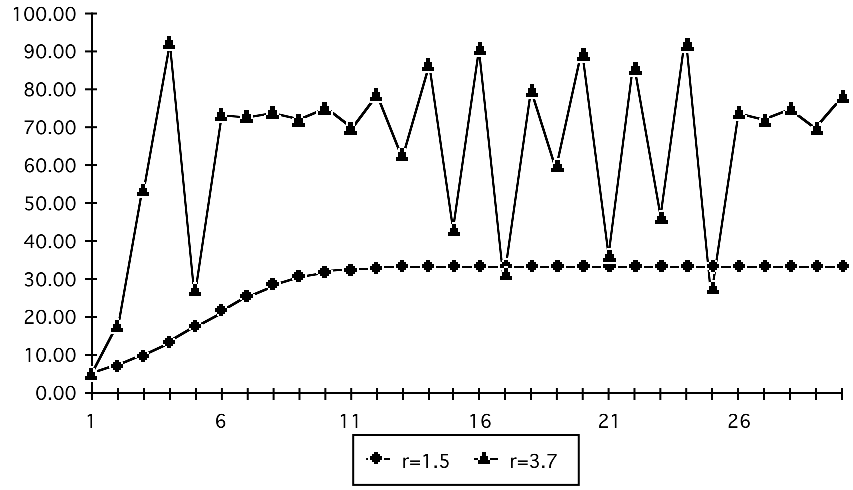


Figure 2.1. Smooth and Chaotic Behavior in the Logistic Model

Chaotic models are also extremely sensitive to initial conditions: arbitrarily small changes in initial values lead to arbitrarily large deviations in later behavior.

Chaotic behavior turns out to be almost the rule rather than the exception in dynamic systems. The essence of these results are summarized by Casti:

> A deterministic mechanism can give rise to random-looking behavior, even when the measurements are exact! Further … it's perfectly possible for a process formulated in terms of stochastic processes to be mathematically identical to a totally deterministic process.(Casti 1989,230)

More generally, researchers have begun to recognize that many dynamic processes will have the characteristic noted by Henri Poincaré:

> A very small cause which escapes our notice determines a considerable effect that we cannot fail to see and then we say that the effect is due to chance. … it may happen that small differences in the initial conditions produce very greater ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes impossible…(Quoted in Gleick 1987,321)

In contrast to the world described by Laplace, seemingly random behavior over long time scales is the norm rather than the exception. This occurs even in systems that are deterministic and homeostatic; in fact the nonlinear self-correcting aspects of the equation are those producing chaotic behavior.

The conditions for chaotic behavior are probably common in international politics. The importance of the logistic equation itself provides one source of chaos, but more generally even the simplest nonlinear finite difference equation system—one containing quadratic terms—can display chaotic behavior. For example, one of the most widely studied chaotic systems, the Henon attractor, is generated by

$$x_{t+1} = 1 + y_t - ax^2_{t+1}$$
$$y_{t+1} = bx_t$$

This is simply a pair of linear equations with a quadratic damping element on x—hardly a convoluted model. Casti (1989) provides an assortment of additional examples for economic, population and biological models.

Lagged responses also seem to create chaos, though in the elementary chaos literature this issue is studied less frequently than nonlinearity because physical systems rarely exhibit the long lags found in social systems. In Schrodt (1981b, 1983) I numerically simulated the logistic with a lagged component and the behavior was similar in many respects to that encountered in chaotic models. Lags are probably a more important determinant of chaos in international behavior than all but the most simple (e.g. quadratic) nonlinearities; most international behavior occurs in systems of the form $x_t = F(x_{t-\partial})$ where ∂ is a random variable.

A model capable of chaotic behavior does not exhibit chaos for all parameter combinations and the onset of chaos is abrupt—an arbitrarily small change in an appropriate parameter value can change a model from smooth or periodic behavior to chaotic behavior. Some systems go through a series of chaotic and stable phases as their parameters change, so neither regularity nor chaos for one set of parameters necessarily implies others will show the same behavior.

As a consequence, dynamic models are caught on the horns of a dilemma. Linear systems do not show chaotic behavior, but they also produce unbounded values when they are unstable and have a very limited repertoire of behavior. The simplest feedback equation that can produce periods of exponential growth but still guarantee bounded behavior is nonlinear—the logistic model—and the finite-difference version of this model can produce chaotic behavior. Since international processes appear largely self-limiting, feedback is all but essential in any credible model. One thus must choose between the unrealistic approximations of linear models or allow the possibility of chaos[16].

This has at least three implications for dynamic models of international politics. First, chaos probably lurks in most of the existing models of international behavior—chaos may not be evident in published parameter combinations, but it is unrealistic to assume that chaotic components are absent. That, in turn, means that the long-term predictions of those models are suspect even if the deterministic specification of the model is itself correct. Forecasts for the global meteorological system—a largely closed system with fixed parameters and well-understood deterministic micro-level principles—become unpredictable after even moderate time periods have elapsed. It seems unduly optimistic to expect that an international system that includes elements of

[16] This distinction accounts for my expectation that chaos theory—despite its faddish character—has important implications for international relations modeling that catastrophe theory—the previous mathematical fad—did not. The popular catastrophe theory models required systems that were homeostatic *and* minimized a quartic (cusp catastrophe) or hexadic (butterfly catastrophe) function *and* used continuous time *and* contained two or more independent, real-valued parameters. The "general" topological results of René Thom basically applied only to mathematical abstractions, and only rarely to empirically realizable systems. Chaos theory, in contrast, applies to models that have been in common use for decades and have realistic features such as quadratic feedback, discrete time, and interdependent parameters.

human choice and poorly understood micro-level principles will be predictable in the long term, even by statistical criteria of accuracy.

Second, the stochastic character of observed behavior is not necessarily generated by the familiar, well-behaved (i.e. uniformly or normally distributed) aggregate processes that are assumed by most common statistical techniques. This would argue for considerably greater care and creativity in dealing with estimation and stochastic modeling.

Third, the sensitivity to initial conditions is a long-term problem, not a short-term problem. The divergence of two chaotic trajectories is gradual, not instantaneous. This argues, however, for determining the behavior of a system on the basis of short-term properties, rather than assuming that the global properties of the system will capture the "average behavior". *If* the underlying deterministic equation of a chaotic process can be ascertained, short-term transitions of the system will be accurately described and usually parameter estimation is possible. In particular, many chaotic systems exhibit statistical regularities called "attractors" that could be usefully studied.

Ordinal time

The concept of "time" is different in dynamic models and in traditional international relations theory. In a differential equation, time is a precise variable capable of infinitesimal division, and the processes modeled are tightly linked to clock time. The *dt* of *dx/dt* is a real number, not a random variable, and attempts to make it a random variable create intractable analytical problems. While some ingenious ways have been suggested for getting around this problem (e.g. Allan 1980; Zinnes 1983), they do not alter the basic problem that differential and difference equations, and most simulations, assume regular time.

In contrast, time in international behavior is largely ordinal: events happen in a sequence, but the timing within that sequence is very loose. Whether a response to a diplomatic communication takes a day or two days is rarely of consequence provided nothing else of importance happens in the meantime. International affairs sometimes exhibits time delays lasting for years—for example, the SALT negotiations, the Law of the Sea negotiations, the withdrawal of the USA from Vietnam and any case presented to the International Court of Justice—without those delays appreciably affecting the eventual outcome. International actors are frequently both immortal and very, very patient, and hence the timing of their activity is largely unpredictable except in issues such as budgets and population growth. Statistical models can deal with this problem using stochastic processes such as the Poisson (e.g. King 1989a), and rational choice models by looking at sequential decision-making (e.g. Powell 1990, Morrow 1994) but it is very difficult in the dynamic modeling framework.

Lack of explicit rationality.

Most dynamic models are *arational* (as distinct from *irrational*) — they lack an explicit process of reasoning. It is *possible* to create rational dynamic models (e.g. Gillespie et al 1976) and it is also possible to models invoking rational decision-making from which one can deduce a simpler dynamic model (e.g. Abelson 1963) but usually this is not done. Arationality obviously weakens the a priori case for the models: foreign policy decisions are made by individuals who are at least processing information, if not acting rationally, and ideally theories of political behavior should incorporate cognitive processes.

Rational choice models, in contrast, have made considerable efforts in recent years to incorporate dynamic elements by using sequential approaches. Most existing dynamic models avoid cognition because the analysis and parameter estimation of dynamic rational models is at best difficult and at worst impossible due to problems of underidentification. Instead, the dynamic modeling tradition places cognition in a black box and tries to fine-tune the box without looking inside. This allows models to be specified and tested, but limits one to predicting outcomes rather than processes.

The absence of an explicit cognitive model may contribute to the lack of closure in the dynamic modeling literature: There is one, and only one, widely accepted model—Richardson's—and even it is somewhat controversial. Differential equation models occur only sporadically in models of behavior other than arms races, and for example there is no dynamic model for the outbreak of war comparable to the arms race model. Simulations vary widely in their assumptions as well as in their parameter estimates (see Hughes 1984, 1985), and one can find a plausible simulation to generate virtually any scenario. While efforts have been made to find standard "modules" for simulations, these have had little success. The lack of closure contrasts to the rational modeling tradition, where a wide variety of behaviors are modeled with a small set of common assumptions.

Difficulties of testing.

Dynamic models are difficult to test. First, there are problems in finding data on processes that remained consistent over a sufficiently long period of time that parameter estimation can be done. Data on international behavior tend to be collected annually—event data are the exception—yet most processes are affected by changes in governments and policies, so it is unreasonable to assume that the parameters of the model would be constant over a period longer than a couple of decades. Simulations involve a very large number of parameters and hence are underidentified even when long time series are available; many of their parameters must be set *a priori.* When these parameters involve obscure or poorly-understood relationships—for example the linkage between resource use and environmental degradation—a priori estimates may be insufficient.

The disjunction between the interval measures used in mathematics and the nominal nature of international behavior poses additional difficulties. Equations deal with quantities, whereas international behavior is usually conceptualized in terms of events or relations. Thresholds and qualitative characteristics such as stability may partially link the interval and nominal measures but these are usually at best approximations.

Overall Assessment of the Dynamic Approach

International behavior is unquestionably dynamic: events and processes occur over time and few if any traditional theories would argue otherwise. The question is how best to model a dynamic process and how far those tools can be pushed.

In physics, the invention of the differential calculus by Newton and Leibniz provided a powerful tool for dealing with dynamic behavior. Physical processes are highly time dependent and the differential equation proved to be an ideal modeling technique for many such processes. To the extent that biological and economic activity is also dependent on such physical processes, differential equations (and their computerized successors) have also proven useful in some domains.

Applying these techniques to international behavior, however, can be problematic: time is only ordinal, variables are frequently nominal, and data and technical constraints complicate empirical estimation. As the implications of chaotic behavior in nonlinear dynamic systems become more fully appreciated, the ability of these models to make long-term predictions is seriously called into question, even when the underlying process is deterministic and fully understood, conditions rarely obtained for most international phenomena.

Differential equations are clearly useful as approximations in some problems, as indicated by the popular success of the Richardson model and global simulations, but whether they will be successful in most problems is quite a different issue. There are a few clear success stories, but whether those necessarily foreshadow further success along the same lines is questionable.

Rational Choice Models

> *I suspect that [Herbert Simon] would have said that a discipline such as economics that finds ordinary behavior surprising probably ought to spend a bit more time looking a ordinary behavior and a bit less time contemplating its theories.*
>
> *James G. March*

> *A gang of Aleutian Islanders slushing about in the wrack and surf with rakes and magical incantations for the capture of shell-fish are held, in point of taxonomic reality, to be engaged in a feat of hedonistic equilibration in rent, wages and interest.*
>
> *Thorstein Veblin*

The "rational choice" approach utilizes models originally developed in economics to explain individual and microeconomic behavior. It first entered international relations in the 1950s and 1960s through RAND Corporation work applying game theory to international conflict, then expanded dramatically in domestic political science during the 1970s: Riker and Ordeshook (1973), Frohlich and Oppenheimer (1978), Abrams (1980) and Ordeshook (1986, 1989) provide book-length introductions. These models are characterized by the not-unreasonable assumption that political processes are the result of individual decisions, and that those decisions are at least partially determined by some form of "rational" calculation involving preferences and constraints, usually expressed in terms of expected utility.

The rational choice models applied to international behavior can be categorized into three general types: game theory, expected utility, and economic competition and maximization models**.**

*Game Theory***.**

The genesis of game theory was in the analytical problems of World War II and it has been used extensively in the study of military problems in international relations.

Military problems tend to be zero-sum and two-actor, for which complete theoretical results exist; cooperative and N-person games have been extensively applied to the problem of negotiation (e.g. arms control: Saaty,1968) though with less analytical success due to the ambiguity of the results in this sub-field. There are a wide variety of books on game theory, most of which provide examples of military/diplomatic applications: Luce and Raiffa (1957) is still an excellent survey; Schelling (1960) specifically addresses the conflict issues; Hamburger (1979), Davis (1983), Shubik (1985), Myerson (1991) and Morrow (1994) provide good text-level introductions. Brewer and Shubik (1979), Brams and Kilgour (1988) and O'Neill (1994) provide surveys of game theory emphasizing its role in defense policy-making. While the original applications tends to focus on simple one-play, complete information games, during the past decade a variety of work has been done with combinations of non-cooperative, sequential, incomplete information and N-person games, particularly in the context of alliances and arms race/control issues: Ordeshook (1989), Downs and Rocke (1990), Niou, Ordeshook and Rose (1989) and Powell (1990) are examples of these efforts.

Expected utility

The expected utility model, which is based on the assumption that individual decision-making involves the calculation of expected loss or gain across a stochastic set of outcomes, is the dominant model of human decision-making in microeconomics, and has been extensively applied to the study of voting behavior. Expected utility is central to many game theoretic analyses but can also be used to construct models of independent choice when the probabilities of outcomes are taken as given or determined by environmental factors. The most notable work in this field is that of Bueno de Mesquita (1981, 1985) and his students; Bueno de Mesquita (1989) provides a survey of this literature.

Economic competition and maximization

The possible convergence between the competition of the marketplace and the competition among nations has been a theme of international relations theory for centuries (see Parkinson 1977) and it is not surprising that theories of economic competition should be applied to international affairs. The bulk of this work has been concerned with arms races and two-nation competition. Intriligator (1971) provides a survey of the techniques, and Intriligator has been one of the more prolific scholars in applying the methods (e.g. Intriligator 1964; Brito and Intriligator 1974; Intriligator and Brito 1984); Busch (1970) and McGuire (1965) also take this approach.

Successes of the Rational Approach

The introduction of concepts.

The most important contributions of the rational models have been conceptual rather than empirical. Formal concepts such as "rational", "zero-sum", "game", "strategy", "utility", "expected utility", "Prisoners' Dilemma", "risk adverse" have proven very useful in characterizing certain international situations. In many cases, however, the *concepts* are introduced without the formal *mathematics*: Schelling (1960), Rapoport (1974), Snyder and Diesing (1977), and Jervis, Lebow and Stein (1985) provide examples of this. These works use game theoretic concepts extensively but employ only a fraction of the mathematical content of works of comparable level in microeconomics. A recent example is Niou and Ordeshook's (1994) assertion that coordination in international affairs—the subject of endless discussion in the "realist-neoliberal debate" of traditional IR scholarship—could be parsimoniously explained by simple equilibrium concepts. This article makes very effective use of game theoretic concepts, but requires only a minimum of algebra.

The early results of game theory also provided two related results that fixed a limit on predictability in international relations: multiple equilibria and mixed strategy solutions. Both concepts provided a formal basis for the observation that international events are often unpredictable, and involve guesses, hedging, feints and so forth. The game theoretic results show that such behaviors are consistent with assumptions of rational behavior, and that in many circumstances predictable behavior would be irrational.

The Prisoners' Dilemma/Chicken analysis of war initiation and arms races.

The single most widely analyzed game in international relations is Prisoners' Dilemma and its variants, particularly Chicken. This mixed-motive game appears to have wide applications in international politics, and is especially important in the analysis of nuclear deterrence and military affairs. The existence of the dilemma is simple to demonstrate, a wide literature on solutions exists, and Prisoners' Dilemma is the formal model most likely to be used in an introductory international relations course.

Axelrod (1984) used the iterated Prisoners' Dilemma to show that tit-for-tat behaviors are selected for in evolutionary situations. Axelrod's results show why the world is not, for the most part, a war of all against all, and why life is not nasty, brutish and short for most people most of the time. This is one possible solution to the problem of explaining the high levels of order found in the anarchic conditions of the international system.

The game theoretic analysis of deterrence.

The problem of nuclear deterrence has been extensively studied using the principles of game theory, and game theory has probably contributed substantially to our understanding of that problem.[17] The nuclear deterrence problem has thus far largely been a two-actor game, it is repeated, and it is sufficiently high on the foreign policy agenda that the unitary actor assumption is reasonable. The failure of nuclear deterrence is also a counter-factual situation, and hence history provides less guidance than it does in other foreign policy decisions, opening a niche for the hypothetical analyses of game theory.

Problems with the Rational Choice Approach

While the rational choice approach can make *some* contributions to our understanding of international behavior, I will argue that it is, for the most part,

---

[17] O'Neill (1994) presents considerable evidence that the influence of game theory in developing nuclear strategy was much more limited than often assumed. Most of the game theoretic analysis followed, rather than preceded, the development of the strategic theories.

unsuitable as the *primary* model for the study of international politics, whatever its merits in studying microeconomic behavior.[18]

Rational choice presents an attractive avenue for modeling and theoretical development because it is a well-developed theory of human decision-making. While most elements of the theory have been borrowed from economics, that discipline is closer to political science than agricultural statistics, meteorology or electrical engineering, to list some of the sources drawn upon by the statistical and dynamic approaches.[19] Economics and political science are the two social sciences dealing with large scale formalized human interaction; psychology deals with individuals; sociology with informal groups; and anthropology with everything.

If one could encompass political behavior and economic behavior—behaviors differentiated in natural language—using a single theory or related set of theories, this alone would be ample justification for accepting a few inevitable inaccuracies introduced in the transition. All models are inaccurate, so it may be better to deal with modifications of an existing theory than to develop a new one.

I contend, however, that we need a separate theory to account for political behavior. The rationale for this assertion is more complicated than the criticisms of the statistical and dynamic approaches and consists of two parts:

- The economics-based approach—particularly revealed preferences and expected utility decision-making (EUDM)—is empirically inappropriate in the international relations environment;
- Political behavior involves fundamentally different calculations than economic behavior—there are reasons why the two are traditionally differentiated—and as a consequence they require different models of decision-making.

These two assertions are virtually independent. There are an assortment of economics-based approaches (see Frey 1983) that reject EUDM but still work within a microeconomic framework where behavior is primarily determined by preferences and constraints, so while EUDM is almost universally assumed in existing rational choice models, one could work without it.

Assertion (2), in contrast, attacks the generalization of the economics framework to *any* political problems: it is as relevant to decisions about Congressional voting as to

---

[18] While the arguments I present here were developed independently (see Schrodt 1989b), many are similar to argument presented at length in Green and Shapiro (1994). Green and Shapiro consider only applications of rational choice to domestic politics but many of their criticisms, particularly on the scientific status of the field, apply equally to international politics. Two differences might be noted, however. First, domestic applications of rational choice usually apply in situations more institutionally structured (e.g. elections, Congress, interest group participation) than those found in international politics. Second, domestic rational choice is usually applied to situations where the individual is the decision-making unit, whereas in international politics that unit is almost always an organization, thus requiring the imposition of an additional unitary actor assumption. Neither difference makes the rational choice approach any more likely to apply in the international realm than in the situations discussed by Green and Shapiro.

[19] For example, three Nobel Prizes in Economics have been awarded to individuals who have made very substantial theoretical contributions in political science: Herbert Simon, Kenneth Arrow and James Buchanan.

decisions about nuclear war. This argument follows paths that have been pursued for decades by Simon, March and others[20], and consequently is central to the larger debates in political science about the relevance of the "positive political theory" approach.

Empirical Problems with the Rational Choice Approach

> *A mathematician may say anything he pleases, but a physicist must be partially sane.*
>
> *Josiah Willard Gibbs*

Expected utility decision-making

The assumption of expected utility maximization, while apparently obvious and common-sensical, has fared poorly in empirical tests. The earliest challenges were the Allais paradox (Allais 1953), which dealt with the valuation of certain versus probable events, and the Ellsburg paradox (Ellsberg 1961), which dealt with the assessment of unknown probabilities; the full-scale assault on EUDM came in the 1980s with the work of Kahneman, Tversky, Slovic and others (see Kahneman, Slovic and Tversky 1982; Hogarth and Reder 1987; Thaler 1991). While there is considerable debate concerning the psychological processes that account for these results, they have been replicated in thousands of experiments, no alternative experimental protocol has been developed that refutes those challenges and thus their empirical validity is seldom seriously questioned.

Particularly relevant to models of international politics are an assortment of studies (usually from the domains of diagnostic medicine, occupational risk or environmental hazards) that deal with threats to human life: in these domains the predictions of EUDM fail with great consistency. These violations of statistical decision-making principles occur even when the decision-makers are trained experts in statistical methodology and/or experienced at making decisions in situations of uncertainty (e.g. business executives, medical personnel, psychologists and judges).

These studies bode ill for EUDM as a model for actual decision-making, particularly in the frequently considered domain of crises. Consider the situation of decision-makers facing a potentially disastrous and unprecedented situation such as nuclear war. The lives of soldiers and possibly an entire nation are at risk. The decision-makers are engulfed in the "fog of war"; information is missing, distorted and created in part through deliberate deception by the enemy. They are short of sleep, emotionally keyed up, and probably influenced by psychoactive chemicals ranging from caffeine and

---

20 See Simon (1985) and Cyert and March (1963). In reading Cyert and March —now over three decades old—one is struck by the fact that the case for microeconomic rationality in organizations was quite thoroughly destroyed in the realm of *economic* behavior, and yet the same arguments emerge, as models of *political* behavior, in the 1970s onward. Cyert and March noted in 1963

> [These] debates … seem to have remarkable powers of reincarnation. They neither die nor fade away with much permanence. Instead, with each publication of new empirical evidence or a new theoretical treatise, the argument is resumed. We will leave the more esoteric features of this remarkable immortality to students of the sociology of knowledge and the metaphysics of wisdom" (pg. 15).

It is as if each generation of social science modelers is newly seduced by the mathematical simplicity of the rational choice assumptions, unaware that is a siren song luring them onto the rocks of empirical irrelevance.

nicotine to mood-alerting prescription drugs. They are concerned about their future in the organization and possibly their place in history; they probably don't know Bayes Theorem from a Banach space. Under such circumstances, is there *any* reason to believe that these individuals will defy all norms of human behavior and engage in expected value maximization? Probably not.

Unfortunately, the EUDM assumption tends to be central and critical for most models of international behavior using the rational approach. While these EUDM formulations are sometimes mathematically elegant, it seems most unlikely that they accurately describe human behavior in the circumstances being modeled. Unless other factors in the international system selects for EUDM in organizations even as it is resisted in individuals—and I will argue against this below—EUDM is weak ground upon which to build a theory.

In economics, these challenges to the fundamental assumptions about the nature of rationality under uncertainty have been taken quite seriously. For example, the Hogarth and Reder (1987) volume resulted from a conference of over sixty scholars, including three Nobel Prize winners, addressing this issue. In contrast, the formal modeling community in political science has tended to dismiss this research as irrelevant.[21] For example, Ordeshook's otherwise excellent *Game Theory and Political Theory* associates the Allais paradox with the entirely unrelated paradox of not voting (Ordeshook 1986,49) and does even mention the Ellsburg paradox or the Kahneman *et al* opus, despite the central position of expected value decision making in game theory. Green and Shapiro express concern that

> The appeal of rational choice in fields like political science, law, and sociology derives, in part, from its reputation for great success in [economics]. Just how well rational actor hypotheses hold up in economics when subjected to empirical scrutiny is debatable... It may be that we are witnessing a curious phenomenon in which rational choice theories are fortified in every discipline by reference to their alleged successes elsewhere. (Green and Shapiro 1994,179-180)

While it is easy to find instances where a science is advanced by postulating processes that remain unconfirmed for a considerable period of time—for example wide-spread acceptance of Mendel's concept of the gene preceded by half a century Crick and Watson's elucidation of its chemical basis—it is much more difficult to find examples where a science has been advanced by holding on to assumptions that have been repeated *disconfirmed*.

I am not suggesting the EUDM is always irrelevant. In situations where probabilities are fairly well known, where the risks are limited, where there is ample time to consider alternatives in a dispassionate fashion (quite possibly informed by operations research explicitly employing EUDM models) and where a situation is repeated, EUDM may in fact be appropriate. Such circumstances obviously exist in international politics: arms control monitoring agreements come to mind; protracted conventional deterrence

---

[21]By the late 1980s a number of researchers in economics, aware of the problems of microeconomic rationality in formal models, began to experiment with computational alternatives involving limits on information, sub-rationality, memory effects and organizational constraints. I have more than once been in the curious position of being attacked by political scientists for rejecting microeconomic rationality, and then having this choice defended by research economists.

systems such as Israel and Syria may provide another example. But for the reasons mentioned, it seems to be a very inappropriate model for situations of crisis, and certainly inappropriate for counter-factual situations such as breakdowns of nuclear deterrence.

Lack of empirical validation

One of the fundamental tenets of the positive economic school is that empirically false assumptions are legitimate provided they lead to empirically correct results[22]. Given the extensive experimental evidence indicating that EUDM is not an accurate description of human decision behavior, one might expect substantial efforts would be undertaken to validate their deductions.

Testing is not, however, a distinguishing characteristic of the rational choice literature—of the three categories of formal models, rational choice models are the least-frequently subjected to empirical assessment. While this situation has certainly improved substantially over the past few years—most notably with the extensive empirical work of Bueno de Mesquita and his co-authors[23]—articles in the rational choice literature remain all to frequently characterized by a surfeit of indifference curves and superscripts and a dearth of estimated coefficients.[24] This absence of empirical testing is due in part to the frequent game theoretic focus on counterfactual situations such as the breakdown of nuclear deterrence—I have no interest in living in a world where such models can be tested!—but for example, none of the five international relations models in Ordeshook (1989) deal with counterfactuals, but none are subjected to empirical tests. As Leontief (1982) and Simon (1985) have pointed out with respect to economics, and Green and Shapiro (1994) note with respect to study of domestic politics, an institutionalized disregard for the empirical world has been a general problem for modelers using rational choice assumptions.

This problem extends beyond the research preferences of the rational choice paradigm to the fact that the basic model is in many ways non-falsifiable: preferences effectively provide infinite degrees of freedom and thus for *any* behavior, one can construct a set of preferences to predict that behavior. As with Alice's Restaurant, you can get anything you want with indifference curves. The Richardson model, in contrast, will only fit certain patterns of arms expenditures (e.g. it fits poorly the pattern of defense

---

22 Friedman (1953) provides the most influential exposition of this principle; most early proponents of the rational choice approach accepted it, though in the past decade it has been treated less favorably. Moe (1979) and Green and Shapiro (1994,30-32) provide reviews of many of these arguments. My own view is that Friedman's principle played a role in the development of mathematical social science analogous to that played by the Piltdown fossils in the development of a theory of human evolution, and should be treated accordingly.

23 See for example Bueno de Mesquita 1981 1985, 1989; Bueno de Mesquita, Newman and Rabushka 1985; Bueno de Mesquita and Lalman 1992. *International Studies Quarterly* 38,3 (September 1994) contains three articles testing models derived from rational choice premises; Fearon (1994) takes on the difficult task of testing a sequential bargaining model. Compared to a few years ago, this is definitely progress!

24 *International Studies Quarterly* 37,1 (March 1993) and *Journal of Conflict Resolution* 37,3 (September 1993) provide several examples. This is not progress...

expenditures by the USA and USSR for the post WWII period). The ability to explain everything explains nothing: such models are tautologies.[25]

At the individual level, Simon observes

> In principle, it should be possible to obtain independent evidence about the nature and shape of any particular person's utility function, as well as evidence of the probabilities that person assigns to events. In practice, this is completely infeasible. In fact, when such experiments have been run, it has generally been found that human subjects do not possess consistent utility functions or probability assignments. (Simon 1985:296)

For example, Tversky and Kahneman (1987,69) note a phenomenon which they call the "violation of invariance": "preferences" revealed in one situation are not necessarily going to apply in another which is mathematically equivalent but either phrased differently or placed in a different context.

The international system is an even more hostile environment for obtaining information about preferences than is the individual. In an election or a consumer choice situation, revealed preference is a rather innocuous assumption: information can be obtained when voters vote or consumers make purchases. In contrast, the potential scope of international actions is so complex, and the pace of international decision-making so slow, that only a fraction of preferences of decision-makers are ever expressed in revealed preferences. In the absence of such information, it is impossible to answer "what if" questions with any certainty.

Achen (1995) argues that these features of international environment render the assumption of rationality in foreign policy decision-making irrelevant:

> The crucial point is that in foreign policy decision-making, alternatives appear once and once only. Even when they have the same name, they remain distinct: To do nothing in Berlin in 1958 is not the same as doing nothing in Berlin in 1961, since they produce different states of the world, and that by definition makes them distinct alternatives. ...
>
> The conclusion, then, is that attention to "the unitary rational actor hypothesis" is misdirected. One cannot test a tautology. Effort should go instead toward finding variables and plausible functional forms which predict state behavior. (Achen 1995, 8 and 15; quoted with permission)

Consistent with Achen's observation, the empirical work of Bueno de Mesquita can be characterized as a test of variables and functional forms that are consistent with, and clearly inspired by, rational choice theories, but they cannot be considered direct tests of those theories.[26] The statistical forms tested do not uniquely follow from rational choice assumptions and are often similar to tests derived directly from realist theories (for example the numerous studies based on the Correlates of War dataset) without the intermediating assumptions of EUDM. While the rational choice approach may have

---

[25] See also Green and Shapiro 1994,34-38.

[26] These tests, while admirable in their extent, are not without critics, e.g. Majeski and Sylvan (1984). In addition, Bueno de Mesquita's rational-choice-based predictions in *U.S. News and World Report* (3 May 1982, pg. 30) have not stood the test of time particularly well.

heuristic value for some researchers, it is clearly possible to construct and empirically assess formal models without it.[27]

A related problem in the empirical assessment of rational choice rests in the amount of information provided by the model compared to that provided by the modeler. Consider, for example, this matrix from Snyder and Deising's (1977) game theoretic analysis of the Berlin blockade crisis:

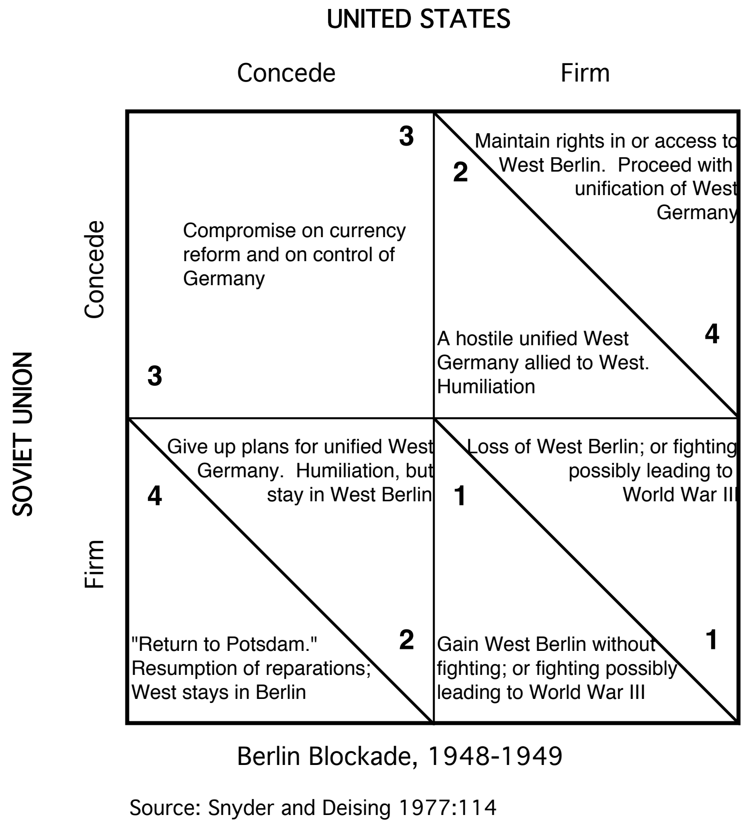


Figure 2.2. The Berlin Decision

Once the Berlin decision has been stated in this fashion, a solution can be derived using game theoretic techniques. Yet from the standpoint of *prediction*, the most difficult problem is not the solution of the game matrix but the construction of that matrix—how were the complex political outcomes in the cells generated?

The specification of most game theoretic models of international behavior presupposes the solution of a very difficult prediction problem involving outcomes. In this respect, international relations differs significantly from most microeconomic decisions and many domestic political decisions, where the space of outcomes is institutionally defined. In international politics, understanding how the players construct the game is at least as important as knowing how they will play the game they have constructed, and the rational choice model provides us with almost no guidance on this issue.

---

[27] For example, only 3 of the 21 empirical studies in Midlarsky (1989) and Wayman and Diehl (1994) employ EUDM (Bueno de Mesquita; Intriligator and Brito; Bueno de Mesquita and Lalman)

Recognizing this problem, in recent years some rational choice modelers have employed models of signaling and incomplete information that explicitly examine the processes by which players might acquire information about each others' preferences and payoffs (see Morrow 1994, Chapter 8; Ordeshook 1989, Morrow 1989; Powell 1990; Banks 1991). This is a decided improvement over the full-information "God's eye view" of classical game theory and is clearly of utility in situations where most of the alternatives are constrained by the characteristics of the political situation (for example a militarized dispute or an arms control negotiation). I nonetheless question whether the game theory approach to incomplete information is the appropriate form when the alternatives are poorly defined (for example in assessing the dissolution of the Soviet Union or Yugoslavia, or the future of Israeli-Palestinian negotiations). Such unstructured predictions were not in the original scope of game theory—which assumed a known set of alternatives—and obtaining them pushes the limits of a technique already on thin empirical ice.

Conceptual Problems with the Rational Choice Approach

As noted above, *if* an empirically accurate theory of international behavior can be constructed and tested using axioms similar to those used to model economic behavior, then there are good reasons to pursue such an approach. These advantages include the practical utility of being able to employ the results of a century of work in mathematical economics and the wider philosophical objective of finding a set of "universal" laws governing human behavior. My contention, however, is that there are fundamental differences between the behavior we describe as "economic" and that we describe as "political", and any models of international behavior must systematically account for those differences.[28]

There is no single definition of what constitutes the "economic" approach, and definitions in books explicitly following the rational choice approach (e.g. Riker and Ordeshook 1973; Bueno de Mesquita 1981) are very diffuse and occupy entire chapters. The following two definitions would seem to capture the key points, however. First, from an excellent review paper by Frey:[29]

> The *homo œconomicus* as dealt with here is certainly *not* the one found in the traditional micro-economic textbooks, where it is assumed that economics actors are fully informed or where the problem of uncertainty is defined away by

---

[28]Lest I appear to be setting up a straw man, this issue is one of the most emotionally charged in contemporary social theory. Economists, and political scientists in the rational choice tradition, almost without exception feel that the assertion that political and economic behavior requires different models indicates ignorance and a lack of appreciation of the full breadth of contemporary mathematical economics. For every example proposed of behavior that is primarily political rather than economic, proponents of the universality of the economic model construct what is, to them, a convincing economic model to explain the phenomenon (also see Zeckhauser's comments in Hogarth and Reder 1987,254; and Green and Shapiro 1994 generally). Political scientists outside the rational choice tradition, again virtually without exception, tend to agree with the assertion and express concern that in attempting to fit political behavior into the Procrustean bed of *homo economicus*, much of what passes for contemporary political science is completely divorced from real world politics.

[29] Frey (1992) provides a published source of these issues.

> postulating certainty equivalents. Rather, the *homo œconomicus* here considered is the much more advanced and generalized model currently used in non-market economics, particularly in the economic theory of politics, or Public Choice [Footnote in original: See in particular the works by Downs (1957), Buchanan and Tullock (1962) and Olson (1965)]. This approach follows the methodological individualism, it strictly distinguishes between preferences and constraints; takes the preferences to be stable and explains changes in behavior by changes in the constraints (e.g. in relative prices); considers marginal changes and stresses substitution; and finally assumes that individuals behave in a consistent ("rational") way. Economic man is thus considered to be resourceful (he searches for and finds solutions, he learns and is inventive), restricted (he is confronted with scarcity and has to choose between alternatives), expectant (he attaches subjective probabilities to future events), evaluating (he has ordered and consistent preferences) and he is maximizing his utility. (Frey 1983,2)

Becker's innovative if controversial work in applying the economic model to various sociological phenomena such as marriage, crime and fertility uses a definition that is quite similar:

> The combined assumptions of maximizing behavior, market equilibrium and stable preference, used relentlessly and unflinchingly, form the heart of the economic approach.(Becker 1976,5)

Following Frey, one can intelligently discuss an economic model even if the classical assumptions of expected utility maximization are wrong and behavior is instead non-linear, information rich, socially influenced and generally far more complicated than EUDM assumes.

Nonetheless, I contend that there is a difference between the behaviors that we call "economic" and that we call "political" which lies in the issue of private (excludable) versus collective goods[30]. "Economics" deals primarily with production and exchange of private goods where the dominant constraints are physical, involving time and resources, whereas "politics" deals primarily with the development of coalitions where the dominant constraints are social, informational and historical.[31] Economics is the study of behavior determined by the interplay of preference and constraint; political science the study of behavior dominated by the formation of coalitions to solve collective action problems

In economic activity, the single individual can be a viable unit; in political activity the individual, alone, is by definition without political power—political power is

---

[30] A collective goods situation is one where goods are non-excludable: once provided, everyone derives benefit. The standard example is a lighthouse; free highways, national defense, and clean air are other examples. The individually rational behavior is to not participate in providing the good and simply let others do the work. If everyone follows this logic, the good is not provided, hence the paradox. The best known modern work on the problem is Olson (1965); Barry and Hardin (1982) provide a good collection of readings on the issue, which has unsurprisingly dominated much of contemporary political theory; Sandler (1992) provides an excellent integrated introduction to both the political and economic issues, including a variety of empirical applications.

[31] The term "coalitions" is used here in the general sense of the collective action literature—any group of actors collectively engaging in a mutually-beneficial activity. The military coalitions founds in international politics are one example but not the only.

achieved only when groups of individuals act in concert. The physical requirements of production and consumption required for survival vary little between individuals, across time, or even across cultures, and therefore the assumption that these may be held as universals in determining economic activity is, prima facie, not unreasonable. But assuming the same consistency in the culturally and historically mediated behavior of political coalition formation is a major step. Economic behavior exists to provide efficiently the basic goods that individuals require to survive; political behavior exists to solve public goods and Prisoners' Dilemma problems required for organizations to exist. Economics in the end is a physical process, politics a social process. Robinson Crusoe, living alone on an island, still had to solve economic problems; he did not have political problems until the arrival of Friday.

To take the extreme case, war is a profoundly cooperative act. Modern warfare requires the coordination of literally millions of individuals to expend a great deal of time and fortune to send a few of their number great distances to kill people they've never met and whose deaths produce no direct benefit.[32] The conflictual aspect of war is the tip of two vast cooperative icebergs: modern war requires political organization. Primitive societies without large hierarchical systems and institutions cannot engage in this form of war, but only in short-lived brawls and organized thievery emphasizing individual heroism; civilization is required to wage war.

Along a similar line, there are many political activities that individuals would not undertake voluntarily, but do because of the threat of sanctions. But where do those sanctions come from?—obviously from the coordinated actions of individuals in the form of the police, a bureaucracy, the community, a neighborhood gang or whatever. Behind every set of anticipated sanctions lies a great deal of institutionalized cooperation; an "absolute" dictator such as a Hitler or a Stalin requires a lot of help. As Hobbes observed, all humans must sleep, and any human who is asleep is vulnerable to destruction: this vulnerability can only be countered through political coordination.

These institutionalized cooperative structures are so common that we tend to take them for granted. A great deal of political theory presupposes the existence of the "state", a reasonable assumption since persistent political institutions are a prior condition for the social stability required to permit the writing and dissemination of works on political theory. Strictly speaking, however, the activities of the "state" require an individual-level justification as collective action: one cannot assume that the state can "force" actions such as military service or taxation without also solving the collective action problem of coordinating those who are going to do the forcing.

Virtually all civilizations have distinguished between behavior that is economic and behavior that is political.[33] We can also recognize political behavior occurring at various levels: the term "office politics" has meaning to anyone who has worked in a large organization, and "office politics" shares some characteristics with politics at higher levels. In "office politics" leaders can be identified; there are agendas of issues under active debate and issues considered resolved; there are crises resulting in policy change,

---

32 As distinct from criminal homicide, which is usually directed either against spouses, friends and relatives or in the course of an economic exchange such as a drug deal or robbery.

33 More so, for example, than they differentiate between behavior that is religious and behavior that is political: Jaynes (1976) argues for a psychological convergence of these two in early human history.

and so forth. These political activities usually occur in the absence of direct economic exchange[34], though a change in economic circumstances may be one of the results of the political activity.

Because economic and political activities differ in their fundamental focus, models of the two types of behavior ultimately will be different, though they will doubtless share some characteristics. The two theories are unlikely to diverge on the issue of *preferences*: one can safely assume these exist in both market and non-market behavior. At issue is whether *maximization* (or what form of maximization) explains the regularities observed in behavior considered political.

Resources are the limiting factor in economic behavior: individuals have a finite amount of time and money and wish to utilize these in a manner maximizing the acquisition of goods and services that increase ones sense of well-being. Utility responds monotonically to resources for most people, and if individuals or organizations are either sufficiently intelligent to know what is good for them, or exist in a system that enhances the survival of those who act as if they were maximizing, a model assuming some form of maximizing principle (not necessarily EUDM) will have predictive value.

In political affairs, in contrast, the limiting resource is coordination rather than maximization. Utility maximization alone rarely provides a sufficient organizing principle to explain political behavior; the solution of the collective goods behavior problem by maximization is not a solution. A predictive theory must therefore rely on additional mechanisms such as institutions, rules, concepts of fairness, memory and so forth, even when it involves preference. As March and Olden note

> Although self-interest undoubtedly permeates politics, action is often based more on discovering normatively appropriate behavior than on calculating the return expected from alternative choices. As a result political behavior, like other behavior, can be described in terms of duties, obligations, role and rules. (March and Olsen 1984,744)

Frequently, the limiting factor in collective action situations is *information* on the current and anticipated compliance of others.[35] Participation in collective action is ill-advised if there is evidence that most people are attempting to free ride, as this decreases the likelihood that the good will be successfully provided, and means that one is investing a disproportionate effort to obtain the good. The attraction of Gandhi, Khomeini or Mandela as political leaders is that they were *not* utility maximizers[36]: they could not be

---

34 Economic exchange does occur in some political situations, but it is usually considered "rent-seeking", "bribery" or "corruption", neither a term of approval. Keeping economic exchange *out* of politics has been a concern of theorists since at least the time of Mencius.

35 Ostrom (1990) provides an excellent analysis of these issues with respect to stable self-governing systems for the maintenance of common-pool resources such as forests, fisheries and irrigation systems.

36 Those defending EUDM through *ex post facto* explanations would doubtlessly argue that these leaders were maximizing utility in that they achieved more by being in opposition than by being bought off. But this assessment is virtually impossible *a priori*: identify, for example, the comparable utility maximizers in contemporary Zaire, Palestine, Bosnia and Rwanda. The chance of Gandhi, Khomeini or Mandela *per se* succeeding was remote; dozens of leaders of comparable stature were either killed (for example Steven Biko in South Africa) or successfully co-opted; and biographies of such leaders seldom emphasize a dominant role for EUDM in their personal philosophies.

bought. As a consequence, they provided nuclei around which political movements could coalesce. Furthermore, the collective goods problem cannot be solved by the acquisition of additional resources alone: if that were true, the British would still rule in India, the Pahlavi regime would still rule in Iran and the apartheid system would be alive and well in South Africa.

If utility maximization were the sole determinant of human behavior, political activity should be characterized a high level of change and disorder one sees in markets. In fact, the opposite is observed: the political world is usually very stable. The mayor of Chicago in 1960 was Richard M. Daley; the mayor of Chicago in 1995 is Richard M. Daley Jr.: would simple chance or microeconomic maximization predict the choice, out of 3.5-million people, of an individual who just happens to be the son of the former mayor?[37] At the organizational level, this order comes about in the form of rules and institutions; at the individual behavior level, it comes about through norms of "fair play", "justice", "morality" and other limits to individual utility maximization when collective action is required.

Overall Assessment of the Rational Approach

This section has been considerably more detailed than the previous two and its length testifies to the power and attractiveness of the various formal models of economic behavior applied to the study of international behavior. While the rational choice approach certainly can—and has—provided important insights, rational choice this has some critical weaknesses limiting its general application to the study of international behavior.

The international system is fundamentally different from a market, where basic information on price and demand is timely, inexpensive and accurate. EUDM can model, in a simplified fashion, the absence of information, but there is little, if any, evidence that EUDM describes actual human behavior, and ample reason to believe it would not describe the decisions made by groups under stress. The decision to go to war is made in a different fashion than the decision to go to lunch.

At the theoretical level, the decisions we characterize as political differ fundamentally from those that we characterize as economic. Maximizing behavior acceptable and encouraged in the economic systems becomes self-defeating in a political system. No one wants to share a foxhole with a utility-maximizer. The slow, ambiguous and quasi-anarchic nature of international interactions make them particularly ill-suited to economic models that assume fast feedback, full information and, implicitly, a known and relatively constrained set of allowable behaviors.

When buyers in a market are dissatisfied with a price, they will refuse to buy the product. In the international system, they might steal the product, manufacture the product, destroy the marketplace or decide, through a change in government, that the product wasn't desirable in the first place. For governments such behaviors are typical; for individuals they would be pathological.

---

[37] A phenomenon hardly limited to Chicago: Kansas's Senator Nancy Kassebaum "just happens to be" the daughter of 1936 Republican presidential candidate Alfred Landon, a fact I became aware of when I heard a retired member of our faculty refer to the senator as "that Landon girl".

Rejection of the EUDM rational choice model currently popular in political science does not mean the rejection of any model of rationality. If anything, the opposite is true. Simon notes:

> Skepticism about substituting a priori postulates about rationality for factual knowledge of human behavior should not be mistaken for a claim that people are generally "irrational". On the contrary, I think there is plenty of evidence that people are generally quite rational; that is to say, they usually have reasons for what they do. … To understand and predict human behavior, we have to deal with the realities of human rationality, that is, with bounded rationality. There is nothing obvious about these boundaries; there is no way to predict, a priori, just where they lie. (Simon 1985,297)

When a road ends in a swamp, one can step on the gas in the hope that the swamp is shallow and dry land lies beyond the next curve, but it is quite likely one will simply spin one's wheels or go deeper into the swamp. Alternatively, in such situations, one can search for another road.

Computational Modeling

The models described above are presumably familiar to any reader with knowledge of contemporary international relations research. Computational models—the focus of this volume—effectively form a fourth general category of formal models. Computational models are distinct in that they are specified algorithmically rather than algebraically; they tend to be less parsimonious than either rational choice or differential equation models, but in common with rational choice, the models usually have an explicit information-processing component.

The computational modeling approach fills in some of the weaknesses in the existing approaches identified above. Specifically, they depart from the prevailing techniques in at least the following ways:

Experimentation with a diverse set of formal structures.

Existing formal models have been strongly influenced by models of physical processes—in rational choice, intermediated by economic modeling—and have generally confined themselves to techniques that could be represented algebraically. The computational approach extends formal modeling to the much wider set of structures and processes that can be represented algorithmically, and generally borrows its techniques from computer science rather than mathematics.

As noted in Chapter 1, the computational modeling efforts in international relations were originally seen as an extension of artificial intelligence research (Sylvan and Chan 1984; Cimbala 1987; Schrodt 1988b; Hudson 1991), but over the years its proponents recognized that the AI agenda was quite distinct from most of the work being done on international behavior. While AI research clearly provided some important tools for the development of computational models of international behavior, the theoretical bases for the political models are primarily derivative of cognitive psychology, organizational behavior and political science.

A focus on discrete events rather than continuous variables, and ordinal rather than continuous time.

This emphasis follows from the first: the tools of classical mathematics function best in the continuous realm; digital computers, by necessity, function best in the discrete. As I've argued in this chapter, most political decisions are framed in discrete rather than continuous terms, and in many instances the computational approach provides a more natural framework for modeling political behavior than that provided by the mathematical approach.

A focus on short-term regularities.

Consistent with the rational choice approach and most statistical studies, but distinct from many dynamic models and simulations, most computational models work with short-term regularities rather than long-term processes or equilibria. Computational approaches—particularly rule-based models—differ from rational choice in their ability to handle complex sequences of context-dependent, inter-related rules.

An emphasis on processes consistent with the information processing limitations of human individuals and organizations.

As will be discussed in detail in Chapter 3, the decision-making approximations found in computational models are usually informed by work in psychology and organizational behavior rather than relying on mathematically convenient "as if" assumptions such as EUDM.

Extensive empirical tests.

In contrast to the rational choice approach, computational models tend to be very data-rich and demonstrate considerable breadth in their substantive foci. For example the topics in Hudson (1991) include the Johnson administration's Vietnam policy, Eisenhower's Vietnam policy, Japanese energy security policy, economic development, United States policy in the Caribbean, international relations in the early Cold War period, the international events recorded in the BCOW and CREON data sets, the Senate Foreign Relations Committee deliberations on the Persian Gulf in the 1980s, the Soviet intervention in Hungary, and Luttwak's theory of *coups d'etat*.

As I hope is evident in this chapter and the next, I am not suggesting the development of computational models to the exclusion of the other types of formal modeling. I am simply suggesting that computational approaches reflect characteristics of political decision-making that are not easily captured with algebraic methods. The use of different models to capture different elements of a behavior is not unusual: note that in the arms race literature at least four distinct formal models have been developed to explain the behavior:

- Differential equations (Richardson 1960)
- Comparative static models of duopolistic competition (Brito and Intriligator 1974)
- Iterated prisoners' dilemma (Brams 1985)
- Dollar auction (O'Neill, 1985)

While it may be possible to equate these through a major exercise in logical gymnastics, they have very different assumptions about the underlying process, although all probably reflect some elements of that process.

At the same time, I should also point out what I consider to be two key weaknesses in existing computational modeling research. The first weakness is the lack of a systematic stochastic framework. As discussed above, there are a number of good reasons to treat international behavior as having a significant random component. Appropriately, both the statistical and rational choice approaches are firmly based in a stochastic conceptualization of politics and make extensive use of probability theory.

This has not been true of computational modeling, a problem the approach shares with artificial intelligence generally. The reasons for this are pragmatic rather than theoretical. First, there is no clear algorithmic analog to probability theory. As will be discussed in Chapter 4, fuzzy set theory and other alternative probability systems such as Dempster-Shafer logic may eventually fill this gap, but currently these do not provide the same level of support for computational models that probability theory provided for the development of game theory and computational statistics in the 1950s and 1960s. The second reason concerns computer capabilities: probabilistic schemes could be implemented using bootstrapping or Monte-Carlo methods, but these are extremely processor-intensive and can challenge even the capabilities of supercomputers. At present, the prototyping and experimentation that characterize computational modeling are practical only in a deterministic mode. I would hope that once this bush clearing is finished (and as computer capabilities increase), computational models will begin to incorporate much more sophisticated stochastic elements to deal with the noise, uncertainty and rational randomness of the political environment.

The second weakness of computational modeling to date is its failure to provide significant new insights into general theories of international behavior. I would attribute this largely to the fact that the field is quite new. After all, two decades elapsed between the development of game theory in the 1940s and the widespread use of game theoretic concepts in political science in the 1960s, and a similar lag occurred in the application of systematic studies of public opinion. Still, caveats are in order so that the approach is not over-sold.[38]

At the present time, computational models are considered successful if they produce plausible results. For the most part, the models are not dramatically more accurate than statistical models working with comparable data, nor are their insights superior to those of informed intuition. Machine-learning techniques—the focus of my own work—are further hampered by their heavily inductive character and consequent need to derive from their data considerably more information about the regularities of political behavior than is required in a typical game theory or hypothesis testing model. In the long term, this induction is a strength of the machine-learning approach, but the early results can be pretty mundane.

Given this, why bother? To this question, one can only respond with Faraday's reply to a query concerning the utility of his experiments with electricity, then a

---

[38] Unrealistic expectations have been a notorious problem for the field of artificial intelligence, which fluctuates between well-funded periods of hype (1955-1965, 1980-1990) and poorly-funded periods of "AI winter" (1965-1980, 1990-present).

laboratory novelty: "What is the use of a new-born baby?"[39] The newborn baby of computational modeling is at present something of an ugly duckling. In the future it may become a beautiful swan, or it may grow into a larger but equally ugly duck, or it may languish and fail to grow at all. In order to assess the future of computational modeling, we must look in detail at characteristics of international behavior that argue for this approach. These are discussed in Chapter 3.

---

[39] Confronted with this same question by William Gladstone, then Chancellor of the Exchequer, Faraday took a more practical approach, "One day, Sir, you may tax it." (Mackay 1977,56)

Chapter 3
Patterns and Politics

It is a profoundly erroneous truism, repeated by all copy books and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations which we can perform without thinking about them.

Alfred North Whitehead

This chapter develops a partial theory of the sources of international behavior that justifies the development of the computational modeling techniques in Chapters 4, 5 and 6. It is based on cognitive and organizational studies of foreign policy decision-making and seeks to integrate three things:

- What international behaviors are sufficiently regular that computational approaches can be used to model them;
- What types of information should be used in those models;
- What processes link this information to the observed behavior;

In behavioralist terms, I will be justifying my choice of dependent variables, independent variables and formal model. In so far as practical, formal models should be consistent with what is known about the components of the system being modeled. A model of cell metabolism can assume the existence of an as-yet-unknown enzyme that facilitates a specific reaction, but it cannot involve a violation of the laws of thermodynamics.

In principle, a detailed justification of this sort should be part of every scientific study. The realities of research dictate that in practice we tend to choose our variables based on the available data, and choose models—whether regression, differential equations or games—based on known techniques (and I certainly do not exclude myself from this characterization). With the luxury of a book-length exposition, tenure and a several years to deal with the problem, I am approaching the justification of computational modeling in a more systematic fashion. I am seeking points of intersection between *some* of the computational methods made possible by contemporary computer technology and data resources, and *some* of the characteristics of international political behavior. My objective is not to provide a universal model of foreign policy decision-making, but to justify the use of a class of models for some aspects of that behavior.

The approach discussed in this chapter has guided my development of computational models in international relations, and it could well guide the development of methods far more sophisticated than those I'm presently able to demonstrate. I suspect that computational modelers, particularly those working with rule-based modeling, have been guided by similar considerations. However, this theory is not solely driven by the availability of one or two very specific methods—it is not the proverbial hammer in search of a nail. Algorithmic specification imposes few constraints; computational models have employed a relatively narrow set of methods compared to those available in artificial intelligence and computer science and this has been dictated (or at least informed) by the nature of the phenomena being studied.

This chapter is somewhat multifaceted, jumping between cognitive, organizational and political explanations. A brief map of the morass into which we are about to plunge follows:

1. Four basic concepts are defined: information, events, patterns and classification.
2. Some characteristics of foreign policy decision-making are discussed: foreign policy requires prediction; it is an “ill-structured” problem; decision-makers focus on predicting small sets of events rather than specific events; and prediction tends to be short-term.
3. Humans are very good at pattern recognition and use recall instead of logical reasoning whenever possible. The international relations environment, due to its inefficiency in providing information, is particularly well suited to the use of pattern recognition. Much of this occurs on a sub-cognitive level: one can find patterns without being able to explain verbally how this was done.
4. Much of the knowledge required to make predictions is stored as stories. A story has a set of associated conditions and a temporal ordering of events. Stories deal with general patterns of behavior such as wars, coups and crises; a general story is matched to a specific international situation using substitution rules.
5. Organizational decision-making is constrained by the necessity of communicating and retrieving information. Information processing is much less efficient when communication must occur between individuals than when all of the processing occurs within an individual. Explicit organizational memory consists primarily of a large number of relatively simple *if...then* rules that can be processed without associative memory. However, communication in an organization can also serve to activate patterns held by its members; the content of communication is therefore influenced by these shared patterns.
6. Individuals and organizations learn patterns and rules primarily by example and through adaptive responses to the international environment, particularly the success and failure of existing policy. Policies are created by evolution rather than by design, rules are not necessarily logically consistent, and the full implications of a set of rules and patterns are unknown.
7. The use of rules, pattern recognition, and adaptation leads to behavioral regularity in the system, particularly in the short-term. The complexity of the behavior in the system is constrained by the informational processing limitations of its constituent parts, and the co-adaptation of competing organizations leads to equilibrium sets of rules.

Definitions

The discussion in this chapter rests on the concepts of *information*, *events*, *patterns* and *classification*. While these terms are common in natural language discourse, formal definitions and some associated concepts will be developed here.

Information

Information provides the ability to discriminate between patterns; in other words, to classify a case as fitting a particular pattern. In information theory, information is "a measure of the *amount of selective work* a message enables its receiver to do" (Krippendorff 1986,13; emphasis in the original).[1] Stated somewhat differently, information is anything that could change a decision.

Decision-making organizations, particularly since the advent of electronic communications, tend to have an abundance of material available that might be information, but much of it is irrelevant to decision-making.[2] Thus a distinction can be made between "data" and "information". Data are potential information; information must make a difference in a decision. In the decision-making world, information is usually in the form of text or numbers, because decision-makers receive almost all of their information from secondary sources.

The information value of an attribute—its ability to discriminate between patterns—is dependent on the situation being considered, so information is domain-specific. For example, electoral laws are probably of interest in a study of governmental stability in a parliamentary system, whereas the ethnic composition of the military is likely to be more important in a government with military rule. Data that are information in one problem may not be information in another.

In the situation of a classification problem (defined below), a *feature vector* is the set of attributes describing a case. The number of features is the *dimension* of the vector. A feature can take any value from a set of values; for example if the feature is the size of a country's defense budget, the allowable values are positive numbers, whereas if the feature is "type of government", the values will be chosen from a qualitative typology such as "liberal democratic", "single party democratic", "military", "monarchy" and so forth. The values a qualitative attribute may take will be called *categories*. The set of all possible values of the feature vector describing a system is called the *state space* of the system.

A term I will use frequently in the context of organizational decision-making is *bandwidth*: the amount of information that can be conveyed in a unit of time. This word carries some unnecessary physical overtones (as does this metaphor…) but has no obvious informal equivalent. A narrow bandwidth means that relatively little information can be conveyed; a wide bandwidth means that a great deal of information can be conveyed.

---

[1] See Pierce (1980) for a general discussion of the development and use of information theory in its original context, communications; Krippendorff (1986) provides an excellent discussion of its applications to social science modeling and analysis.

[2] In the policy literature, this is often called "intelligence": Fain, Plant and Millroy (1977,78) note the official definition of intelligence in the *Dictionary of United States Military Terms for Joint Usage*:

> Intelligence is the product resulting from the collection, evaluation and analysis of all available information which concerns foreign nations or activities and which is immediately or potentially significant to planning and decision-making.(Fain, Plant and Millroy 1977,78)

Note that this definition uses the term "information" to refer to the larger concept that I call "data".

Events

As I will argue in more detail below, the focus of computational models, and foreign policy decision-making generally, is on "events".[3] An early paper by McClelland, one of the first event data researchers, provides the following definition

> Event-interaction is meant to refer to something very discrete and simple—to the veritable building blocks of international politics, according to my conception. The content of diplomatic history is made up, in large measure, of event-interactions. They are the specific elements of streams of exchange between nations. Here are a few examples for hypothetical Nations A and B: Nation A proposes a trade negotiation, Nation B rejects the proposal, Nation A accuses B of hostile intentions, Nation B denies the accusation, Nation B deploys troops along a disputed boundary, Nation A requests that the troops be withdrawn, … Each act undertaken by each actor as in the illustration is regarded as an event-interaction. (McClelland 1967,8)

In the simplest terms, an event is an interaction between two entities in the international system. Gerner et al (1994,95) provide a more formal definition:

> An *event* is an interaction, associated with a specific point in time, that can be described in a natural language sentence that has as its subject and object an element of a set of *actors* and as its verb an element of a set of *actions*, the contents of which are transitive verbs.

Whether a particular interaction is or is not an event depends on the individual or organization and the substantive domain of the decision.

Actors are the persons, organizations and places that might affect a decision. The decision-maker is interested in only certain interactions between these actors, and these can be described by transitive verbs such as "apologize", "met with", "endorsed", "promise", "accuse", "threaten", "attack" and so forth.[4] Multiple verbs might signify the same category of behavior, either because the words are synonyms within the language (e.g. "grant", "bestow", "contribute", "donate") or because the behaviors, though linguistically distinct, are politically equivalent, as with Most and Starr's (1984) "foreign policy substitutability". These equivalence sets vary with the individual and the specific problem being considered.

Pattern

While Margolis' assertion that "Pattern recognition is all there is to cognition" (1987,3) is stating the case excessively, pattern recognition is an important foundation for

---

[3] For surveys and critiques of event data research in international relations, see Andriole and Hopple 1984; Azar and Ben-Dak 1975; Burgess and Lawton 1972; Daly and Andriole 1980; Gaddis 1987; *International Studies Quarterly* 1983; Laurance 1990; Munton 1978; Schrodt 1994; and Peterson 1975. Merritt, Muncaster and Zinnes (1993) provides a survey of current work in the field.

[4] That is, a verb which can take a direct object and indirect object. For some events, the second actor is the direct object of the sentence ("Syria accused Israel…"); in other cases it is the indirect object ("Saudi Arabia promised economic aid to Syria").

a computational model of politics. The essence of pattern is *coassociation*: something is "patterned" when one can infer unknown features of a case on the basis of known characteristics. The term "plaid shirt" is a pattern that matches a large number of objects differing in color, weight, size and so forth, but it excludes a much larger class of objects: "plaid shirt" will not match a car, a tie, the planet Mars, or even a striped shirt. The social pattern "Thanksgiving dinner" would, for most people growing up in the United States, invoke a very specific set of objects, events and social interactions and these would be distinctly different from the set of events invoked by the patterns "Christmas", "Passover" or "Fourth of July".

Coassociation is established through *repetition*: A cluster of attributes must be encountered repeatedly and can be expected to be encountered in the future. In political analysis patterns usually have a set of historical instantiations but they could also be counterfactuals—for example the patterns leading to the outbreak of nuclear war—established through stories or contingencies. In the post-WWII period, the pattern "U.S.-sponsored change of government" would include the historical examples of Guatemala 1954, Iran 1954, Grenada 1983, and Panama 1989, but in some contexts it might also include the counterfactuals Cuba 1961 and Iraq 1991.

This does not mean that all patterns are distinct. For example, the pattern "U.S.-sponsored change of government" *might* include the elections in Nicaragua in 1990. The U.S. at times claimed credit for this, and certainly approved of the outcome, but the techniques employed were quite different than the situation in Panama in 1989 or Kuwait in 1991. The classification of a case depends on how it compares to the universe of cases: If Nicaragua is compared to changes in Eastern Europe in 1989, then the US was actively involved; if compared to Grenada, Panama or Kuwait, the U.S. was not.

Coassociation allows one to fill in missing, hidden or noisy features on the basis of known features, and to identify obscure things from obvious ones. For example, when aerial surveillance of Cuba in 1962 revealed the construction of Soviet-style missile launching sites—conveniently constructed in Cuba in a fashion identical to sites in the Soviet Union—U.S. intelligence agencies inferred that the USSR intended to deploy missiles targeted at the United States. In this instance, easily obtained physical evidence was used in place of difficult-to-obtain evidence on Soviet policy. This inference involved patterns of Soviet Cold War policy as well as the physical patterns of missiles and missile sites; for example the concurrent Soviet construction of a soccer field in Cuba did not imply the impending deployment of Soviet soccer teams.

Concepts or *typologies* are simply a shorthand for coassociational groupings:

> [W]hat delimits a particular class of objects, qualities, actions or relations is not some sort of ideal example. Rather, it is a list of qualities. … the language of everyday life makes arbitrary, overlapping and less than all-inclusive divisions of experience. It is by means of such lists of qualities that we identify doors, windows, cats, dogs, men, monkeys and other objects of daily life. … Our lives do not present fresh objects and fresh actions each day. They are made up of familiar objects and familiar though complicated sequences of actions presented in different groupings and orders. (Pierce 1980, 119-122)

The political science lexicon develops concepts such as "death squad", "urban bias", "*dependencia*", "peacekeeping operations", "dirty float", and "multinational corporation" as it finds clusters of attributes and behaviors that are usefully grouped together.

Classification

Patterns are used to differentiate between cases. A *classification problem* is the task of assigning a case to one of a finite set of classes, which may include "unknown". While one does not normally think of human decision-making in terms of classification, a large number of "expert" decisions—for example almost all diagnosis, repair, configuration and approval tasks—can be formulated as classification problems.

The best-known classification problem in the AI literature is that of medical diagnosis: a patient has a variety of symptoms (e.g. fever, muscle cramps, headache, no cough) and needs to be classified into a specific disease category (e.g. cold, flu, food poisoning). A "disease" is simply a pattern of attributes dealing with the patient's health, and some of these features are useful in determining the future state of health (e.g. "Patient has a cold and will recover full in 7 days with treatment versus a week without treatment" or "Patient has acute appendicitis and without treatment will die").

Many political decisions are classification problems. For example, if country X appeals to country Y for assistance, choosing the appropriate response is a classification problem for Y. Decision-makers in Y will consider the features of the case, such as the history of past relations, the reasons for the request, and the policies of other actors in the international environment, and then choose a response, for example whether to send economic aid, arms, troops, aid through multilateral agencies, or no aid at all. The set of possible responses may be small—for example a Yes/No answer—or large. Typically a complex problem will be decomposed into a set of more specific problems: for example an aid decision might first involve the type of aid, then the amount, then the timing and side-conditions.

The utility of a feature in a classification problem depends on its cost, reliability and ability to discriminate between the categories to which a case might be assigned. The ideal feature provides a high level of discrimination can be acquired at low cost and is reliable. A feature is redundant—provides no new information—if replacing it with a single value makes no difference in the ability of the remaining features in the vector to classify. For example, in evaluating the capabilities of military hardware, one is unlikely to be interested in whether a tank can make carrier landings or whether a jet aircraft can ford streams.

In classification problems, even missing data may carry information useful. For example individuals who fail to answer survey questions are not a random sample of the population. In some contexts this could provide information. If individuals who refused to answer a survey question regarding their income level were also less likely to support tax increases, the missing value would still provide information on the support in a population for tax increases.

Problem-Solving and Prediction in the International System

Fundamental to any model of foreign policy decision-making is the task of prediction: anticipating the likely consequences of an action or absence of action.[5] Prediction is simply temporal coassociation of the current features of a case with its future features. The passage of time turns a prediction problem into a classification problem; in other words, viewed in retrospect, one can ascertain how information available at time *t* coassociated with behavior at some time *t+k*.

This predictive element of problem solving in foreign policy tends to be implicit, rather than explicit, in many formal models. Prediction is relatively straightforward in numerical models because the range of future numbers is typically bounded, and in any case are numbers. For example defense expenditures are usually substantially less than 100% of GNP; interest or economic growth rates rarely change more than a couple of percentage points from year to year.

In contrast, modeling discrete political choice is more difficult because the universe of possible events is so varied. The following is the list of the first 12 international events reported on Agence France Presse on 14 January 1992.

U.N. to consider membership applications of former Soviet republics
Ukraine and Russia to divide up Black Sea fleet this week
U.S. farm exports to Australia up sharply
Rover car firm considering production in Russia
Israeli and Palestinian negotiators hold their first formal session,
Arabs still divided a year after Gulf war
Italy, Austria said ready to recognize Macedonia
Jakarta to welcome a UN official but not probe team
Palestinians ask Japan to pressure United States
U.S. warns Pakistan to destroy components for two nuclear bombs
First visit by a Thai premier to Hanoi
Indonesia and Papua New Guinea sign defense accord

In the absence of additional information, every one of these events—and tens of thousands of others—is a *possible* consequence of, say, the USA accusing Iraq of hiding nuclear weapons material. However, a human analyst will typically focus on a tiny subset of those events. In this set, the only plausible events that might be a consequence of U.S. criticism of Iraq's nuclear weapons program are the similar criticism against Pakistan and possibly the statement about the Arabs being divided over Iraq.

More abstractly, consider the universe of equiprobable events generated in a system with 200 actors and a 60-category event-coding scheme. This produces 2,400,000 possible dyadic interactions with an entropy (see Chapter 5) of 14.7. Suppose in constructing an expected utility model, this choice space is reduced to a decision tree with 7 branches and an assigned probability distribution of [0.05, 0.1, 0.2, 0.3, 0.2, 0.1,

---

[5] Researchers who have done multiple, systematic studies on political problem solving include Voss (Voss et al 1983 1986; these studies cover general social science problem solving), Purkitt and Dyson (1986, 1987), and Boynton (1988); Purkitt (1991) provides an excellent review of this literature from a computational modeling perspective; Vertzberger (1990) provides an encyclopedic survey of these issues from the standpoint of foreign policy decision-making.

0.05], which has an entropy of 1.76. The construction of the model has reduced the complexity of the problem by 99.9997% in terms of the choice space, and 88% in terms of entropy. The modeler did the bulk of the reduction of uncertainty, not by the model.

Ideally a model of foreign policy decision-making should specify how the set of possible political outcomes is reduced to that much smaller set of likely outcomes without those outcomes being specified a priori. At present most existing formal models with qualitative domains, as opposed to numerical domains, do not do this. As noted in Chapter 2, while these models may be useful in illuminating some aspects of foreign policy problem solving, they miss most of the process.

Ill-defined problems

The political world is an "ill-defined problem" in the taxonomy established by Ashby (1956). In an excellent article reviewing such systems, Moray defines these as[6]

> …an ill-defined system is one whose state-transition matrix cannot be known, either because some states are inaccessible, or because some of the transition probabilities are inaccessible, or because the matrix is not time-invariant. … The list of variables in the state vector may not be constant, new variables may appear, old ones disappear, and the transition probabilities may alter from time to time. … Finally, it should not be overlooked that the ill-defineness of an observed system may be the result not of chance but of intelligent hostility. (Moray 1984, 11,14)

Ill-defined problems present a substantial change challenge to traditional mathematical theories of decision and control (see Selfridge, Rissland and Arbib 1984). Mathematical models usually deal with systems that are well-defined. The control mechanisms of an automobile engine function consistently because variables such as the ratio of fuel to air and the timing of ignition sparks have a predictable effect; the operation of the machine can therefore be optimized. Similarly, a space probe can be sent millions of kilometers to rendezvous with a planet because the laws of gravity do not change, and the system can be described using a few numbers dealing with mass, position and velocity. Most of the commonly modeled physical systems—and microeconomic systems—are well defined: that is why they are studied with models.[7]

Early researchers attempting to model political systems hoped that such systems would also be sufficiently well defined that simple mathematical methods could be used. For example, this belief presumably accounted for the emphasis on cybernetic approaches in the work of early behavioralists such as Easton, Deutsch and Almond, and certainly

---

[66] Also compare this to Almond's discussion of "the ontological properties of politics":

> [politics is] not readily amenable to cause-and-effect 'clocklike' models or metaphors. Basically, this is because the behavioral repertoires of elites and citizens are not fixed. The actors in politics have memories; they learn from experience. They have goals, aspirations, and calculative strategies. Memory, learning, goal seeking and problem solving intervene between 'cause' and 'effect, between independent and dependent variable. (Almond 1990,35)

[7] For example, while a stock market is complex and has stochastic elements, its state-space is numerical and its transition matrix is sufficiently predictable that computerized trading programs can operate effectively.

accounted for the attraction of economic analogs to rational choice modelers. In the international system, however, the ill-defined aspects of the system—starting with the pervasiveness of Moray's "intelligent hostility"—tended to overwhelm the well-defined aspects.

Yet while traditional optimization methods fail on ill-defined systems, humans are remarkably successful in dealing with them:

> Despite [all of their cognitive failures] it is clear that humans do, given enough practice, frequently manage rather well to control systems which all the evidence suggests they should not be able to. [By] interacting with an effectively ill-defined [system] a human is able (to adopt the language of optimal control theory for an example) not merely to adjust his gain matrix, but actually to identify the state variables, construct the gain matrix, construct the predictor and even construct the appropriate estimator—and top change them all when the ill-defined [system] changes its properties. (Moray 1984, 18-19)

Ideally, a formal model of foreign policy decision-making should be able to identify the credible consequences of an action or a sequence of actions, since the task of prediction is inherent to determining the outcomes of various policy options. This is a very difficult and in many ways unsolved problem. Human analysts, nonetheless, are able to solve it to the extent that complex political activity is possible.

Short-Term Set Prediction

Political analysts are usually not concerned about point prediction—figuring out precisely what will occur. Instead, political prediction deals with reducing the very large set of *possible* future events to a much smaller set of *plausible* events. For example, in Schrodt (1987a) I give an example of a set of predictions generated by a single international event, the 1985 US diversion of an Egyptian aircraft carrying individuals accused of hijacking the Italian cruise ship *Achille Lauro*. These predictions involved such things as the diplomatic responses of Egypt, Italy and the PLO, the involvement of Italy and Egypt in the US interception (an event that had already occurred but which was unknown to me at the time), and the eventual fate of the accused hijackers in Italy. These predictions proved accurate and were typical of the type of predictions that any political scientist familiar with the situation might have made.[8]

Point prediction in politics is unnecessary because the individual or organization can prepare for multiple contingencies simultaneously: for example when playing chess, one doesn't prepare for an attack on a single piece, but against multiple possible attacks on multiple pieces. Circumstances also exist where it is rational for an opponent to play randomly; random behavior can also arise from vagaries in health, complex bureaucratic

[8] I have noticed that in the media, the role of the political scientist is usually to deal with the future: historians deal with the past, journalists with the present, and the political scientist is asked "what happens next?".

interactions, natural disasters, equipment failure, communications breakdowns and other circumstances.[9].

Predictions of international behavior tend to be short term rather than long term for two reasons. First, the branching of contingencies, particularly those with random components, causes the set of credible outcomes to expand exponentially with time.[10] The human brain, unable to cope with an exponentially increasing set of options, drastically trims this set, often using rules that attempt to bound the true outcome with "best possible case scenarios" and "worst possible case" scenarios. Second, human planning in the competitive situations typical of politics is continually and deliberately disrupted by the actions of opponents. Thus while one may have a long-term plan in mind, only rarely is it fully implemented. These factors lead to the primacy of short-term planning, and short-term regularity, over long term planning. Long-term forecasting can still occur, but with considerably less detail concerning specific actions and the timing of actions than short-term forecasting. Day-to-day decision-making is relatively myopic.

In contrast to the extensive attention paid to the accuracy of economic[11] and election forecasts (Gelman and King 1992; *The Political Methodologist* 5,3 (Summer 1994)), few formal tests are available in the published literature on the accuracy of international relations forecasting. Jensen (1972) tested the predictive abilities of 171 foreign policy experts (categorized as being journalists, State Department, Defense Department, and academic) on a set of 25 questions.[12] Answers were a loose set of

---

[9] During the 1980s, a favorite term paper topic in my U.S. Foreign Policy class was predicting the future of the Clark Air Force Base in the Philippines. None of the dozens of carefully researched analyses correctly predicted the actual outcome: the base was abandoned after being buried in ash by the Mt. Pinatubo volcano.

[10] From Thomas Pynchon's novel, *Gravity's Rainbow*:

> It occurred to him to focus on the European balance of power. … He started in on a mammoth work entitled *Things That Can Happen in European Politics*. Begin, of course, with England. "First," he wrote, "Ramsay MacDonald can die". By the time he went through resulting party alignments and possible permutations of cabinet posts, Ramsay MacDonald had died. "Never make it," he found himself muttering at the beginning of each day's work, "it's changing out from under me. Oh dodgy, very dodgy". (Pynchon 1973,77)

[11] See for example the numerous articles on this topic in the *Journal of Business and Economic Statistics*. *The Economist* (13 June 1992,75), quoting a study by economist Victor Zarnowitz of the University of Chicago, notes an average improvement in the mean absolute error of forecasts of about 20% in U.S. Treasury forecasts of GDP, inflation and the current account in the coming year when comparing 1979-84 to 1985-91 (for example, 1979-84 predictions for March inflation averaged 1.4%; those for 1985-91 averaged 1.2%), though it is not clear how much of this is due to improvements in the models versus changes in the economy.

In a less systematic study, in 1984 *The Economist* (3 June 1995,70) also queried four former finance ministers of OECD countries, four MNC CEOs, four Oxford students and four London garbage collectors about economic trends for the 1985-1994 period. The predictions were not particularly accurate and as expected most errors could be linked to extrapolating trends from the early 1980s. As to the impact of expertise, the finance ministers as a group were the least accurate; the CEOs and garbage collectors tied for the most accurate groups.

[12]A total of 51 questions were asked but many were contingencies—"If war between X and Y breaks out, then…"—that did not obtain so the accuracy of the prediction could not be evaluated.

categories, such as “unlikely”, “relatively likely” and “very likely”, which depended on the question. The overall success rate was about 64%, which is considerably better than chance but hardly outstanding. The Jensen study dealt with predictions on the order of one to five years, rather than with the shorter-term predictions, where I suspect the accuracy would be higher.

In an informal exercise at the end of September 1990, I had the undergraduate students in my international conflict class make detailed predictions about the state of the Iraq-Kuwait crisis as of the final day of class in mid-December.[13] A large majority of the students (and the instructor) got the prediction wrong, expecting the outbreak of armed hostilities in late October or early November. However, four students—about 10% of the class—predicted the situation in December almost exactly, including the use of the United Nations, the continued allied buildup, and the coalition-building activities of the Bush administration. Short-term prediction is definitely possible.

The literature on strategic deception actually provides a fascinating “exception that proves the rule” on the importance of patterns in predicting political behavior.[14] A successful deception involves the elaborate creation of a “world” that makes sense to the organization being deceived. Whaley, summarizing the reasons for the success of the Barbarossa deception that preceded the German invasion of the Soviet Union in 1941, concludes:

> I suddenly realized that the Wohlstetter model [of strategic surprise being due to noise] was a quite inappropriate representation of BARBAROSSA surprise. Stalin (and almost everyone else) had been surprised not because the warnings were *ambiguous* but precisely because German intelligence had managed to *reduce* their ambiguity.... [Hitler’s] cunning “ultimatum” stratagem served to eliminate ambiguity, making Stalin quite certain, very decisive, *and wrong*.... Stalin’s false expectation was the direct effect of Hitler’s campaign to manipulate his victim’s information, preconceptions, and decisions. (Whaley 1973,242; italics in original)

Wohlstetter’s (1962) study of the successful Japanese surprise at Pearl Harbor emphasized the signal/noise problem in information processing, whereas in a strategic deception, the signal of the adversary’s actual intentions is hidden not through noise or secrecy, but by an alternative explanation for events.

For example, part of the deception in Barbarossa involved convincing the Soviets that German troop movements were part of a plan to invade Britain. In reality, that plan had been abandoned months earlier. To further confuse matters, the Soviets were led to believe that Barbarossa itself was a deception. directed against the British, protecting German plans to invade Britain (Whaley 1973, 173). In the successful “Bodyguard” deception plan preceding D-Day the deception was fine-tuned to the knowledge and pattern recognition of the German intelligence agencies, which the British could monitor because they had broken the German communications codes.

---

13 See Kugler, Snider and Longwell (1993) for a more systematic analysis of real-time predictions in this crisis.

14 The well-studied examples include Germany's Operation Barbarossa in WWII (Whaley 1973), the Allied deception of Germany prior to the D-Day invasion (Brown 1975), and the Egyptian-Syrian deception prior to the 1973 October War (Handel 1976)

Finally, strategic deception was effectively used during Operation Desert Storm in convincing Iraq that the U.S. would attempt an amphibious landing from the Persian Gulf. In addition to several well-publicized practice landings, a fine twist to the deception were leaks to the media reminding them of the inter-service cooperation problems encountered by United States forces during poorly-coordinated invasion of Grenada. The media, Iraqi military (and this author) were thus convinced that the Marine Corps would "insist" on an amphibious operation. In fact, the Goldwater-Nichols Act in 1986 had eliminated the coordination problems, no amphibious operation took place, and Iraqi defensive preparations on the Kuwaiti coast were wasted. Egypt and Syria used a similar strategy against Israel in 1973, playing on Israeli assumptions that Arab states would never be able to politically coordinate a joint attack.

Ironically, predictions can structure political behavior irrespective of their accuracy. Because planning requires prediction, the complexity of political behavior is constrained by its participants' ability to predict. From a purely statistical standpoint, the political world is very predictable and adjusts itself to reduce uncertainty.

We perceive a very uncertain political world because we disproportionately notice its uncertain aspect. In fact, we experience a truly *random* world only in nightmares, bad *avant garde* films, postmodern exegesis and after ingesting controlled substances. Imagine, for example, a mob of pitchfork-wielding peasants marching on a government building and pulling down a fence. Ah, a violent demonstration, we conclude. But the peasants take the fence posts, build fires, and roast sausages using the pitchforks. Aha, a picnic. But from the building come a blare of trumpets and a young man wearing a tuxedo appears beside a young woman in a long white gown. Hmm, well, maybe this is a wedding. Seeing this, the peasants put down their sausages, level their pitchforks at the couple and rush forward. Well, maybe it's a riot after all, with a break for lunch. But then the woman raises her hand and the peasants stop, form large circles and dig frantically at the pavement trying to plant turnips.

Such a scenario reads like a bad dream, perhaps induced by a dinner of sausages and turnips. But even this sequence is far from truly random—from my brief description, the reader has probably invoked an elaborate mental image using associated patterns of social behavior. For example, how wide is the street? What are the peasants wearing? What does the facade of the building look like? Does the gown have lace and a veil? I supplied none of these details but your mental image probably contains them.

"Random" is not the same as "unexpected" When we speak of "anarchy", "disorder" and "unpredictability" in social behavior, we are rarely referring to true randomness but instead to transitions among highly patterned modes of behavior. A "riot" is a form of regularized behavior with predictable characteristics (e.g. shouting, large crowds, looting, attacks on police, etc.) as much as a "parade" is a regularized behavior. The amount of information required to specify a riot is probably similar to that required to specify a parade: the rules for proper behavior in a riot (that is, riotous behavior) are merely different from those governing behavior in a parade.

Pattern Recognition

> *It is becoming increasingly apparent that biological systems are much more complex than the technological systems usually considered by the control engineer. A technological system is usually designed on the basis of predesignated criterion of stability and response which are expressed in some analytical form. Physiological systems, on their other hand, have evolved slowly, continuously, adapting the performance of specific tasks to a wide variety of conditions. ... Natural selection is not saddled with an expediency demand. The evolution of physiological control systems might, therefore, be expected to result in optimal systems chosen with the complexity of description not at all entering as a limiting factor.*
>
> *B. Pennock and E.O. Attinger.*

Pattern recognition is the ability of an individual to consider a complex set of inputs, often containing hundreds of features, and make a decision based on the comparison of some subset of those features to a situation that the individual has previously encountered or learned.[15] Pattern recognition is useful in the analysis of politics because certain sequences of political behavior occur on multiple occasions under relatively predictable circumstances. In problem solving situations, *recall can substitute for reasoning*. For example, chess involves a well-defined, entirely deterministic system and should be solvable using purely logical reasoning. Chess-playing computers use this approach, but Chase and Simon (1973) found that human expert-level chess playing is done primarily by pattern recognition.

Humans possess very large, albeit imperfect, long-term memories. Failures in the fidelity of this memory are compensated by the fact that it is associative: we can recall a large amount of information from a small number of features, even in the presence of noise. Information about the fruit "apple" can be invoked by a smell, a taste, a variety of objects, a variety of words (e.g. "apple", "Red Delicious", "Winesap", "cider") and a variety of social memories, as well as by the noisy stimuli "Aqple" or the rainbow-colored trademark on an Apple computer.[16]

Pattern recognition is dependent in part on what the decision-maker is already focusing on and the extent to which information fits with patterns already in memory; this source of misperception is a major emphasis in the foreign policy decision-making studies of Jervis (1976) and Khong (1992). Vertzberger, summarizing a very large literature in psychology[17], observes

---

[15] The literature on pattern recognition in human problem solving goes back to the 1960s, for example Newell and Simon 1972, Simon 1982, Margolis 1987. A substantial literature on pattern recognition also exists in the AI literature (for example, Fu 1974, 1982; Patrick and Fatu 1986; Devijver and Kittler, 1987), though the bulk of the AI literature on pattern recognition deals with *spatial* patterns—for example distinguishing a tank from a truck—rather than temporal patterns.

[16] [2nd edition]: The *original* Apple Computer logo consisted of rainbow-colored stripes, a motif intended to remind the viewer that the earliest Apple][computers were capable of driving a color video monitor. When Apple updated its industrial design in the late 1990s, the logo shifted to a solid, pearl white.

[17] I've not included Vertzberger's references in the quote; see the original.

> The contexts of prior knowledge or expectancies also generate salience. … Information that has already gained attention in the past and penetrated the perceiver's cognitive system continues to attract attention,… even if its information value has declined.…
>
> Information consistent with expectations is better remembered and more accurately rated than inconsistent information. … In particular, information that fits into existing schemata—that is, cognitive structures of organized prior knowledge abstracted from specific experiences—is noticed earlier, considered more valid, and processed much faster than information that contradicts or does not fit into any particular schema. (Vertzberger 1990,60)

Kahneman, Slovic and Tversky (1982) also deal extensively with these issues; a variety of formal models for this type of memory have been developed, particularly in the work of Abelson (1973), Anderson (1983) and Kohonen (1984).

International relations theory has begun to reemphasize the importance of non-statistical patterns. The "international institutions" and "regimes" literature (Krasner 1983) is quite explicit on this point. Robert Keohane, in his 1988 presidential address to the International Studies Association, notes

> ..."institution" may refer to a *general pattern* or *categorization* of activity or to a *particular* human-constructed arrangement, formally or informally organized. Examples of institutions as general patterns of behavior include Bull's "institutions of international society" as well as varied patterns of behavior as marriage and religion, sovereign statehood, diplomacy and neutrality. ...What these general patterns of activity have in common with specific institutions is that they both meet the criteria for a broad definition of institutions: both involve persistent and connected sets of rules (formal or informal) that prescribe behavioral roles, constrain activity, and shape expectations. (Keohane 1988,383)

While neither Keohane nor most of the international institutions literature have provided formal definitions of these patterns, their emphasis on the importance of pattern in international behavior is unmistakable.

Recall is preferred to reasoning because working memory[18], which must be utilized in deductive reasoning, is slow and constrained to handling only a few items of information at a time. The long term memory used in pattern recognition, in contrast, is effectively unlimited in capacity[19] and works very quickly—on the order of seconds—even when solving a complex associative recall problems across thousands of potential matches.[20] Purkitt notes:

---

18 Earlier known as "short term memory".

19 See Newell and Simon (1972), chapter 14. Newell and Simon argue that the capacity of associative memory is effectively unlimited because the amount of time required to store items is sufficiently long that life-span, rather than memory capacity, is the constraint.

20 As noted in Chapter 1, it seems that the brain actually invokes a pleasure response when solving associative recall problems. The two most popular televised game shows in the United States during the 1980s and 1990s were *Jeopardy* and *Wheel of Fortune*: both are associative recall games. Crossword

> Generally speaking, the power and complex of human cognition is tied to the almost unlimited capacity of humans to store and combine vast amounts of information in long-term or associative memory. … Research has also demonstrated that the active processing of information is a serial process within the limited capacity of working memory. In moving from the level of pieces of information to the level of factors or indicators, it is now clear that individuals can only systematically process information on a few (probably two or three) factors without explicit cognitive aids (e.g. algorithms). (Purkitt 1990,40)

Associative memory is vast, effortless and quick; logical processing is limited, painful and slow. Consider the following four questions:

- Describe three uses of military force by the United States during the 1980s
- Who was attorney general during the 1963 Cuban Missile Crisis?
- What is 15% of $22.30?
- Prove the Pythagorean Theorem.

The answer to the first question will come to the international relations specialist more or less immediately "from memory". However, the method by which the answer was determined cannot be described—for example, were all instances of military force in memory searched, all actions by the United States, or all international events in the 1980s? One cannot say; this is instead done using "subcognitive processing", discussed below. The answer simply appeared, without conscious operations, in a couple of seconds. Similarly, the second question can be answered quickly despite being stated in a factually inaccurate manner.[21]

In contrast, most people can articulate the algorithm used to solve the third question. This may be general-purpose schoolbook multiplication ("multiply 2 by 5, carry the one…") or a specialized algorithm developed because the problem is commonly encountered when calculating restaurant tips ("divide the total by ten by shifting the decimal point to the left, then divide that by two, then add these two amounts"). Failing these, one can solve the problem on a calculator, and in any case the manipulation of the information can be verbalized without difficulty. The final problem involves the logical manipulation of only a few axioms from plane geometry, virtually every literate person has learned the proof in high school geometry, and yet its solution is difficult for most people.

The latter two problems are far less information intensive than the first two—this is why the third can be solved on a calculator—but require deductive processing. In fact,

---

puzzles fall into the same category; as does the board game *Trivial Pursuits*. All of these games involve the rapid recall of specific information out of a huge memory on the basis of partial cues.

It may not be coincidental that these games gained popularity during an "Information Age" when information became available in surplus. Demonstrating skill at *Trivial Pursuits* or *Jeopardy* may be the contemporary equivalent to county fair competitions of physical strength such as wrestling and weight-lifting in an earlier age when physical strength was economically important.

21 Robert Kennedy, and the crisis was in 1962, not 1963. The fact that the attorney general was the President's brother and actively involved in the crisis aids in the recall; I suspect most people could not answer the same question for the U-2 crisis.

the first two problems require a very large amount of historical information and complex associative links. The first problem could be solved using a large electronic database such as NEXIS but constructing a query that duplicates only the three examples usually retrieved by experts (Grenada, Lebanon and Panama) is quite difficult. A slightly more difficult query, e.g. "Indicate three changes in the NATO alliance between 1952 and 1972" becomes almost impossible even for NEXIS. But such questions are nothing more than typical college examination questions and barely worthy of consideration as expert political knowledge.

Origins of Pattern Recognition

In all likelihood, the human brain evolved with a strong bias towards pattern recognition rather than deductive reasoning. This natural environment is comprised of two systems: the physical and the biological. Many aspects of the physical world can be usefully described by deductive axiomatic systems, and an information-processing system operating solely in a law-governed world would be able to survive with purely deductive reasoning; examples would include computer viruses and programmed trading systems.

The biological world, in contrast, is exceedingly complex and arbitrary. It is a world of individuals constructed from complex feature vectors made of DNA, with billions of components, and selected solely by the ability of their ancestors to reproduce, oftentimes in unusual circumstances such as the aftermath of asteroid collisions. Such a world cannot be described deductively in any practical sense, but because it is very repetitive, pattern recognition is an effective information-processing strategy. If one Tyrannosaurus Rex tries to devour you, the next one is likely to as well. Since critical decisions must be made in real time ("Is the object approaching me sometime I can eat, something that will eat me, or something I can ignore?"), evolution will select for high recall speeds under noisy environmental conditions. It does not select for theorem proving or the minimization of quartic polynomials.

This neural bias would emerge early in the biological record, well before the development of primates, or mammals, or even vertebrates. *Homo sapiens sapiens* is endowed with sophisticated pattern recognition capabilities honed through eons of evolution, and it is unsurprising that this capacity is put to use in social behavior. Deductive reasoning, in contrast, is a comparatively recent development and is much more difficult. While we are very proud of deductive reasoning, it is not necessarily more useful, particularly when dealing with social behaviors that may also have some evolutionary roots.

Anderson and Rosenfeld trace the pedigree of this idea to William James:

> As James points out [in *Psychology (Briefer Course)* (1890)] emphatically in several places, the brain is not constructed to think abstractly — it is constructed to ensure survival in the world. … [The design principles are:] do as good a job as you can, cheaply, and with what you can obtain easily. If this means using ad hoc solutions, with less generality than one might like, so be it. We are living in one particular world, and we are built to work in it and with it. (Anderson and Rosenfeld 1988, 1)

Pattern recognition, unlike deduction, is easy. Purkitt (1991) discusses research on the variety of ways that the limitations of working memory cause decision-makers to take cognitive shortcuts; these in turn affect communication and policy formulation.

An important consequence of the survival value of pattern recognition is a brain is biased in favor of recognizing, rather than rejecting, patterns. As Hugh Kenner (*Byte*, November 1989,486) puts it, “The computer is a difference detector. The human mind is a similarity detector.” The survival costs of fleeing in terror from a dimly perceived and ultimately nonexistent threat are substantially lower than the risks of not fleeing a genuine threat. The ability of the brain to perceive patterns in random data is the bane of statisticians and the true salvation of TV evangelists, but arises quite naturally from the necessities of survival in a noisy environment.

Substantial parts of the brain are specialized for the social tasks of recognizing faces and that most cognitively complex of all social interactions, language. It would not be surprising if the brain had in addition some specialized hardware for handling at least some basic political interactions, for example the social hierarchies present in many vertebrates. Human associative memory may be able to handle, subcognitively, complex episodic political information such as precedent retrieval in part because the brain evolved in part to handle comparable problems.[22]

Subcognitive Processing

Studying associative recall is problematic because the process occurs in the non-verbal, unconscious or subcognitive[23] part of the brain: it is a form of information processing that we can do but not articulate. In the foreign policy field such reasoning is typically called “intuition” or “feel.” A typical example of this approach to foreign policy analysis is found in the following quote from Kissinger:

> Statesmanship requires above all a sense of nuance and proportion, the ability to perceive the essential among the mass of apparent facts, and an intuition as to which of many equally plausible hypotheses about the future is likely to prove true.(Kissinger 1979,39)

Bull’s defense of the classical method states that the core of that approach is “that general proposition about [international politics] must derive from a scientifically imperfect process of perception and intuition” (Bull 1966,361).

Subcognitive information retrieval involves nothing mystical; the process can be, and has quite extensively been, empirically studied (see Collins and Smith 1988; Reber 1993). For example, Gilovich presented Stanford political science undergraduates with a

---

22 Masters (1989) provides a thorough discussion of the possible connections between evolution and political behavior; Axelrod (1984) and Simon (1990) note that evolution may have predisposed humans to altruistic behaviors, a definite change for the bellicose "Social Darwinism" of a century ago.

23 This term is from Douglas Hofstadter (1985) via Holland et al (1986). The word is attractive since unlike "unconscious" it implies active information processing; it avoids the Freudian overtones of "subconscious" and it is more general than the term "nonverbal". Jackendoff (1987), while dealing with an entirely different set of domains, provides an excellent discussion of subcognitive information processing and a guide to much of the relevant psychological literature in the linguistic and visual perception domains; Springer and Deutsch (1981) give a semi-popular review of the related literature on split-brain experiments.

hypothetical crisis and asked them to make a decision about intervention. Some of the students were provided cues to suggest that the crisis was similar to Munich; others were cued to a similarity to Vietnam; a third group was provided neutral cues. As expected, students cued to make the Munich analog "recommended significantly more interventionist strategies than subjects in either the Vietnam or the neutral groups." (Gilovich 1981,805). However, the groups did not significantly differ when *asked* about the similarity of the hypothetical crisis to Munich or Vietnam:

> Thus, though the subjects made recommendations consistent with specific historical episodes, they were unaware of the influence that these episodes apparently had on their decisions. This raises a very interesting issue, one in need of further research, concerning the different levels of awareness at which this process [of comparison] may operate. (Gilovich 1981, 806)

One can engage in very complex information processing without being aware of how one is doing it, *even using introspection*. Lashley notes:

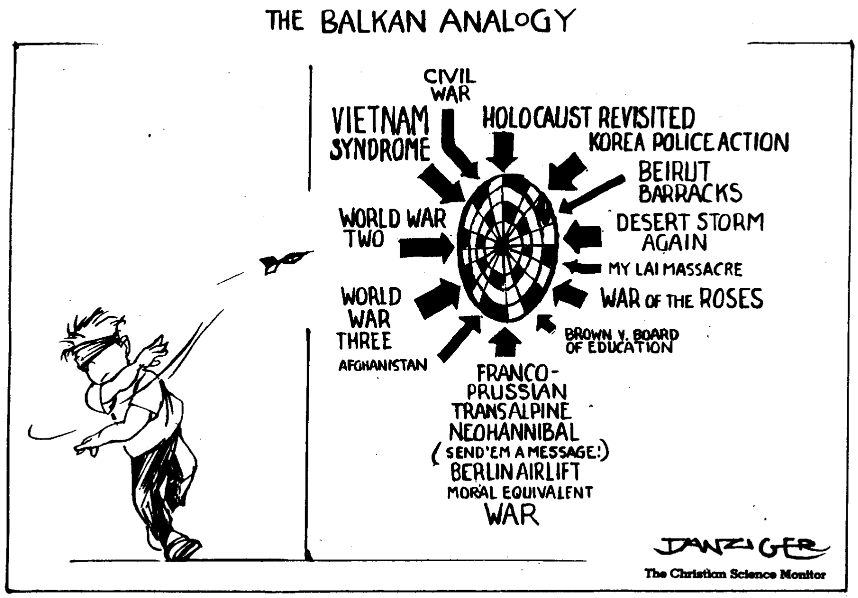
> *No activity of the mind is ever conscious.* [Lashley's italics] This sounds like a paradox, but it is nonetheless true. A couple of illustrations should suffice. Look at a complicated scene. It consists of a number of objects standing out against an indistinct background: desk, chairs, faces. Each consists of a number of lesser sections combined in the object, but there is no experience of putting them together. The objects are immediately present. When we think in words, the thoughts come in grammatical form with subject, verb, object, and modifying clauses falling into place without our having the slightest perception of how the sentence structure is produced. (Lashley 1956,4; quoted in Jackendoff 1987,45).

As Lashley points out, the most obvious example of subcognitive processing is language. An illiterate speaker with no formal linguistic knowledge can continuously and effortlessly assemble grammatically correct sentences consistent with a grammar of several dozen, if not several hundred, rules. This skill is learned entirely by example and can be observed in any average three-year-old child, in any culture. One can also construct grammatically correct sentences through rules—as is done when one is learning a second language or in a machine translation system—but this is slow, awkward and not characteristic of fluency.

Wittgenstein notwithstanding, a long series of empirical experiments demonstrate that one can know things one cannot say, where "know" is defined as the successful completion of complex information-processing tasks and "say" means verbalize. One of the consistent lessons from the machine learning literature is that verbal problem solving protocols provided by human experts are usually much more complicated than the set of rules logically sufficient to solve the problem. An expert usually perceives that much more information is required to solve a problem than is logically needed, and when asked why will be unable to describe what that additional information provides other than "feel".

A likely explanation for this is that the expert actually does most of the problem solving using associative pattern recognition and therefore cannot articulate the process. The verbal protocol represents what the expert *thinks* he or she is thinking, rather than providing an actual description of the problem-solving process. Protocols are very useful in determining the *information* that an expert uses, but they cannot provide an accurate step-by-step description of the information *processing*.

Knowledge Representation



Christian Science Monitor 5 May 1993

Because the state-space of political behavior is so complex, political analysis is a very information-intensive process. The information in used international politics is of three major types:

- *Declarative knowledge*: Static facts, definitions, and associations. Declarative knowledge deals with "what is" from the standpoint of the decision-maker;
- *Stories, precedents and sequences*: Ordered sequences of events, usually with an historical instantiation. These guide long-term behavior and can be used to infer the behavior of others. Sequences are strategic: "what to do";
- *Rules and heuristics*: *If...then* rules. These deal with short-term behavior but do not provide any long-term structure. Rules are operational: "how to do it"

There is some overlap in the categories, in particular a story can be seen as a very complicated heuristic, and an historical instantiation of a story is a form of declarative knowledge. Rule and stories are often stored as general "templates" that use the declarative knowledge and "substitution principles" to match specific sequences of events. Declarative knowledge and stories will be discussed in this section; rules will be discussed in the context of organizations and in Chapter 4.

Declarative Knowledge

Declarative knowledge is primary factual, linking specific actors to general concepts. Declarative knowledge is the sort one finds in an encyclopedia or almanac:

> The Republic of Niger is ruled by a military President and Supreme Military Council, advised by a council of 20 ministers. Islam is the dominant religion; French is the official language, with Hausa and local languages widely spoken. … Economically Niger depends on agriculture and mining; about 90% of its workforce are herders or farms; the country is almost self-sufficient in food. (Stonehouse 1985,234-235)

In addition to this "book knowledge", decision-makers use a great deal of "common sense" knowledge about the physical world: for example the statement "There is a drought in Niger" plus the information above would lead one to conclude that its food imports would increase, but this inference requires the additional knowledge that agriculture requires rain, human beings require food and that in the absence of rain a country dependent on agriculture would need to find alternative sources of food, and food is a commodity traded across international borders. Similarly, lack of concern over the statement "There is a drought in Kuwait" would require the additional knowledge that Kuwait is a city-state at the edge of a desert and is not dependent on agriculture.

Stories and Precedents

A "story" or, more formally, an event sequence, is a set of temporally ordered events and an associated context or set of preconditions. These are usually based on simplified versions of history, though some are based on counterfactuals (see Fearon 1991). Stories are easily transmitted and stored by individuals and the use of stories is universal in human culture. Whether sitting around the dying embers of a Neolithic campfire or sitting in the departure lounge of an airport, humans find relaxation in a good yarn.[24]

The use of stories as a means of knowledge representation is strongly associated with the work of Roger Schank (e.g. Schank and Abelson 1977; more recently Schank 1990):

> The form of memory organization upon which our arguments are based is the notion of episodic memory…organized around personal experiences or episodes rather than around abstract semantic categories.… [O]ne of the principal components of memory must be a procedure for recognizing repeated or similar sequences. When a standard repeated sequence is recognized, it is helpful in 'filling in the blanks' in understanding. (Schank and Abelson 1977, 18)

---

24 Listening to stories and story-telling appears to be one of the dominant forms of human social interaction: an alien anthropologist might conclude that humans exist primarily as information transmission devices. Monitor the lunch-time conversation among professionals and observe just how much time is spent relaying stories. For example, I'm writing this after final exams, and for the past two weeks the departmental lunch-time conversations have been taken up primarily with (a) plagiarism and cheating schemes; (b) idiosyncratic excuses that students present when requesting incomplete grades. Topic (a) is pure information exchange, as one obtains an extensive repertoire of techniques and the means of detecting them, much as bacteria exchange snippets of DNA that provide resistance to antibiotics. Topic (b) is usually given with a request for a normative evaluation—"Now what would you do in this situation?"—but is transmitted through a story, just as the typical student excuse is a story ("I have this uncle in Wichita who I'm real close to and last week…"). Much of the memory and culture of an organization exists in its stories.

Common story patterns are assigned to general categories, for example "crisis", "war", "coup", or "revolution". At the lowest level of aggregation, the elements of a story sequence are events—interactions that can be described by transitive verbs—but typically stories are constructed hierarchically, with complex sequences can be built out of simpler subsequences. Conversely a large amount of detail can be compressed into a few statements by absorbing the details into commonly known patterns.

For example, the Cuban Missile Crisis could be described in very general terms by the sequence:

USSR builds missile launchers in Cuba
USA discovers missile launchers
USA blockades Cuba
USA and USSR negotiate
USSR promises not to deploy missiles in Cuba
USA promises not to attack Cuba

This rendition is very simple but still sufficient to distinguish the Cuban Missile Crisis from, say, Desert Storm or the SALT negotiations. The event "USA blockades Cuba" could be expanded to

President Kennedy convenes Executive Committee of the National Security Council
ExComm considers six possibilities: do nothing, bomb, invade and blockade, negotiate internationally, negotiate with Castro.
"Do nothing" option is rejected because…
"Bomb" option is rejected because…
and so forth.

In the international conflict literature, Lebow's "justification of hostility crisis" provides an example of a general episodic structure.

1. Exploit a provocation to arouse public opinion.
2. Make unacceptable demands upon the adversary in response to this provocation.
3. Legitimize these demands with reference to generally accepted international principles
4. Publicly deny or understate your real objectives in the confrontation.
5. Employ the rejection of your demands as a *casus belli*. (Lebow 1981,29)

Lebow develops this sequence using crises such as the Austria/Serbia in 1914, Japan/China in 1931, Germany/Poland in 1939 and USA/Vietnam in 1964. The sequence also fits nicely the actions of Iraq towards Kuwait in the summer of 1990. Table 3.1 shows the relevant Reuters headlines and lead sentences for the events that match the justification of hostility sequence:

Table 3.1. Justification of Hostility Crisis: Iraq/Kuwait 1990

July 17, 1990: RESURGENT IRAQ SENDS SHOCK WAVES THROUGH GULF ARAB STATES
Iraq President Saddam Hussein launched an attack on Kuwait and the United Arab Emirates (UAE) Tuesday, charging they had conspired with the United States to depress world oil prices through overproduction.

July 23, 1990: IRAQ STEPS UP GULF CRISIS WITH ATTACK ON KUWAITI MINISTER
Iraqi newspapers denounced Kuwait's foreign minister as a U.S. agent Monday, pouring oil on the flames of a Persian Gulf crisis Arab leaders are struggling to stifle with a flurry of diplomacy.

July 24, 1990: IRAQ WANTS GULF ARAB AID DONORS TO WRITE OFF WAR CREDITS
Debt-burdened Iraq's conflict with Kuwait is partly aimed at persuading Gulf Arab creditors to write off billions of dollars lent during the war with Iran, Gulf-based bankers and diplomats said.

July 24, 1990: IRAQ, TROOPS MASSED IN GULF, DEMANDS $25 OPEC OIL PRICE
Iraq's oil minister hit the OPEC cartel Tuesday with a demand that it must choke supplies until petroleum prices soar to $25 a barrel.

July 25, 1990: IRAQ TELLS EGYPT IT WILL NOT ATTACK KUWAIT
Iraq has given Egypt assurances that it would not attack Kuwait in their current dispute over oil and territory, Arab diplomats said Wednesday.

July 27, 1990: IRAQ WARNS IT WON'T BACK DOWN IN TALKS WITH KUWAIT
Iraq made clear Friday it would take an uncompromising stand at conciliation talks with Kuwait, saying its Persian Gulf neighbor must respond to Baghdad's "legitimate rights" and repair the economic damage it caused.

July 31, 1990: IRAQ INCREASES TROOP LEVELS ON KUWAIT BORDER
Iraq has concentrated nearly 100,000 troops close to the Kuwaiti border, more than triple the number reported a week ago, the Washington Post said in its Tuesday editions.

August 1, 1990: CRISIS TALKS IN JEDDAH BETWEEN IRAQ AND KUWAIT COLLAPSE
Talks on defusing an explosive crisis in the Gulf collapsed Wednesday when Kuwait refused to give in to Iraqi demands for money and territory, a Kuwaiti official said.

August 2, 1990: IRAQ INVADES KUWAIT, OIL PRICES SOAR AS WAR HITS PERSIAN GULF
Iraq invaded Kuwait, ousted its leaders and set up a pro-Baghdad government Thursday in a lightning pre-dawn strike that sent oil prices soaring and world leaders scrambling to douse the flames of war in the strategic Persian Gulf.

Source: Reuters

Stories are generalized into ideal cases that I will call “templates”. When a decision-maker refers to the danger of a coup in El Salvador, this usually refers not to a specific coup but coups in general[25]. In order to apply a template to a specific case, a decision-maker uses substitution principles in combination with historical or idealized sequences of international events to create analogies:

Analogy = precedent + substitution principles

Substitution principles are primarily based on declarative knowledge about the actors involved—for example does the actor have allies; is it a major power; where is it located—though they may also involve contextual knowledge about the historical circumstances of the story (for example, did the story occur before, during or after the Cold War period).

Consider the template

[Tension between X and Y]
[Political instability in X]
[Y invades X]
[X consolidates power and repels Y’s invasion]

If X=Iran and Y=Iraq, this describes the initial phases of the Iran-Iraq War circa 1980; if X=France and Y=assorted European monarchies it describes Europe circa 1790; if X=Russia and Y=assorted capitalist states it describes the allied intervention in the Russian Revolution in 1918-1920; if X=Bulgaria and Y=Serbia, it describes the Serbo-Bulgarian war in 1885. Sometimes these underlying general patterns are discussed explicitly, more commonly they are used implicitly in arguments based on precedent and analogy.

Substitution principles often derive simply from the natural language content of a statement itself—for example in the case given above, X and Y would be any pair of mutually antagonistic states. However, the allowed substitutions might be specific to an individual or organization. For example, when United States decision-makers accepted the Munich analogy as a guide to dealing with Vietnam, the substitutions

Southeast Asia 1965 = Europe 1938
North Vietnam = Germany
Ho Chi Mihn = Hitler
Ngo Dinh Diem = Churchill

was at least implicit in the argument and occasionally it was explicit. North Vietnam’s preferred analogy

Southeast Asia 1945 = North America 1775
French Indochina = British colonies

25 An exception occurs when there is a clear and obvious precedent: for example U.S. policy towards Ferdinand Marcos in the Philippines was discussed in terms of Marcos as "another Somoza", the Nicaraguan dictator whose fall led to the establishment of the Sandinista regime opposed by the United States.

Ho Chi Mihn = George Washington
Ngo Dinh Diem = Benedict Arnold

was not accepted. The U.S. also invoked the Munich analogy (endlessly…) when dealing with Iraq in 1990-91; but this analogy was not invoked when dealing with Israel after 1967 or with Turkey after the invasion of Cyprus in 1974.

Stories have a number of advantages as a knowledge representation structure. First, they solve the combinatorial problems that exist when only short-term rules or heuristics are used: a story reduces the set of possible outcomes to something reasonably finite. This allows prediction through pattern recognition: a story is a pattern where the early parts of the story are temporally coassociated with the later parts.

Weigley provides an interesting case of precedent over-ruling deductive argument:

> In 1966 Walt Rostow called President Johnson's attention to the effects of sustained aerial attack on Germany's petroleum facilities late in World War II and argued "With an understanding that simple analogies are dangerous, I nevertheless feel it is quite possible the military effects of systematic and sustained bombing of [petroleum supplies] in North Vietnam may be more prompt and direct than conventional intelligence analysis would suggest." The intelligence analysis in question indicated that North Vietnam depended so little on petroleum … that bombing … would not much affect the war in the South or compel North Vietnam to make peace. But the Joint Chiefs agreed with Rostow's analogy, and so the aerial campaign against North Vietnam's petroleum was attempted. (Weigley 1973,387)

The bombing campaign eventually failed largely for the reasons suggested in the deductive intelligence analysis, but the analogical argument prevailed[26].

Schank and Abelson relate sequences—their term is "scripts"—to the fundamental process of "understanding"

> In order to understand the actions that are going on in a given situation, a person must have been in that situation before. That is, understanding is knowledge-based. The actions of others make sense only insofar as they are part of a stored pattern of actions that have been previously experienced. … Understanding is a process by which people match what they see and hear to pre-stored grouping of actions that they have already experienced. New information is understood in terms of old information. (Schank and Abelson 1977, 67)

In international politics, understanding includes information that has been learned or deduced rather than experienced, but the principle is the same. "Understanding" a political situation means fitting observed events into a pre-existing event structure. The

---

26 Rostow seems particularly fond of analogical argument: Wirtz's (1989) discussion of analogies in the Vietnam War opens with a discussion of Rostow's approval of a memo comparing North Vietnam in 1967 with the American Confederacy in 1863-64. Majeski and Sylvan's extensive research on Rostow emphasizes his use of rules and heuristics, but many of those rules were complex scripts based on historical analogies.

analyst's "intuitive feel" is simply the subcognitive ability to match a set of observed events to sequences he or she already knows by using substitution principles.

Stories are also a means of inferring "motive": A motive is the end point of a sequence. Associated with the motive is a series of sequences that terminating in a specific end-point.[27] Motive sequences can be used either inductively or deductively. Inductively, one has a set of facts that could match any of several different end points (e.g. is an arms control proposal intended to reduce arms or to weaken oneself prior to an opponent initiating hostilities?); the decision-maker then searches for information to differentiate between those sequences. Deductively, an end point can be assumed and one can seek to differentiate between the various paths that might be used to reach the end point and thwart them (e.g. you know an opponent is trying to weaken your alliances, but how?). In both instances the use of stories dramatically reduces the information-processing task by identifying only those items of information necessary for prediction.

Finally, stories provide means of correcting for noise and missing information. This is particularly important in the political environment that is subject to low information and deliberate deception. Pennington and Hastie observed in an experimental study of individuals summarizing trial evidence:

> …spontaneous interview protocols do exhibit story structures. … Juror's stories were not simple lists of evidence. They always contained most components of the episode schema (initiating events, psychological and physical states, goals, actions and consequences) in appropriate causal relations. Jurors typically inferred missing components for these structures when they were not contained in direct testimony at trial. Evidence not related to a story of what happened was systematically deleted from discussion. (Pennington and Hastie 1986,252; quoted in Boynton 1991).

Stories carry complex information in a compact and easily usable form.

The most widespread application of stories in political analysis is found in the use of history or precedent.[28] In foreign policy discourse, precedent is most likely

---

[27] One of the purposes of myth is to provide idealized sequences that terminate in a known end-point so that the early parts of the sequence can be used to infer motive. "Little Red Riding Hood" is a complex sequence carrying the messages "Obey your mother; don't talk to strangers; and don't wander about in the woods"; *MacBeth* and *Hamlet* carry the message "Those of avarice and deceit will come to no good in the end." Brunvald's study of urban legends finds that they "repeatedly [have] the quintessential urban legend twist—the poetic justice, or got-what-he-(or she)-deserves twist" (Brunvald 1984, xiii). This provides a deterrent function—"you might think you're smart but God'll get you…"—that undoubtedly accounts for the prevalence of urban legends: They provide a message we to wish passed along to others.

Less subtly, in preliterate societies where knowledge is transmitted in part by itinerant storytellers, one typically finds an ample supply of legends where an apparently helpless old wanderer is a hero, god or ruler in disguise. Ulysses' return in the *Odyssey* is perhaps the most familiar in early Western culture; *Luke* 24:13-35 provides a variant in the Christian tradition. Such stories have clear and immediate utility to the story-teller.

[28] The arguments for precedent from a computational modeling perspective are reviewed in Mallery and Hurwitz (1987), Anderson (1981) and Mefford (1991). A variety precedent-based models were developed in a series of papers by Alker and others during the 1970s (Alker et al 1972, 1977, 1980). The models also are closely related theoretically, though not necessary in technique, to an extensive body of AI research on analogical reasoning (see Prieditis 1987) and case-based reasoning (see Kolodner 1988,1993; Mefford

encountered as a *justification* for action—in other words, it is invoked as an empirical regularity. A precedent such as "Munich", "Pearl Harbor" or "Vietnam" would be invoked as something to be avoided, not to be implemented[29]. The precedents used repeatedly are incorporated into the standard operating procedures of the organization and in doing so become rules.

In contrast, in legal reasoning and in much of the case-based reasoning literature a precedent is a *plan*, a series of actions that *should* be taken. In this sense a precedent is merely the complex antecedent clause of an *if...then* clause with a complex consequent. These two uses of precedent are different but not incompatible: over time, the use of precedents as plans causes them to become empirical regularities. The utility of precedent is an empirical issue: pattern recognition is very effective in chess but useless in Lotto. Decision-makers may study the politics of the Napoleonic period for guidance on contemporary international affairs; they do not accord the same respect to Napoleonic surgical techniques or gunnery practices.

Organizations and Rules

> *One man alone can be pretty dumb sometimes, but for real bona fide stupidity there ain't nothing can beat teamwork.*
>
> *Edward Abbey*

International behavior is primarily the product of bureaucracies rather than individuals. An individual may influence the direction of a foreign policy, but the implementation is still left to bureaucracies.[30] In everyday language this organizational interaction is simplified—"Hitler decided to attack Poland"—but in virtually all cases (and certainly in systems which have a strong democratic and/or bureaucratic component) individuals are constrained to choose from a very small set of options that have been made available through a bureaucracy. While the "Great Man" [sic] model attempts to

---

1991). Khong (1992) provides an extensive analysis of the use of analogy in foreign policy decision-making, based primarily on the Vietnam decisions. Neustadt and May (1986) provide an elaborate set of general guidelines and heuristics on the practical use of history and analogy for decision-makers, lavishly illustrated with examples from the post-WWII period.

29 This can occur for normative as well as pragmatic reasons: for example Allison emphasizes the strong negative impact of the Pearl Harbor analogy on Robert Kennedy's assessment of the option of bombing Cuba during the Cuban Missile Crisis:

> "I now know how Tojo felt when he was planning Pearl Harbor" Kennedy wrote during an ExComm meeting, and he would later write "America's traditions and history would not permit … advocating a surprise attack by a very large nation against a very small one". (Allison 1971,197).

30 The focus here is on the Weberian "rational-legal" bureaucracy that dominates foreign policy decision-making structures in the modern era. This applies even in "revolutionary" situations: For example, the taking of American hostages in Iran during 1979-80 was outside the normal range of international behavior, but the attempts at resolving that issue were very much normal, including the unsuccessful rescue attempt and the eventually successful mediation by Algeria and other international agents. Similarly, the rhetoric surrounding the Iran-Iraq war and the personal animosity between Saddam Hussein and Khomeini provided personalistic aspects to that conflict—as would the animosity between Saddam Hussein and George Bush in 1990-91— but the military interactions were bureaucratically quite conventional.

allow the cognitive processing of an individual replace bureaucratic decision-making in an organization, the individual is still dependent on an organization to supply (and filter) information and implement decisions. Behind every Great Man is a well-entrenched bureaucracy pleased to have someone else taking responsibility.

The shift from individual to organizational information processing produces a paradox: An organization, in order to obtain the capacity for large-scale information acquisition and policy implementation, must make substantial compromises in its information processing capabilities. Specifically, an organization must work within the bandwidth constraints of language, and it cannot directly invoke associative recall. This substantially reduces the dimensionality of the state-space used to describe the problems being considered. Sylvan, Majeski and Milliken note in their study of U.S. decision-making concerning Vietnam:

> [W]e found, at least for the troops decisions, that at any given time there were three policy lines: a main one and two subsidiary ones. Other policy decisions were of course put forward, and on rare occasions they were discussed at high levels; but as a general rule, the president and his top advisers simply did not consider more than three principal lines at a time (Sylvan, Majeski and Milliken 1991,333)

Because of these factors—and consistent with Purkitt's (1991) discussion of the limitations of working versus long-term memory—the knowledge used by an organization is usually simpler than that used by the individuals who comprise it.

The sequential processing of *if...then* rules can be done within the constraints of working memory so this, rather than associative pattern recognition, is the preferred mode of information processing in organizations. In much of their behavior, the bureaucracies are not acting *as if* they followed rules; they are instead *explicitly* following rules and are expected to do so, rule-following being a *sine qua non* of bureaucratic behavior.[31] This section will focus on the interaction between the explicit rules of an organization and the pattern recognition used by its human components; a detailed discussion of use of rules to model organizational behavior will be deferred to Chapter 4.

Argument, Understanding and Tacit Information

Some of the early computational modeling projects assumed that due to the rule-oriented nature of bureaucracies, one would be able systematically to extract an organizations rules and precedents from a sufficient quantity of debates, formal regulations and internal memoranda, and from these rules one could simulate much of the decision-making process. Based on the subsequent success of rule-based systems in replacing some routine managerial decision-making in business, this was not an unreasonable proposition. In fact, had the computational modeling projects focused on *routine* State Department activities, such as the processing of visa requests or arranging golf games for visiting members of Congress, the approach might have worked.

Instead, because an extensive set of documentary evidence was required, the typical focus of these projects was on critical decisions such as Vietnam. In crisis decisions

---

31 Thorough treatments of the role of rules and heuristics can also be found in Majeski (1987), Mefford (1991), and Sylvan, Goel and Chandrasekran (1990); the approach also permeates the Kahneman, Slovic and Tversky research.

explicit rules proved insufficient. Sylvan and Majeski's study of Vietnam decision-making quickly encountered the problem of "tacitness" and "tacit cultural rules" where many of the key elements required to understand the bureaucratic debate were absent from the recorded discourse:

> For example, the first few months of 1961 were marked by the eruption of a non-Vietnamese war in Indochina. The country, of course, was Laos, a subject on which Kennedy's advisers spent literally hundreds of hours. Yet the Vietnam documents of that time barely mention Laos. It is only after the crisis fades that explicit references to the Laos settlement are made, particularly in the form of 'lessons'. ... Our heuristics reflect very little of situation-specific interpretations, and this means we miss many of the allusions and other references in any given text. (Sylvan and Majeski 1986,11)

For example, fundamental theme of anti-communism was never verbalized because it was a shared assumption:

> Most people employ a host of common sense rules for getting along in daily life; only a small handful of these will ever be verbalized. In part, this is because of their obviousness; in part, due to the embarrassment that would attend someone who reminded others about them. … If rules are shared in the [bureaucratic] culture as a whole (e.g. communism is bad), they will never (or almost never) be made explicit. (Sylvan and Majeski 1986,10)

Boynton noted a similar problem in trying to formalize the construction of narratives in Congressional hearings on the 1983-84 Lebanon policy:

> The apparent reason a narrative account was not constructed was that members of the committee knew what had happened. They had been arguing about Lebanon policy for well over a year, and they could take the knowledge of the events for granted. … My tentative conclusion is that narrative reconstruction is a fallback mode of reasoning. When the subject is well known [it] is not necessary (Boynton n.d.,8)

Because understanding involves matching observed events to a pattern, the function of political discourse is to provide sufficient information—in the forms of declarative knowledge, event sequences and substitution principles—to cause the audience to understand (i.e. pattern match) the situation in the same manner that the individual transmitting the information understands it. Political information transfer attempts to stimulate pattern recognition in the mind of the audience and thereby trigger a desired behavior. This process can occur between competing organizations as well as within them: Signaling in a conflict situation involves exchanging messages with an opponent in an attempt to get the opponent to take, or refrain from taking, certain actions.[32]

In an organization, rules are invoked by patterns. When a pattern matches the antecedent of a rule, it directly or indirectly causes the implementation of policy responses, for example the deployment of military units, granting of aid or lifting of trade

---

[32] As noted in Chapter 7, a plan of strategic deception does this through the systematic use of false information that simulates a pattern of events that does not actually exist.

sanctions. The patterns conveyed in political discourse are keys attempting to fit a lock,[33] and the response may occur directly or as one part of a chain of rules.

Heuristics

Cyert and March's (1963) classic, *A Behavioral Theory of the Firm*, examines organizations as information processors. Almost three decades old, this work constructs and tests formal models that nowadays would be considered expert systems or rule based models.[34] Cyert and March emphasize that rules can take a variety of forms, including, "rules of thumb", professional expertise (e.g. rules used by a carpenter), rules for the distribution and storage of information (e.g. memos, reports and records), legal constraints and requirements, contractual agreements (e.g. union work rules), "common industry practice", "tradition" and so forth. In the international political arena, "regimes" provide similar sets of expectations.

Because international politics is a complex problem-solving environment, heuristics—simple rules used to partially solve complex problems—are of particular importance. Purkitt observes:

> To cope with limited cognitive capabilities, individuals selectively process information and use a limited number of heuristics or mental rules of thumb as cognitive aids in their effort to manage information. This apparently universal reliance on intuitive heuristics to solve all types of problems seems to be due to the need to compensate for the limitations of short-term memory and information processing capabilities. By using intuitive mental heuristics, people can develop a problem attack plan which permits them to develop a subjectively acceptable problem solution. (Purkitt 1991,43)

Heuristics are frequently very specific; their power comes from their quantity. Most estimates put the number of heuristics used by expert human problem-solvers in the tens of thousands.

Majeski provides an example of four such heuristics on U.S. decisions to use military force

A. Increasing military involvement is possible if such action does not lead to a losing military situation.

B. If acts are interpreted as a case of aggression, then steps must be taken to halt such behavior.

---

[33] With respect to finding political analogies for the three biological information processing systems—genetic, neural and immune—this process is the immune. The immune system responds to harmful foreign proteins in the body by first be recognizing them as harmful. This is done by anti-bodies, which are highly differentiated to match specific proteins characteristic of various potential threats. Once an antibody has identified an intruder, the immune systems goes into action and destroys it. In a similar fashion, rules are solutions waiting around for problems to fit their antecedent clauses; when triggered they provide a specific response to the problem.

[34] Compare, for example, the models in chapters 7 and 8 of Cyert and March with Kaw (1989) or Tanaka (1984). Cyert and March were dealing with economic organizations but the approach has been very influential in studying foreign policy, particularly through Allison (1971). Sniderman, Brody and Tetlock (1991) provide an extended discussion of the role of heuristics in mass decision-making.

C. Start at the most minimal [force] level necessary, and progress step by step openly when a step has not succeeded in achieving the objective.

D. The initiation or escalation of military action requires justification. (Majeski 1987,74-80)

Rational choice and balance of power theories are both heuristics in the sense that they are relatively simple; they come with a complex set of side-conditions; and they are intended as general rules to guide decision-making, without providing a complete specification of actions to be taken. To the extent that the decision-makers share an heuristic in a political system—for example balance of power in 19th century European diplomacy or the Chicken game in 20th century nuclear deterrence—it reduces uncertainty and becomes self-validating.

When conflicts do occur among logically inconsistent heuristics or the heuristics fail to accomplish the desired results, a problem is "kicked upstairs" and resolved laboriously through serial processing in an individual, or in an organization becomes a "crisis" requiring managerial intervention. The decisions we are most aware of involve serial processing, which is conscious, and crisis decision making, which is both conscious and dramatic. However, most information processing most of the time is actually done by these lower-level methods that go unnoticed so long as they are smoothly functioning.

Learning and Adaptation

> *The foundations of economic analysis since the 1870s have been the rationality of individual behavior and the coordination of individual decisions through prices and markets. There has been a steady erosion of these viewpoints [and now] the rationality of individual behavior is also coming under attack. ... What I foresee is a gradual systematization of dynamic adjustment patterns both at the level of individual behavior and at the level of interactions among economic agents. ... For a century, some economists have maintained that biological evolution is a more appropriate paradigm for economics than equilibrium models analogous to mechanics.*
>
> *Kenneth J. Arrow*
>
> Science*, 17 March 1995, pg. 1617*

Learning and adaptation are such basic human activities that they have been taken for granted in much of the informal bureaucratic politics literature. These are somewhat distinct activities. Learning primarily involves the acquisition of static information—declarative knowledge and stories—that have the potential for being used in problem solving. Adaptation involves the dynamic development of rules and substitution principles that determine when to use the information. Conceptually, learning occurs with the knowledge base, adaptation occurs with the pattern recognition mechanism.[35] In practice, it is difficult to disentangle the two because of the

[35]International relations, and political science generally, usually instructs using historical examples followed by principles. This contrasts to the practice in axiomatic knowledge domains (mathematics, physics, statistics microeconomics) of teaching general principles followed by idealized examples.

subcognitive nature of pattern recognition in human decision-making, but in a computational model the two processes are usually quite distinct.

Decision makers use the "lessons of the past" to formulate policy, continually modifying those policies depending on their success or failure.[36] For example, the CIA overthrow of Mossadegh in Iran in 1953 was used as a precedent for the overthrow of Arbenz in Guatemala in 1954; those successful interventions were then used as the model for the ill-fated Bay of Pigs invasion against Castro in 1961. The failure at the Bay of Pigs was attributed in part to the low level of U.S. military support; this was corrected by the massive use of US troops in invading the Dominican Republic in 1965 and Grenada in 1983.

When the CIA planned the Bay of Pigs invasion, it was considered safe because of its similarity to the earlier successes in Iran and Guatemala. This pattern match was, in retrospect, incorrect. Had the Marine Corps been involved in the planning, analogies might have been made instead to the problems of amphibious landings encountered in WWII and this analogy would have led to greater skepticism about the operation, or at least better preparation. In this framework, the *knowledge base* used by the CIA was deficient because it contained "overthrow" precedents to the exclusion of "invasion" precedents.

Alternatively, one could argue that the Bay of Pigs situation had been incorrectly understood because important variables were not considered: the *pattern matching process* was in error. If that process indicates that two things should match but they produce different outcomes, the organization should change the matching criteria. Arguably this occurred historically: the failure at the Bay of Pigs was attributed to the lack of sufficient force and the lack of a clear United States commitment, and therefore the next two US interventions in the Third World—the Dominican Republic and Vietnam—involved the overt use of large numbers of troops, quite possibly with WWII, Korea and NATO in mind as precedents.

Cyert and March note that history of any organization involves adaptation to improve performance:.

> In order to examine the major attributes of short-run adaptation … we need to take a new look at the … ways in which [standard operating] procedures change. Standard operating procedures are the memory of an organization. Like any other memory or learned behavior, [they] change over time at varying rates. Some rules seem to change frequently; others appear to have been substantially unchanged for as long as the organization has existed. … When we ignore adaptation with respect to decision rules, we are simply saying that relative to the time span being considered those rules are given. (Cyert and March 1963, 101)

An individual learns patterns by observing sets of behaviors and classifying them as "I've seen this before" or "I've not seen this before". Sets of behaviors previously encountered reinforce existing patterns—they provides evidence that the pattern is

---

Political science uses the pedagogical principles of second language acquisition rather than mathematics. The popular case study approach in addition combines the acquisition of static information with the refinement of pattern recognition.

36 See for example the discussions in May 1973; Neustadt and May 1986; Vertzberger 1990; and Khong 1992.

common and therefore worthy of attention. Frequent patterns can be generalized and stored as templates to avoid retaining redundant information.

Novel sets of behaviors, in contrast, are provisionally added to memory. If they are not reinforced, they are eventually discarded as erroneous or statistically unlikely; they are retained if reinforced.[37] Finally, if observed behavior appears to match an existing pattern but further experience shows that the pattern predicted a different outcome than actually occurred, then either the matching criteria are changed or one seeks additional information that would have differentiated the two patterns. Schematically, basic approach to adaptation is shown in Figure 3.1.

Most patterns of international behavior are acquired through the study of history or by observing the contemporary experience of others, not through personal experience. International politics changes very slowly, providing limited new information, and is a risky environment. All organizations would prefer that patterns involving highly undesirable outcomes—for example losing a war—be learned indirectly rather than from experience. The prerequisites of professional diplomacy—typically graduate-level training in some combination of history, political science and economics—insure that general patterns of historical and contemporary political behavior have been learned before an individual even joins a foreign policy bureaucracy; the organization then supplements these with specialized courses of study, “war stories”, formal organizational history, shared counterfactuals and so forth.

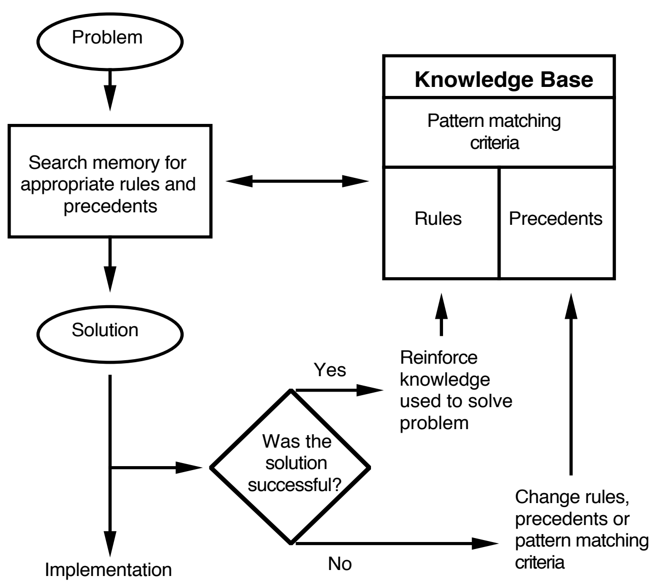


Figure 3.1. Organizational Adaptation

---

37 For example the US intervention in Lebanon in 1958 does not appear to have set an important precedent for the United States, though comparable interventions in Africa were quite common for Britain and France in the 1960s and 1970s.

Sources of Knowledge

Experience

Direct experience is the source of most of our expectations about human behavior: lessons from this source are intensely and sometimes painfully acquired from infancy onward. Experience plays a major role in understanding such politically relevant individual psychological states as fear, pride, and greed. Considerable direct experience is also acquired in the mundane day-to-day operation of a foreign policy bureaucracy and the normal relations between states.

In international politics, the role of experience is limited because the major events that capture the attention of theorists and are the focus of long-term planning are rare. Few states experience a major war more often than once every two decades; crises that might involve war are also fairly infrequent; and major international agreements such as SALT and GATT occur only once every decade or so.

History

In the absence of direct experience, foreign policy experts spend a substantial amount of time studying history, and classical writers on international relations continually admonish them to do so.[38] This history is typically partitioned into autonomous event sequences and stories—for example the 1914 crisis and the Cuban missile crisis—that provide archetypal patterns of behavior.

History enters the foreign policy community in two ways. The first is through formal book learning—foreign policy experts rarely have less formal education than an M.A. in the humanities or social sciences, and Ph.D.s are not uncommon. The second source is organizational memory consisting of the event sequences considered important within the organization. These become acceptable analogies that will motivate action in the organization. History is reinterpreted over time, and the lessons differ with individuals. For Ronald Reagan, the United States involvement in Vietnam was a glorious, selfless, and humanitarian endeavor; this interpretation was not universally shared in either the foreign policy community or the military.

Trauma.

History is sampled selectively. While a sequence of events could probably be found somewhere in history to justify almost any course of action, such scattered applications of history are rarely observed. Instead, certain historical sequences are used much more commonly than others.

In the public justification of policies, the most frequently invoked historical events tend to be traumatic. A traumatic sequence that vaguely fits a situation will supersede a less traumatic sequence that fits the situation more closely, particularly in public debate. This could be due either to the fact that the general public only remembers traumatic historical events, or to the fact that traumatic events invoke sequences leading to highly undesirable outcomes. For example, the two most important “lessons” of history for the WWII generation of US foreign policy analysts were “Pearl Harbor” and “Munich”; the

---

[38] Gaddis (1992/93) presents an excellent debate of the merits of the historical and behavioral approaches; Gaddis (1987) is a more typical traditionalist critique.

"lessons" for Soviet analysts of the same period appear to have been "Hitler" (a variation on the earlier lesson "Napoleon") and "Stalingrad" (Western powers can not be trusted to aid to the Soviet Union even when they share common interests).

Myth and Counterfactuals

History is not the sole source of patterns of behavior: counterfactuals are also important and perhaps more so in rhetorical situations. Myths, children's stories and fairy tales are the archetypes of counterfactual conveyors of social lessons. For example, the most widespread counterfactual in United States popular history is probably the quick-draw shoot-out at high noon between the good guy and the bad guy. Western writers from Mark Twain to Louis L'Amour consistently failed to find *any* original instances of this behavior in American history:[39] it is a total fabrication of nineteenth century popular fiction. Yet this image is popularly invoked in editorials and editorial cartoons to justify the US policy position in militarized crises such as the invasion of Grenada in 1983, the bombing of Libya in 1986, the invasion of Panama in 1989 and Operation Desert Storm in 1991.

Myths vary in their message. The lesson of both the *Illiad* and the *Odyssey*—stories known to every educated person in Europe and North America—is rugged individualism. Much of the *Illiad* focuses on Achilles, a petulant and egocentric, but tolerated, jerk. In the *Odyssey*, Ulysses returns home to triumph, although all of his companions perish. In contrast, the lesson of the great South Asian epic, *The Ramayana*, is one of the loyalty of a leader to his wife, his family and his followers. The underlying messages of these epics different substantially, and such differences could have profound implications for the types of behavior that might be anticipated in a political interaction.

Probably the most important, and most easily justified, application of counterfactuals has been in the analysis of nuclear stability. The international system has had no experience with the outbreak of war between two nuclear powers, and the unilateral use of nuclear weapons at the end of WWII provided little useful guidance on the issue. Instead, the entire policy analysis on such major issues as the deployment of anti-missile defenses and multiple warhead missiles was done with theory, analogies and counterfactuals.[40]

One interesting offshoot of these counterfactuals in foreign policy analysis was the development of an extensive popular literature on the outbreak of nuclear war. This was initially apocalyptic, such as *On the Beach* (Shute 1957), *Failsafe* (Burdick and Wheeler 1962) and the film *Dr. Strangelove*. Later a popular literature evolved emphasizing the technical sophistication of the non-nuclear weapons and the organizational control of the NATO and WTO militaries, and promised large-scale European war without nuclear catastrophe: Sir John Hackett's *The Third World War* (1978) and Tom Clancy's *Red Storm Rising* (1986) are the two best-selling examples of this genres. Life imitated art when Vice President Quayle used *Red Storm Rising* to justify the military effectiveness

---

[39] That is, instances of shoot-outs that were not themselves inspired by the pulp fiction (notably Ned Buntline) of the day; Twain mentions several such instances in *Roughing It*.

[40] See Fearon (1991) for a general discussion; Gaddis (1990) provides a succinct summary of these arguments based on a conference entitled "Nuclear History and the Use of Counterfactuals" with particular emphasis on Elster's (1978) analysis.

of "stealth technology". While Quayle's use of fiction was ridiculed, from a strictly logical perspective it was little different from the formal arguments underlying the nuclear weapons programs of the superpowers.

To be an effective player in a bureaucracy, one must internalize and effectively use the organization's counterfactuals. Counterfactuals justify and reify existing organizational norms and procedures, they direct debate and hence smooth the operation of the bureaucracy, and they provide a means of minimizing the effectiveness of "outsiders" who have not mastered the internal culture.

Culture

Closely related to the issue of myth is that of culture. From the perspective of knowledge representation, culture is a series of expected behavioral sequences, both social and political. As noted above, some of these expectations are explicitly learned through myth or history, but a myriad of additional patterns are acquired from childhood onward.

To take a simple nonpolitical example, in some cultures, a polite business conversation almost always begins with an inquiry about the other's family; not asking would be considered rude. In other cultures, the topic usually would not be mentioned in a business conversations. In still other cultures, typically those with high infant mortality rates, asking about family is considered extremely impolite. This is only one bit of cultural knowledge; a culture contains thousands.

"Culture shock" occurs when one is surrounded by unfamiliar social patterns.[41] While this is usually associated with a shift in language, it can just as easily occur simply through a shift in social environment: most academics, for example, would probably feel exceedingly uncomfortable at a meeting of unemployed autoworkers, unless they happened to be studying autoworkers and knew what behaviors were expected. We apparently monitor the social environment subcognitively, and when observed behaviors deviate sufficiently from what we expect, a signal is triggered that says "let's get out of here". Over time, the new patterns can be learned and the signal is suppressed.

As discussed above, organizational culture is particularly important. Sylvan and Majeski make this point with respect to bureaucratic decision making:

> The rules in which we are interested are most usefully characterized as cultural. Specifically, the rules represent the everyday, common sense understanding of the member of a particular subculture: high bureaucratic officials in mid-twentieth century America. Those officials know, as a matter of routine, all matter of things about the way their world works: these vary from methods to ingratiate oneself with one's boss to an understanding that the Russians are obviously an enemy who has to be watched very carefully. It is this taken-for-granted quality of cultural rules that many observers in international relations find so irksome.
>
> Insofar as the rules we are studying represent the common sense knowledge of policy makers, the rules will tend to be tacit. … In general, we can say that if rules are justificatory (e.g. making a legal case) they will be often be made explicit. If

---

[41] For the strict positivist who argues that culture shock is a phenomenon we should not admit into scientific discourse, I can provide a long and quite interesting list of places where it can be intersubjectively experienced…

> rules are very general (e.g., what are our aims in Southeast Asia), they will be made explicit during crises. If rules are widely shared in the culture as a whole (e.g. communism is bad), they will never (or almost never) be made explicit. (Sylvan and Majeski 1985, 16-18)

Illusion

The human brain is very good at using patterns to filling in gaps in observations. This is very useful—in fact essential—when dealing with noisy stimuli, but it also means that one is likely to learn things that are not true. Vertzberger notes

> Issues that are uppermost in the information processor's mind have a priming effect and exert disproportionate influence on the interpretation of incoming information, biasing it in terms of these particular concerns even if the bias is not justified.
>
> The complexity of integrating multiple indicators into a unitary assessment of the situation (e.g. the adversary's intentions) encourages a motivational bias towards deriving a single indicator from one's conceptual framework. [With this simplification] not only does the whole process of searching for and evaluating information become simplified, but the probability that multiple indicators will produce inconsistency … is preempted and avoided as well (Vertzberger 1990,66-67)

Illusion can also arise from the hierarchical structure of memory. For example, if one mentally visualizes a zebra, one seems to have has a very clear image of the animal. The mental image is sufficient to distinguish a zebra from a horse, or to infer the missing parts of a zebra partially blocked by a tree, or to identify a zebra distorted by fog, shadow or rain. The pattern appears to be very complete and robust. However, its incomplete nature is apparent when more detailed questions are asked: for example, how many stripes cross the zebra's back; where on the zebra are the stripes horizontal?

This same type of illusion occurs with political knowledge. For example, students are usually astonished to find that during the 1980s China, considered a military "superpower", had a military budget less than that of Japan, usually considered to be militarily insignificant.[42] This could be true even for experienced analysts; Laurance reports an experiment where:

> in-house military analysts were used to help create a set of country scores for the military capability (naval in this case) of various LDCs. When we validated these scores with a set of outside experts, most were surprised that the PRC had scored so high. Upon reflection we realized that the high scores were a function of the expectation that the PRC would be very capable, since at the time they were in the 'communist enemy' category. In addition the PRC was high on the collection and analysis priorities list of the intelligence community. More information existed on the PRC navy, compared to say Malaysia, Singapore and India (Laurance 1990,125)

---

[42] For example, ACDA reports China's 1988 military budget as $20-billion; Japan's as $24-billion (ACDA 1989). While this calculation may involve some dubious use of exchange rates, Japan was still outspending China by around a factor of ten in terms of military expenditure per capita.

This illusory aspect of pattern recognition may account for some of the difficulty in locating precise historical precedents in political arguments. When a human processes a political argument, a single historical precedent—for example Munich—will associatively activate a complex story whose details are filled in by memory. From a human standpoint, the precedent was an important part of the argument. Objectively, however, the historical reference is only a small part of the *text* of the argument and a computer, without associative recall, finds little use of precedent.

Adaptation

A major focus in Cyert and March is organizational adaptation:

> Without denying the substantial abilities of organizations as problem-solving and decision-making institutions, … a business firm is constrained by the uncertainty of its environment, the problems of maintaining a viable coalition, and the limitations on its capacity as a system for assembling, storing and utilizing information. As a result, the theory outlined in this volume characterizes the firm as an *adaptively rational* system rather than an *omnisciently rational* system. (Cyert and March 1963,99; italics in the original).

The basic adaptive mechanism proposed by Cyert and March is corrective feedback under satisficing norms, a process similar to that in Figure 3.2. If the organization's output is approximately what it desires, only marginal changes to rules occur. Rules change when policies fail due to a new environment, stochastic "bad luck"[43] or when the rules never worked in the first place. The basic process of adaptation is "bureaucracies do not make the same big mistake twice".

The second general characteristic of adaptation is the *reapplication* of solutions that worked in the past: in the rule-based JESSE simulation (Sylvan, Goel and Chandrasekaran 1991) this is called "compiled reasoning". Regularized behaviors evolve within an organization through an adaptive process of mutation and selection, and hence there will be common roots to apparently disparate behaviors. Even the most cursory examination of international law shows an evolutionary process at work: compare, for example, international maritime and international aviation law, or international postal law and international telecommunications law. Successful new laws, often as not, are modifications of existing laws, rather than created *de novo*, and given the choice between two equally effective solutions to a problem, organizations will gravitate to the more familiar.

Because of this continual adaptation, the interaction effects of the rules of a complex system are effectively unknown. Adaptation creates a system that is not necessarily logically consistent, nor is its behavior easily predictable in all circumstances. In a well-functioning bureaucracy in a stable environment, the same input should produce the same output each time, and the more frequently an organization is confronted with a particular situation, the more efficient its response. However, when an organization is confronted with a novel input, its response may be both inefficient and unpredictable, which has obvious implications for crisis behavior.

---

[43] For example, one of Cyert and March's case studies involved rule changes initiated after a fatal accident involving antiquated machine controls.

Adaptation provides "path dependence": International politics is not like a game of chess where future plays can be evaluated solely on the basis of the current board position. In a competitive situation, the fact that a clever set of actions has provided outstanding gains over an opponent is a good reason to assume that the same set of actions will *not* produce that result the second time around.[44] Van Creveld notes that in war,

> ...the underlying logic is not linear but paradoxical. The same action will not always lead to the same result. The opposite, indeed, is closer to the truth. Given an opponent who is capable of learning, a very real danger exists that an action will not succeed twice because it has succeeded once. (van Creveld 1991,316)

How a situation came into being may be as important as its static characteristics; deductive reasoning from first principles alone is not sufficient.

A potential problem in adaptation is the phenomenon of *superstitious learning*. As defined by Lave and March, this occurs in

> …situations in which the environment is indifferent to the individual's behavior; it dispenses rewards or punishments on some basis other than a response to his actions. The individual, however, will not necessarily be able to detect this indifference…the individual seems to behave as if he believes the environment is responsive, and he learns in the usual way. [This] encompasses some of the most widespread, and most interesting, forms of human behavior (Lave and March 1975,294)

Superstitious learning is so called because it includes behaviors normally considered "superstitious": a person who avoids swamps out of fear of evil spirits will still obtain the benefits of avoiding contact with malaria-carrying mosquitoes. But the phenomenon goes beyond conventional superstition to encompass any situation where there is a coassociation between a behavior and a positive result without a causal linkage. As will be noted below, this is particularly problematic in international politics where feedback is exceedingly slow and there are multiple potential paths of causality. An organization may have adapted to an environment that it believes to be firmly under its control, only to find in a crisis that this control was entirely illusionary.

The supposed dominance of the "China lobby" in the 1950s and 1960s and the "Israel lobby" in the 1970s and 1980s are two cases in point. Both were considered at the time to have established complete control over their respective foreign policy domains (opposition to recognizing the Communist regime in China; support for Israeli government policies), yet when directly opposed by a shift in Presidential policy (Nixon

---

[44] At appropriately expert levels, this is true even in chess: Rapoport begins *Fights, Games and Debates* (1974:2) with a story of how a chessmaster, on discovering a novel variation of a standard opening, waited ten years before springing it in tournament play, knowing it could be used but once.

The effectiveness of precedent may vary substantially depending on whether one is in a *cooperative* or *zero-sum* situation. In a zero-sum situation, the fact that something happened once reduces the likelihood that it will happen again. In a cooperative situation—for example coordination games—the opposite occurs: a successful solution provides an anchoring point for other successful solutions. Failure to distinguish these two types of cases could cause precedent to look either more or less effective than it actually is.

in 1970; Bush on loan guarantees in 1991), they proved completely powerless. This result should have been expected given the low salience of either issue to the vast majority of the electorate, but both groups had persuaded themselves—as well as their detractors—that policy was under the control of the lobby.

A final problem to learning by adaptive feedback in international politics is what March and Olsen (1983) call "the inefficiency of history": international behavior occurs with glacial slowness. Experimentation, the conventional means of increasing feedback, is all but impossible: A state can experimentally modify its own policies—at some risk—but not that of its neighbors and not with any controls. In contrast the Federal government can try income support in Newark but not in Baltimore and compare the results. The "experimental" war, alliance or embargo has undoubtedly occurred, but it not the norm.

Consequently, feedback on a policy may require years or decades, and by the time the feedback occurs, the system may have changed sufficiently to invalidate the appropriateness of the policy to the current environment. Galbraith somewhat sarcastically observes:

> As practiced, [foreign policy] is for many participants the least demanding intellectually of public occupations. That is because, the outcome of almost any action or series of action being largely unknown, the individual who knows little is not greatly handicapped as compared with the individual who knows much. (Galbraith 1984,40)

In the international system, historical anachronisms such as Berlin and Hong Kong in the 1980s or the Austrian-Hungarian and Ottoman Empires in the 1910s are hardly atypical. The international institution of sovereignty reduces feedback to a minimum: one may engage in foolish behavior for years without penalty so long as one does not impinge on the affairs of ones neighbors. When changes finally do occur, they are frequently the result of a catastrophic transformation of the system such as WWI or WWII rather than a smooth movement towards an equilibrium.

International Behavior as a Self-Organizing Process

Pattern recognition, rules and adaptive behavior induce regularity in international behavior. In essence, those forms of decision-making create a substantial degree of self-organization so that rule following and pattern recognition are effective decision strategies in a world of rule-following pattern-recognizers. Meanwhile, the ill-defined character of international politics and the high dimensionality of a pattern-based system reduce the effectiveness of deductive optimization.

This argument will be developed in three parts. First, individual and organizational memories change slowly. Second, information constraints and the necessity of maintaining collective action coalitions force organizations into regular behaviors. Finally, adaptive ruled-based systems are usually at local maxima with respect to their selection of rules, so stable patterns of behavior will be observed for long periods of time.

If, to paraphrase Pope, the proper study of bureaucracies is other bureaucracies, they will find, to paraphrase Pogo, that we have met the enemy and they are us. While bureaucracies are severely constrained in their ability to analyze, they are blessed by the fact that their opponents vary little in this regard. This, in turn, often simplifies behavior to a mutually agreeable level.[45]

Memory

The first source of regularity in a system based on pattern recognition is memory. Individuals and organizations spend a great deal of effort selectively preserving the records of the past. In individuals, most of the key patterns of political behavior are learned relatively early in life—usually before the age of 30—and afterwards change little except in response to traumatic events. New experiences build on that base but are strongly affected by it because individuals are more likely to pay attention to information consistent with what they already know than they are to novel information.[46]

The persistence of individual memory has additional implications for organizational communication: Individuals with similar memories will share greater amounts of tacit information and therefore will be able to communicate more effectively with each other than with individuals who do not have those memories. An outstanding example of this was found in the "generational politics" that doomed most Communist regimes by the late 1980s. The ruling bureaucracies in countries such as the Soviet Union, China, Albania and East Germany proved completely incapable of transferring control to people who had not shared their politically formative experience of WWII and the early Cold War; the hyper-centralized Leninist bureaucracies could not function without this irreplaceable base of tacit information.[47] Ironically, generational memories also facilitate communication between opposing bureaucracies; the stability of the Cold War deterrence system and the successful cooperation within the Bretton Woods

---

[45]Another argument for regularity could be made using a variant on the controversial "anthropic principle" of cosmology. The gist of this principle is that the answer to "why do the cosmological physical constants have values they have?" (e.g. why does water ice float on liquid water, whereas most solids sink in their liquids, etc.) is "because they've got to be that way for humans to exist to observe them". Formally,

> The observed values of cosmological quantities are not equally probable but they take on values restricted by the requirement that there exist sites where carbon-based life can evolve and by the requirement that the universe be old enough for it to have already done so. (Barrow and Tipler 1986,16).

The approach is controversial in cosmology because it is self-referential; *any* discussion of human behavior by humans is self-referential, this criticism is less relevant in political science.

Thus one answer to "why expect regularity in human behavior?" is "because humans are sufficiently organized ask the question." In other words, the social behaviors allowing us to create books and universities and go to conferences require a great deal of regularity and pattern. Human life will on occasion be nasty, brutish and short, but it is highly regularized compared to the life of a minnow or an amoeba. If we did not have a reasonable expectation that certain behaviors would be rewarded or reciprocated with some consistency, the social organization required to ask such questions would not be possible.

[46] See Vertzberger (1990,137ff) or Lebow (1981,101ff) for discussions and references to the relevant literature.

[47] Mao recognized this problem explicitly but his solution, the Cultural Revolution, seems to have had the opposite of its intended effect, producing a traumatized generation intent on rejecting centralized systems and eager to embrace capitalism.

international economic system were undoubtedly enhanced by the shared traumas of WWII and the Great Depression.

The written rules of bureaucracies, which are preserved physically rather than subcognitively, are potentially immortal and can transcend by decades the circumstances that created them.[48] By selective recruitment and retention of new members, a bureaucracy can attempt to extend its culture beyond the limits of human mortality. In the case of religious organizations, this can be done successfully for centuries, though most organizations are far less steady. Because many rules remain unchanged except when they produce undesirable consequences, a set of rules dealing with an unusual set of circumstances—for example those invoked in a crisis—may persist for a very long time without challenge.

Information

The bandwidth limitations on organizational communication restrict the complexity of interactive political processes over time. Suppose that an actor transmits a complex signal in an attempt to trigger a novel pattern of behavior in another actor. Only part of the signal invokes a response from the target, so the return signal will be simpler than the stimulus, and the initiating actor's bureaucracy will further simplify its response to the response. Schematically, the process looks like Figure 3.2, where the segments of the boxes indicate various parts of the organization; shading indicates whether they are active in the process, and the lines indicate communication. Two organizations, interacting over time, will reach a level of interaction compatible with their information processing capabilities. Because those capabilities are relatively limited, the behaviors and hence the observed patterns will also be limited. If part of the organization consistently sends signals that fail to elicit a response, eventually (except in unusual circumstances) that agency will turn its attention to other issues: It takes two to tango, particularly in international affairs.

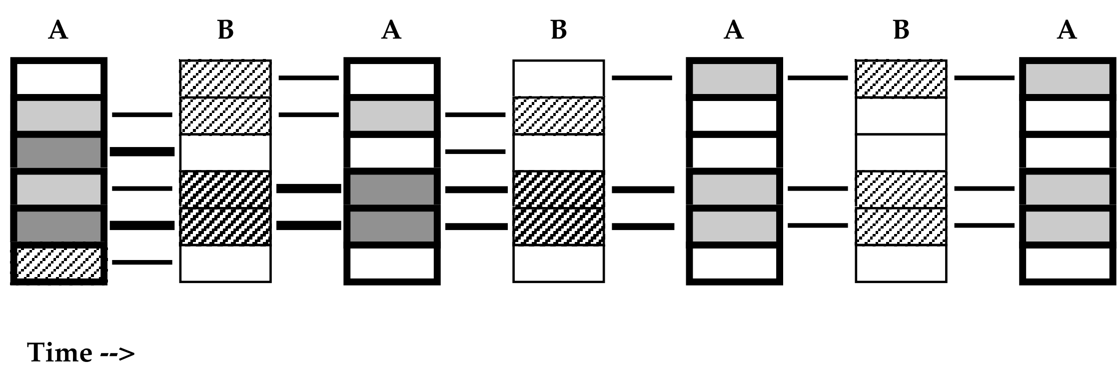


Figure 3.2. Interactive Adaptation

48 Morgenthau (1973,130) quotes a story related by Bismarck in 1859 of a Russian sentry in St. Petersburg posted in the middle of a grassy plot to protect the location of a flower admired by Catherine the Great a century earlier. Morgenthau, meanwhile, is repeating the story (which sounds suspiciously like an urban legend...) a century after Bismarck. University regulations and governance, of course, provide no end of examples of anachronistic rules and rituals.

An interesting example of this simplification can be found in the large number of crises that did *not* occur during the early Cold War. The list of areas involved in persistent military or political conflict is relatively short: Greece, West Berlin, Quemoy and Matsu, and Korea. A much larger number of crises did not occur, including a number that appear more important in geo-strategic terms than the actual crises:

Iran	Soviets back down after invasion
Yugoslavia	Soviets do not contest anti-Soviet Yugoslavian government receiving US aid
East Berlin, Hungary	US does not contest Soviet use of military force against anti-Soviet demonstrators
Kuril Islands	US and Japan do not contest Soviet occupation
Finland, Austria	Neutrality accepted by US, USSR
Tibet	Chinese occupation not contested

This list of non-crises could be continued, but it illustrates the basic point is that in order for a conflict to persist, both parties must continue it. Perhaps the most striking example in this list is the contrast between the treatment of Quemoy and Matsu and the treatment of the Kuril Islands. Both are strategically insignificant; one was allowed to become a major crisis; the other never became an issue for the United States, though it immediately became a major issue for Japan once the Cold War had ended.

Cyert and March note, with considerable empirical support, the importance of reduction of uncertainty on organizational function:

> Organizations avoid uncertainty: (1) They avoid the requirement that they correctly anticipate events in the distant future by using decision rules emphasizing short-run reaction to short-run feedback rather than anticipations of long-run uncertain events. ... (2) They avoid the requirement that they anticipate future reactions of other parts of their environment by arranging a negotiated environment. They impose plans, standard operating procedures, industry tradition, and uncertainty-absorbing contracts on that environment. In short, they achieve a reasonably manageable decision situation by avoiding planning where plans depend on predictions of uncertain future events and by emphasizing planning where plans can be made self-confirming through some control device. (Cyert and March 1963, 119)

Competing organizations, over time, will adaptively organize a largely predictable environment. Some change occurs but the level of instability is low relative to what it might be.

Coalition Stability

An additional information constraint occurs due to importance of collective action in politics. Coalitions providing collective goods are inherently unstable, since the individually rational decision is always to defect. They consequently are difficult to maintain because of the information cost of determining who is cooperating in the coalition and the equally heavy cost of sanctioning defectors. Coalitions, once established, are not maintained by a tenuous C/C solution to the N-person iterated prisoners' dilemma, but instead by a complex, actively enforced, and culturally reinforced set of laws, norms and institutions. In the case of legal systems and bureaucracies, these are explicitly rule following. Cultural norms are much less explicit but still encourage regularity.

Since many of the collective goods provided by political systems—whether physical (highway systems), service (education) or philosophical (freedom of religion)—are meaningful only over a long time horizon, political coalitions must remain stable for long periods in order to be effective. At any point in time, most of the observed political coalitions will be those that have successfully evolved mechanisms for conserving themselves; coalitions that do not solve this problem will be short-lived.

With the increased democratic accountability of foreign policy found in the late 20th century, these constraints extend from the domestic to the international arena. The combinatorial alliance switching of the 18th century balance of power systems does not operate in the democratic systems of today. This is reflected in the problems that Egypt experienced in changing a single international link with Israel from hostility to neutrality, despite Egypt not having an effective elected legislature, having been handsomely rewarded for its new policy by a superpower, and having a central position within the Arab world. In a similar fashion, United States continues to go through periodic problems in its relations with China that can be traced to the over-stated anti-PRC policies of the late 1940s.

Optimization

The final argument that might be made against regularity is the issue of "cleverness". The international system consists of individuals who have foresight and can plan, and they are not restrained, much less condemned, to repeat the past. Such planners will constantly seek to exploit weak points in any system of interactions. Can these clever planners, using deductive logic, dominate the lazy subcognitive pattern recognizers and the plodding rule-followers? In the worst-case scenarios—the Pearl Harbors of history—the planners win, at least in the short-term. But this is not the norm for two reasons.

Dimensionality

The high dimensionality of the state-space of political behavior complicates deductive optimization. Deductive reasoning uses the slow, serial processes of working memory supplemented by memory aids such as written records. It is literal and explicit rather than associative and error correcting. In a situation with low dimensionality, quality information, and well-understood dynamics, deductive reasoning is very effective: ask the computers that run your car. Remove any of those characteristics and human pattern recognition, with its rapid, associative access to tens of thousands of patterns, quickly becomes a contender. An international problem is likely to have high dimensionality, a great deal of noise, poorly understood dynamics and will usually involve a situation for which historical precedents exist. Because political problems are in addition often ill-defined, deductive reasoning can be at a substantial disadvantage.

The crossover point from deductive to recall-based reasoning may occur when the dimension of the state vector of a complex process exceeds the infamous 7±2 item limit of working memory (Miller 1956). While a cleverly deduced behavior can outfox a pattern recognizer once, this cannot be easily repeated: "Fool me once, shame on you; fool me twice, shame on me." Because pattern recognizers are thoroughly primed through myth, legends, organizational culture, social play and other pattern-generating mechanisms, it is difficult to fool them even once.

Local equilibria

The continual adaptation of organizational rules might appear to create a situation where the behaviors of the system are in constant flux. In fact, under fairly general conditions, the opposite will usually be observed because the sets of rules used by various actors adapt to an equilibrium.

This adaptation deals with rules rather than behaviors.[49] The choice of rules followed in foreign policy is *co-adaptive*; it is dependent on the choices of other actors in the system. Formally, if one is looking at two organizations, one organization can improve its performance by adjusting the values of the N rules it controls directly, but its payoff is also dependent on the C parameters of the competing organization:

$$\begin{array}{lccccccccccc} \text{Organization 1} & n_{11} & n_{12} & n_{13} & n_{14} & n_{15} & n_{16} & n_{17} & c_{11} & c_{12} & c_{13} & c_{14} \\ & \downarrow & \downarrow & \downarrow & & & & & \uparrow & \uparrow & \uparrow & \uparrow \\ \text{Organization 2} & c_{21} & c_{22} & c_{23} & n_{21} & n_{22} & n_{23} & n_{24} & n_{25} & n_{26} & n_{27} & n_{28} \end{array}$$

Co-adaptation involves the first organization incrementally modifying n_{1j} in a hill-climbing fashion but its payoff depends on some of the n_{2j} of the other organization. This situation is similar to a two-person game where the row player controls the choice of row strategies but the payoff is dependent on the choices of both the row and column players.

In order to maintain or improve its level of performance, a co-adapting organization may be forced to respond to the changes in c_{1j} induced by its opponent's changes in n_{2j}; this type of interdependence is routine in international affairs. This process may nonetheless have an equilibrium, provided there are points where both organizations are at local maxima. In the terminology of Maynard-Smith (Maynard Smith and Price 1973, Maynard Smith 1982), this is called a Nash equilibrium on a co-adaptive landscape, because of its similarities to a Nash equilibrium in game theory.[50]

Kauffman and Levin (1987) show that under a broad set of conditions, there are a large number of local maxima in these situations and the likelihood of finding a co-adaptive Nash equilibrium is high. Kauffman and Levin look at the problem of rule adaptation as a walk on a "landscape" where each rule corresponds to a point on the landscape and the "height" of that point is a measure of the utility of the sets of rules. The object of adaptation is maximizing utility—reaching the highest point on the

---

[49] Axelrod's (1984) study of emergent strategies in the iterated Prisoners' Dilemma provides a good example of co-adaptation in a set of rules; Howard's (1971) metagame approach, where a game is played over strategies rather than behaviors, is another example. The analysis presented there is based on work done in evolutionary biology; see Kaufmann (1993) in addition to the work cited in the text.

[50] A co-adaptive Nash equilibrium is local—there are no alternative sets of rules accessible from the current set without crossing an area of lower fitness, given the current rule choice by the opponent. In contrast, a Nash equilibrium in game theory places no restrictions on the proximity of other strategies.

landscape from a given starting point—and the walk is constrained by allowing only a limited number of rules to change at any one time. For example, if an initial rule set is

0 1 0 0 1 1

and movement is constrained to one bit at a time, the system could move to the rule sets

1 1 0 0 1 1	0 0 0 0 1 1	0 1 1 0 1 1
0 1 0 1 1 1	0 1 0 0 0 1	0 1 0 0 1 0

but not to any others.

Optimization proceeds by "hill-climbing": modified rules are maintained only if they are an improvement over the original. Under fairly general conditions[51], Kauffman and Levin show

> Along an adaptive walk, the expected number of *fitter neighbors* decreases by *half* when each improved variant is found. That is, the *numbers of ways "uphill"* is cut in half at each improvement step (Kauffman 1988,129; italics in original)

The analysis also shows that the number of local maxima in such systems is very large, these local optima are reached very quickly (that is, after only a few adaptations), and the choice of the local maximum is strongly dependent on the starting location.

Translated into organizational terms, this means that modifying a small part of the organizational repertoire will usually fail to improve its performance. The system has in all likelihood at a point where the interacting heuristics reinforce each other such that small changes will be ineffective. This does not mean, however, that the system is in an *optimal* configuration; to the contrary, it is most likely at a local maximum. Major successful transitions to equilibria with higher utility are possible but would have to occur due in one of two circumstances:

First, a major shift could occur through simultaneous change of a number of rules. In geometrical terms, this means a large "jump" in the Boolean space that, unlike the small jumps assumed by Kaufmann and Levin, might represent an improvement. The reorganization of the European political and economic system at the end of WWII is one example of such a shift.

Second, change in a single critical rule might trigger additional changes throughout the system.[52] For example, one could argue that the collapse of Communism in Europe in 1989-1991, as well as the subsequent breakup of the Soviet Union and Yugoslavia, were all the consequence of a single rule change: the decision, circa 1987, that the Soviet Union did not need to militarily control Eastern Europe in order to protect itself. This led to the renunciation of the Brezhnev Doctrine, which led to the collapse of Communism in Eastern Europe, which then triggered the dissolution of the Soviet Union itself.

---

[51] These apply for an "uncorrelated landscape", which means that values at one point have no relationship to those of adjacent points, adjacency being defined by the transition constraint. This is unrealistic by itself, but as Kauffman and Levin point out, a correlated landscape with large step sizes (i.e. changes in multiple bits are allowed) is equivalent to an uncorrelated landscape with small step sizes. The former situation — correlation but large allowable steps — probably characterizes most organizational choice situations

[52] This argument is pursued in more detail in Kaufmann and Johnsen (1992) and Kaufmann (1993) using a coevolutionary model more complex than that discussed here.

These cascades of change cause the actors in the system to go through a period of unsettled co-adaptation and learning, but the system will eventually settle into a new local equilibrium and new regularities. Based on this model, one would expect international behavior to be characterized by long periods of stability interspersed with short periods where a number of rules in the system are changing simultaneously.

Because of the lack of information concerning the utility that might result from changing a set of rules, precedent is a useful guide to decision making in a co-adapting system. *Ceteris paribus*, an organization's history provides a relatively accurate map of the value of points on the utility surface that have been visited in the past. Positive precedents (e.g. the WWII Allied alliance; Marshall Plan) show points that are desirable; negative precedents (e.g. the Great Depression, Munich, Vietnam) show points to avoid. It is generally safer to use known points to set policy rather than trying to explore new policy combinations, just as it is unwise at night to explore an area characterized by steep cliffs and an unstable surface with a weak flashlight.

Precedent failures—for example the Bay of Pigs or Lebanon 1982-83—occur when one follows directions that were plausible at one point on the landscape but inappropriate for the point that one actually occupies. Dead reckoning only works if one knows ones starting point. Precedent and organizational memory may also restrain the ability to *search* the landscape for new rule combinations because of the perceived dangers of experimentally evaluating certain points.

Conclusion

> *When, at the end of the last century, Wilhelm Wundt made the first serious attempt to turn psychology into a science, strangely enough [it] failed to develop along the lines of biology. Although Darwin's teaching was common property by then, although comparative methods and phylogenetic investigation were established procedures in all the other life sciences, they were persistently ignored by the new experimental psychology. It followed strictly the example of physics where, at the time, the atomic theory was paramount.*
>
> *Konrad Lorenz*

As noted at the beginning of this chapter, the theory of foreign policy decision-making developed here is specifically designed for implementation in computational models. Its objective is developing a model of politics consistent with human behavior that can be modeled algorithmically. Duplicating human *understanding* is impossible for the foreseeable future, until we provide machines with the full gamut of human experience, including adolescence (particularly adolescence...), but predicting human *behavior* is quite possible, just as we can predict the behavior of chickens without being chickens.

Because the modern physical sciences were built on the dual foundations of mathematics and experimentation with simple systems, the axiomatic/deductive approach has been enshrined as "scientific", and most attempts to systematically study political behavior assume that the axiomatic approach is the ideal. Axiomatic knowledge is quite useful in studying systems where the relationships between variables are relatively simple and information is relatively easy to obtain. Simple physical systems, simple economic relationships and artificial logical constructs such as mathematics and legal systems are all susceptible to analysis by axiomatic techniques.

Axiomatic knowledge fails, however, in systems that are either too complex to be modeled deductively or where critical information is missing. This is true for many physical and biological systems as well as social systems. When axiomatic knowledge fails, humans substitute a combination of pattern recognition and experimentation. The cathedral of Notre Dame was built without a theory of gravity; a baseball pitcher can throw a curve ball without knowledge of second-order differential equations, adenosine triphosphate or Bernoulli's law.

As will be discussed in more detail in Chapter 7, in trying to become scientific, the behavioralist approach to the study of international behavior took as a model the most abstract of the axiomatic sciences—physics and mathematics. In the absence of any major theoretical breakthroughs in the understanding of international behavior per se, this was probably premature. The study of international politics deals with human decision-makers whose pattern recognition abilities are exquisitely developed and who have constructed a political environment based on those abilities.

Classical international relations theorists such as Waltz and Bull have been bludgeoning formal modelers with this observation for decades. The classical criticism that international affairs are too complex for mechanical models is correct, but the cure suggested by traditionalists—abandoning formal models for a return to the glory days of informed intuition—tosses out the baby with the bath. Information-rich computational models provide one possibility for systematically modeling international behavior without sacrificing the advantages of formally specified theories.

Chapter 4
Rule-Based Models

McNamara asked [Chief of Naval Operations George] Anderson what he would do if a Soviet ship's captain refused to answer questions about his cargo. At that point the Navy man picked up the Manual of Naval Regulations *and, waving it in McNamara's face, shouted, "It's all in there." To which McNamara replied, "I don't give a damn what John Paul Jones would have done. I want to know what you are going to do, now."*

Allison 1971,131

Rule-based modeling efforts are currently the most common form of computational model. While much of the impetus for their development came from the success of expert systems in artificial intelligence in the early 1980s, rules are generally well suited to the study of politics since much of political behavior is *explicitly* rule-based through legal and bureaucratic constraints. Laws and regulations are rules: these may be vague, and they certainly do not entirely determine behavior, but they constrain it considerably.

Analyses of the Cuban Missile Crisis (Allison 1971) or US involvement in Vietnam (Gelb and Betts 1979), for example, repeatedly observe that the military options were limited by the standard operating procedures of the forces involved. The rules governing policy are usually complex and idiosyncratic rather than regular and parsimonious: the response of the United States to the kidnapping of a citizen in Ethiopia is very different than the response to a kidnapping in Italy. Systems of informal rules—"regimes" in international relations terminology (Krasner 1983)—impose additional constraints. In the Cuban Missile Crisis, Kennedy did not consider holding the family of the Soviet ambassador hostage until the missiles were removed but in some earlier periods of international history—for example relations between the Roman and Persian empires circa 200 CE—this would have been considered acceptable behavior.

Rule-based models reach inside the "black box" of the classical stimulus-response model (Snyder, Bruck and Sapin 1969). These models receive information describing the international environment, process that information using a set of rules, and produce a specific response as their output. While the internal workings of many of these models are complex, and the models may involve multiple agents or submodels[1], their overall structure generally reduces to that shown in Figure 4.1.

As noted in Chapter 3, the concept of modeling organizational behavior using logical rules is not new. Cyert and March (1963,chapters 7 and 8) develop several models—each involving about 100 rules—of organizational behavior in the microeconomic domain; cognitive maps of foreign policy (Axelrod 1976) are similar in structure and level of complexity to rule-based models; and the "operational code" approach (George 1969) attempts to summarize organizational behavior using a small number of rules. The expert systems

---

[1] Figure 4.1 is a very simple schematic but it underlies a number of more complicated models that, at first glance, seem to have little in common. Compare the diagrams, for example, in Thorson and Sylvan (1982,551), Hudson (1987,116), Sylvan, Goel and (1991,87), Majeski (1989,134) and Job and Johnson (1991,228).

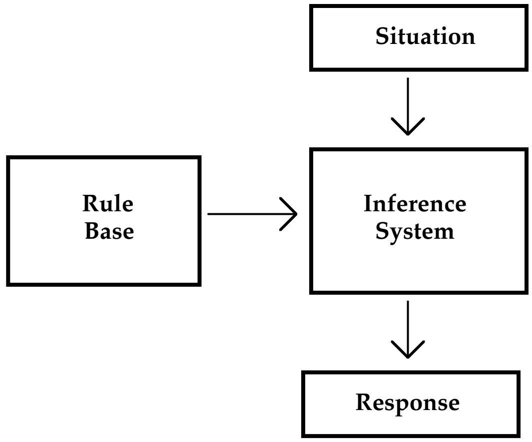


Figure 4.1. Structure of a Rule-based Model

developed in the 1970s and 1980s differed from these earlier efforts in using several thousand rules and hundreds of distinct data objects, an increase in complexity of at least two orders of magnitude.[2] This legitimated the idea of modeling behavior using very large and very specific systems of rules—in contrast to the prevailing norms of parsimony—as well as contributing the concept of using "knowledge engineering" to obtain information about rules from subject matter experts.

There are a number of rule-based models in the published international relations literature; these are listed in Table 4.1. This table is not comprehensive but includes most of the published models as of 1994, as well as some papers referring to models that are still under development[3].

---

[2]Benfer, Brent, and Furbee (1991) provide an excellent text-level introduction to expert systems oriented toward social science problems; Garson (1990) provides a shorter overview with extensive bibliographical coverage of both the literature and software. Buchanan and Shortliffe 1984; Buchanan and Smith 1989; Giarrantano and Riley 1989; Hashim and Seyer 1987; Hayes-Roth, Waterman, and Lenat 1984; and Klahr and Waterman 1986 are additional examples of the rather extensive literature on this subject.

Programming forward-chaining rule-based systems is straightforward and can be done in a variety of languages; Levine, Drang and Edelson 1990; Pederson 1989; and Sawyer and Foster 1986 are text-level introductions that emphasize basic algorithms. Backward-chaining systems are typically done in Prolog, a language designed with that application in mind; Clocksin and Mellish (1981) is the classical introduction to that language; .

[3] In addition to these published models, there are apparently a variety of additional models in the governmental and classified literature. However, my impression is that while AI methods have been extensively incorporated into *military* simulations and expert systems (see for example Andriole and Hopple 1988) there has been very little systematic effort in *policy* simulations.

Virtually all computer simulation invoke rules to some extent but there remains a fairly clear distinction between numerical and rule-based simulation in terms of the types of variables used. The two types can be distinguished by their output: if this is a line, then the simulation is probably numerical; if the output is a large set of discrete events, it is probably rule-based. Some hybrids exist—for example Majeski (1989) predicts budget request levels even though the internal structure of the model is rule-based.

Table 4.1. Rule-based Models of International Behavior

Model	Foreign Policy Domain	Type
Alker and Christensen 1972	UN peacekeeping	classification
Bonham and Shapiro 1976	US reactions to Middle East crises	cognitive map
Anderson and Thorson 1982	Saudi Arabian foreign policy	event driven
Thorson and Sylvan 1982	Cuban missile crisis	event driven
Tanaka 1984	Chinese foreign policy (CHINA_WATCHER)	classification
Banerjee 1986	Simulation of social structures	multiple agent
Hudson 1987, 1991	Foreign policy reactions	event driven
Majeski 1987	War initiation recommendations	classification
Schrodt 1988a	Simulation of balance of power system	multiple agent
Job, Johnson and Selbin 1987	US policy towards Central America	multiple agent
Kaw 1989	Soviet military intervention	classification
Katzenstein 1989	PRC relations with Hong Kong	classification
Majeski 1989	US defense budgeting	classification
Stewart, Hermann and Hermann 1989	USSR support of Egypt	classification
Mills 1990	Sino-Soviet negotiations	classification
Job and Johnson 1991	US policy towards Dominican Republic (UNCLESAM)	event driven
Sylvan, Goel and Chandrasekaran 1991	Japanese energy security policy (JESSE)	multiple agent
Taber 1992 Taber and Timpone 1994	US policy in Asia (POLI)	event driven
Katzenstein 1992	PRC relations with Taiwan	multiple agent

In its simplest form, a rule-based model is just a large set of *if…then* statements processed using Boolean logic. For example, a typical rule from Job and Johnson's UNCLESAM program has the form

IF US Posture to the Dominican Republic government > 4
and Stability Level >= 5
and Stability Level Change > 0
THEN
Increment US Use of Force Level by 1

The published foreign policy models typically have a few dozen rules. Mefford (1991) describes in considerable detail rule-based models that go beyond basic *if…then* formulations; the boundaries between these and other types of models—for example case-based reasoning (Kolodner 1988), machine learning systems and to an extent, expected utility models—are often fuzzy.

Rule-based models involve both *procedural* knowledge and *declarative* knowledge. For example, the rule

IF (X is an ally of Y) AND (Y is attacked by Z) THEN (X will attack Z)

would be procedural knowledge;

USA is an ally of South Korea

is declarative knowledge. In a model containing this information, the input

South Korea is attacked by North Korea

would generate the response

USA will attack North Korea

The expert systems literature distinguishes between "forward chaining" and "backward chaining" as a method of logical reasoning. Forward chaining starts with a set of characteristics about a situation—a feature vector of independent variables—and applies these as needed until a conclusion is reached. The example given above uses forward chaining, as does the ID3 machine-learning algorithm discussed in Chapter 5.

Backward chaining, in contrast, starts with a possible conclusion—a hypothesis—and then seeks information that might validate the conclusion. The information used could be in the form of other rules whose consequence, if true, would validate some antecedents of the hypothesis under consideration, or declarative knowledge in the system, or, in an interactive system, requesting specific additional information.

Backwards chaining is a much newer technique than forward chaining, and was adapted from earlier AI work on automated theorem proving. The technique was popularized during the early 1980s for several reasons. First, it was the method used in MYCIN, a medical diagnosis program that did much to popularize expert systems (Shortliffe 1976). Backward chaining shares characteristics with human expert reasoning: for example a doctor attempting to diagnose a disease will tend to collect information based on a small number of hypothesized diseases, rather than ordering every possible test. Second, backward chaining was embodied in the PROLOG programming language, which was widely publicized through the [ultimately unsuccessful...] Japanese "Fifth Generation" AI project (Feigenbaum and McCorduck 1983). Third, backward chaining lends itself easily to providing explanations as to why information is being requested, since the backwards-chaining algorithm is pursuing an identifiable line of inquiry that ends with a specific solution.

Backwards chaining has obvious applications in political analysis. For example, an analyst studying peasant insurgency in the southern Philippines might be interested in which outside groups could be financially assisting the insurgency. The analyst will form a set of hypotheses and then seek information that might confirm these. For example, the analyst might first ask whether the group's motivations are primarily religious, political or economic. If evidence shows the group to be primarily Moslem, the analyst would look for evidence of support by Saudi Arabia or Iran, whereas if the group seems to be primarily Maoist, Saudi Arabia and Iran would not be likely sources of support and different evidence would be sought. In backward chaining, reasoning is hypothesis-driven rather than data-driven.

Despite the popularity of backward chaining in the AI literature, most foreign policy simulations are forward chaining; Katzenstein (1989,1992) and Mills (1990) are exceptions. Most foreign policy models are attempting to simulate the policy processes that tend to be input driven. These models are also predictive (or post-predictive) rather than diagnostic. A model designed to advise policy rather than simulate it would presumably make more use of backwards chaining to support or refute hypotheses, and Mills (1990) provides such a model in a simple form.

Categories of Rule-Based Models

Rule-based models tend to be complex, idiosyncratic and difficult to summarize because any given model is likely to embody a number of different mechanisms. Nonetheless, it is possible to classify the existing models into three broad categories based on their structure and objectives. This section will provide a brief description of each of those categories along with a few examples. My comments on the individual models highlight only a few of their most salient features; the reader is urged to consult the original articles for greater detail.[4]

Classification Models

The simplest form of rule-based model is the classification tree. Each of the nodes of the tree contains a question, and the branches of the tree correspond to answers to those questions. Classification trees are familiar as a means of everyday knowledge representation. Field guides for birds, wildflowers, mushrooms and other natural objects allow one to "key out" an identification through specific questions. For example, to identify a dandelion, a typical field guide would go through a series of questions such as:

---

[4] I've differentiated these categories according to the process being modeled; interestingly, one gets a similar breakdown based on the formal algorithmic structure: classification models use trees; event-driven models use production rules; and the multiple-agent models use frames and other complex knowledge representation structures.

Color: yellow
Blooms in: spring, summer
Flower type: compound
Leaf shape: [series of shapes]
etc.

Eventually one reaches a point where the identification is reduced to a small set of candidate species and the final classification can be done by comparing the actual plant to a small number of pictures. This is a forward chaining process.

A good key starts with the most obvious information—that with the least acquisition cost—and then proceeds to the more detailed information that provides the most discrimination among the possible cases to be classified. In its use of information, a key differs substantially from a biological classification. For example, a yellow dandelion has more in common biologically with a white daisy or blue chicory than it does with a yellow jewelweed. Nonetheless, because information on "color" is acquired at virtually no cost while the biological concept "compound flower" is more difficult to convey, "color" is the characteristic used first in the key. The classification key approach is also commonly used in guides for the diagnosis and repair of cars, computers, children, household appliances and so forth.

In principle, a tree with N levels of questions and K branches per question can classify K^N categories; e.g. a binary tree (with two branches per question) with ten levels of questions can classify 2^{10}=1024 categories. While the complete tree involves $K^{(N-1)}$ questions, only N questions are required to categorize a case, so a tree is computationally efficient as a means of storing classification information. Furthermore, such systems capture some of the characteristics of human problem solving, specifically the application of idiosyncratic rules and pattern recognition. Since rules can be stored and accessed through associative memory, and only the small feature vector needs to be retained in working memory, this approach is quite compatible with human cognitive constraints. Rule-based classification substitutes a principle of *modal* (most common) behavior in small samples for the statistical principle of *average* behavior in large samples. Rule-based systems effectively partition the universe of cases into very small subsets determined by the independent variables of the feature vector, and then predict or classify the dependent variable using the modal value within each of those subsets.

The earliest use of a formal classification tree model in international relations was the Alker-Christensen-Greenburg model (1972, 1976) of United Nations dispute resolution. Some cognitive maps are similar in structure to classification models; Bonham and Shapiro's (1976) cognitive map model of a foreign policy analyst's recommendations concerning a crisis in Jordan is another early example of such a model.

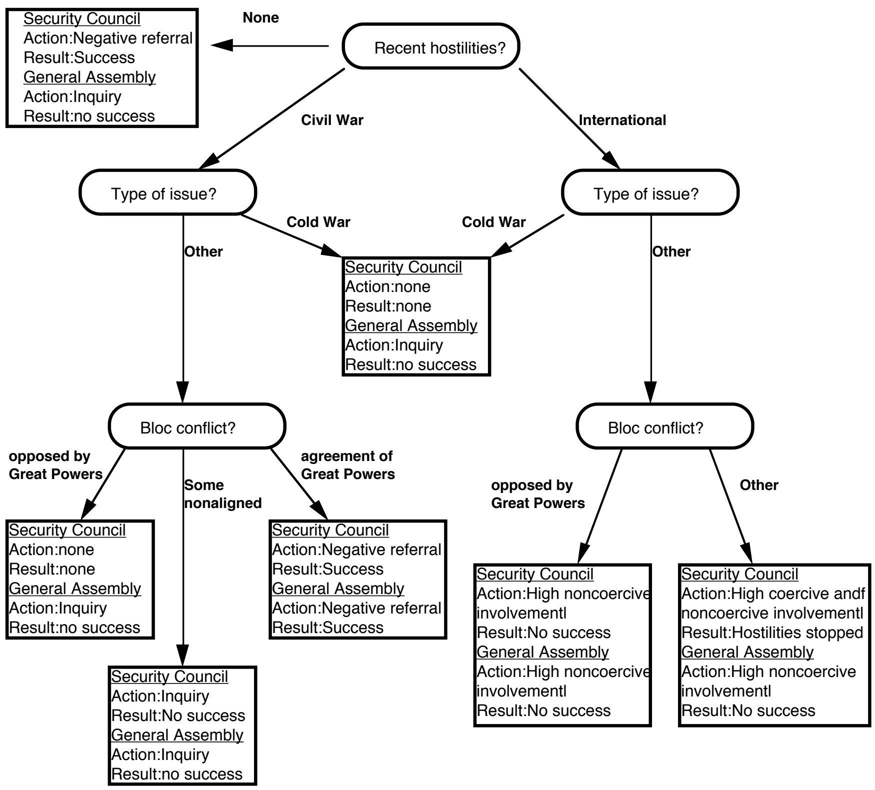
Figures 4.2 and 4.3 show parts of the classification trees from Alker and Christensen (1972,214)[5] and Kaw (1989,409) respectively. In each case, a variety of qualitative characteristics of the situation are used to determine branching behavior within the tree; the model determines qualitative outcomes. In both of these models—as well as the other instances listed in Table 4.1—the structure of the tree was determined

---

[5] This classification tree is only one part of the Alker and Christensen effort, which also involves a numerical simulation and a very innovative system of learning and precedents in addition to the rule-based component. This part of the model is due solely to Alker.

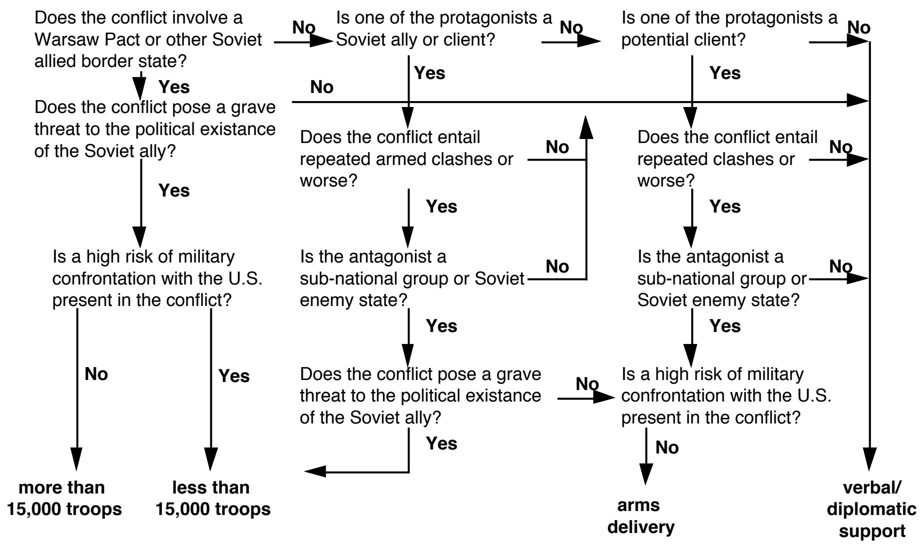
by the researchers through study, intuition and experimentation, though as discussed in the next chapter, algorithms exists that allow a tree be to developed directly from a set of data.

Despite their conceptual simplicity, classification models have proven to be able to successfully model behavior in a variety of domains. Furthermore, most of these models are relatively simple: for example Kaw's model uses 11 rules to predict 403 cases; Alker's uses 7 rules to predict the actions and outcomes in 90 types of UN peacekeeping actions. While the potential exists for tautological modeling through the excessive use of rules, this is not the case in the existing models, where the number of cases considered is typically at least ten times the number of rules used in the classification.



Source: based on Alker and Christensen 1972:214

Figure 4.2. Alker Models of UN Peacekeeping Actions



Source: Kaw 1989:409

Figure 4.3. Kaw Model of Soviet Intervention

Event Driven Models

The second general category of rule-based models focuses on events; these models tend to use a production system approach and simulate the behavior of a single actor.[6] The rules in event-driven models deal with how an actor will respond to various situations, and typically also include the actor's own anticipatory model of the actions of other actors in the system.

The production system approach is based on the cognitive model developed by Newell and Simon (Newell and Simon 1972; also see Simon 1979, 1982). Holland et al succinctly summarize this approach:

> [T]reat problem solving as a process of search through a *state space*. A problem is defined by an *initial state*, one or more *goal states* to be reached, a set of *operators* that can transform one state into another, and *constraints* that an acceptable solution must meet. Problem-solving methods are procedures for selecting an appropriate sequence of operators that will succeed in transforming the initial state into a goal state through a series of steps. Some methods are highly specialized for solving problems within a particular domain (for example, a procedure for solving quadratic equations), whereas others are generalists that can be applied across a board range of problem domains.
>
> A handful of general methods have been described (Newell 1969) which are closely related to each other. The most representative general method is "means-ends analysis", a procedure consisting of four steps:
>
> 1) Compare the current state to the goal state and identify differences.
>
> 2) Select an operator relevant to reducing the difference.

---

6 Production system models that deal with multiple actors will be discussed below.

3) Apply the operator if possible. If it cannot be applied, establish a subgoal of transforming the current state into one in which the operator can be applied. (Means-ends analysis is then invoked recursively to achieve the subgoal).

4) Iterate the procedure until all differences have been eliminated (that is, the goal state has been reached) or until some failure criterion is exceeded. (Holland et al 1986, 10)

As with the classification trees, this approach is motivated by the human cognitive constraints of an effectively unlimited memory but a modest capacity for logical reasoning. Production system models can be implemented and tested on even small computers with relative ease, and this approach characterized much of artificial intelligence research prior to the 1980s.

While few researchers would accept the Newell-Simon approach as a complete description of human problem solving behavior, it forms a good point of departure. Carbonell's (1978) POLITICS model, which is a very simple ideological simulation, is a production system model, as were the early foreign policy modeling efforts at Ohio State (Thorson and Sylvan 1982; Anderson and Thorson 1982). In the production system model

> The decision module of the simulation maintains a data structure containing the current description of the environment.… called the ambient information structure (AIS). … The AIS contains symbols [representing] a description of the current environment. Examples of state knowledge elements include: "Significant military threats exists", "The US is willing to provide arms", "Nation X is hostile"… At any point in time, the AIS contains those state knowledge elements that describe the environment from the perspective of Saudi Arabia.
>
> The behavior generating part of the simulation consists of a set of decision rules. The contents of the AIS determine which rules are to be executed, and, thus, which behaviors the simulation will exhibit. (Anderson and Thorson,182-183)

When a decision rule is executed, it modifies the AIS, which may then trigger additional decision rules. Other decision rules periodically run automatically; for example in the Anderson and Thorson simulation, a module decides on the oil production for every "month" of the simulation.

In event driven models of foreign policy, rules are organized primarily by the actions that actors in the system might take or else rules respond to declarative knowledge dealing with the attributes of other actors (e.g. is it an ally or adversary). In addition, each simulation has a variety of "objects" to which the rules may refer.[7]

---

[7] I am standardizing the terminology here, which is anything but standard in the field. In Thorson and Sylvan (1982), objects are called "vocabulary" and events are called "semantic kernals"; in Taber (1990) objects are called "actors" and events "verbs". This absence of a standard vocabulary stems from the fact that knowledge representation concepts have yet to jell within their parent discipline of computer science. As Doyle points out:

> Past work [in AI] was often creative but rarely precise, since all that mattered was that the general idea of the work be suggestive enough to other, usually personally known, researchers so that they could figure out how to place it within their own favorite vocabulary and methodology. (Doyle 1988,20)

Various combinations of events and objects, modifying the model environment as a result, trigger the rules. Taber summarizes this process

> The production rules allow reasoning from an interpreted event (or from a change in the regional situation that might not be strictly an event) to some policy output through inference chaining. These rules take the general form of IF-THEN statements. Some take situations as their conditions (IF an important enemy attacks an important friend…), while others takes goals as their conditions (To oppose the attacker…). In addition, some give goals as their consequences (…, THEN oppose the attacker), while others given executable actions (…, THEN attack the attacker). Specific executable behaviors are recommended by reasoning through an inference chain from the input event to the output policy recommendation. (Taber 1990,5)

Table 4.2 provides partial lists of the type of actions considered by Thorson and Sylvan (1982) and Taber (1990,n.d.) respectively

---

In contrast, the behavioralist adoption of statistical techniques and the rational choice adoption of microeconomic techniques borrowed from mature fields.

Table 4.2. Objects and Events from Event-Driven Models

Objects and Events from Thorson and Sylvan (1982)

Objects

Actors: US, USSR, CUBA, BERLIN, TURKEY, US-SHIPS, USSR-SHIPS
Type: MISSILES, ARMED-FORCES
Intensity: HIGH, MEDIUM, LOW
Channel: PUBLIC, PRIVATE

Events

Type I: NOACTION, RUNBLOCKADE
Type II: BLOCKADE, INTERCEPT, INVADE, MINE, NUKE, SEIZE, MILMOVE
Type III: MILBUILD, AIRSTRIKE, ATTACK
Type IV: WARN, OFFER, ACCEPT, NEGOTIATE, CONDEMN

Objects and Events from Taber (n.d.)

Objects

Actors: afghanistan, albania, bangladesh, bhutan, brunie, burma, cambodia, ceylon, chine, east germany, … , timor, us, ussr, west germany, west irian, west pakistan [50 actors in total]

Events

Verbs: attacks, physically threatens, verbally threatens, threatens, invades, assaults, aggresses against,…, extorts value from, economically threatens, accuses of violating agreement, accuses of breaking truce, …, breaks truce, violates treaty, violates agreement
[90 verbs in total]
Context Statements: problem is getting worse, problem is getting better, regional war, regional war involving us, war, war involving us, world opinion is against us, target is weakening

The event driven approach is attractive for at least three reasons. First, the *structure* of the rules is actually simpler than the trees found in the larger classification models or the complex knowledge representation structures found in multiple agent models. The use of natural language verbs to describe events means that the rules can oftentimes be derived directly from policy statements made by the decision-makers being modeled: for example Taber is able to use content analysis to assign confidence factors to rules comprising the various paradigms in POLI by looking at the number of times a rule was used by speakers in the *Congressional Record*.[8]

Second, the *output* of an event-driven simulation can be more complex than the outcome of a classification model or a numerical simulation, and will often correspond fairly closely to the actual historical record. In other words, the actions predicted by an event-driven simulation will look similar to those recorded in a journalistic account, albeit with a somewhat limited vocabulary. Numerical simulations, in contrast, usually

[8] POLI implements three different decision-making paradigms: militant anti-communism, pragmatic anti-communism and isolationism.

compress a variety of behaviors into an aggregated score rather than predicting discrete behaviors. For example in the Thorson and Sylvan (1982) model of counterfactual outcomes in the Cuban Missile Crisis, the policy outcomes involve specific events such as "Soviets launch a medium intensity airstrike against Turkey", whereas a numerical simulation would typically show an increase in a conflict variable consistent with "medium airstrike" without specifying the target or means of attack. A number of models have demonstrated that a relatively small number of rules are sufficient to produce remarkably complex and realistic behavior.

Finally, both the rules and output of an event driven simulation are easier to interpret than comparable elements in a numerical simulation. Some translation is required to go between natural language and the formal specification, but the formal language —typically LISP—involves verbs and proper nouns: an airstrike is "airstrike", Turkey is "Turkey".

Multiple Agent

The multiple agent models are the most complex form of rule-based model and, in contrast to the other two categories, have little in common with each other except for the use of frame-based specifications. Many of these models are highly complex and while they rest on a rule-based structure, their primary focus is on the interactions of the component parts of the model rather than on the rules per se. Multiple agent models are also more likely to try to achieve process validity as well as outcome validity.

The simpler form of a multiple agent model allows two or more event driven models to interact with each other; Banerjee (1986)[9] and Schrodt (1988a) are examples of this approach. The behaviors modeled are simplified—in particular the decision rules and allowable actions are quite limited—but because multiple agents interact, the system can exhibit a self-contained evolution over time, rather than focusing on a single decision or crisis. These models are similar in their objectives—though much simpler in structure—to numerical global models such GLOBUS but they deal with qualitative rather than quantitative behavior.

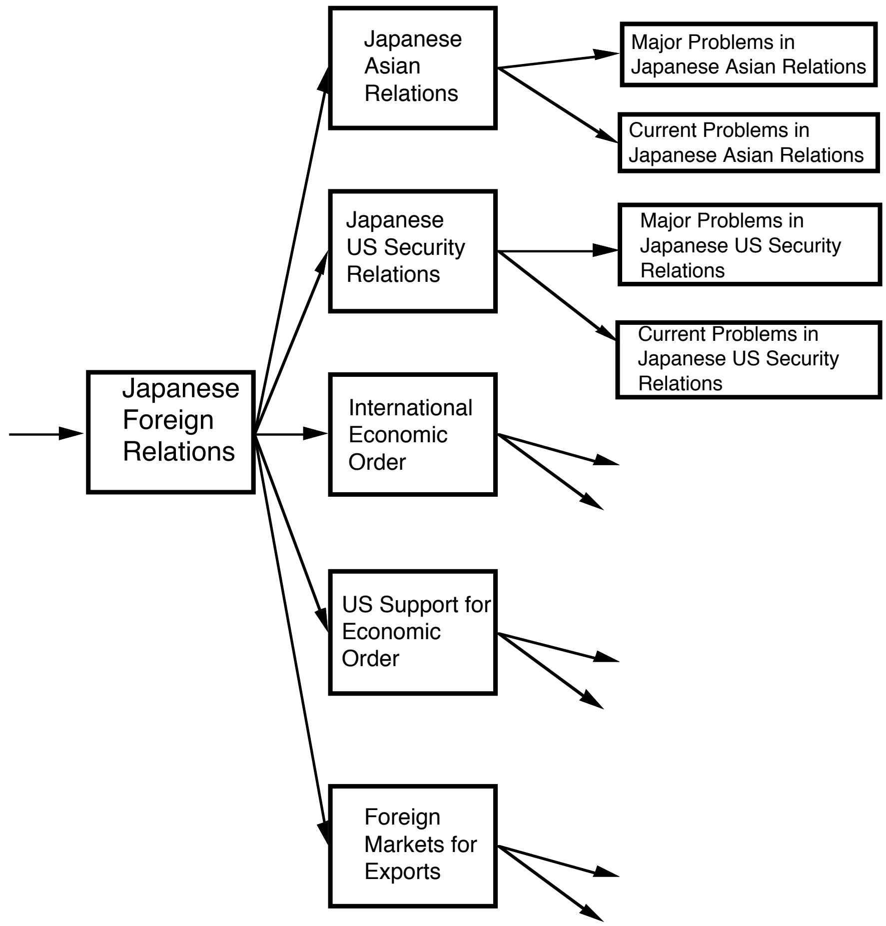
More complex are systems that attempt to model the internal debate within a bureaucracy. While the simpler event driven models can generate "debate" through the application of various production rules or by using backwards chaining, a multiple agent system explicitly models multiple paths of bureaucratic consideration.[10] For example, in Sylvan, Goel and Chandrasekaran's (1990, 1991) JESSE model of Japanese energy security policy—easily the most complex completed multiple agent model in the published literature—the problem of foreign policy is decomposed into a set of subproblems, as illustrated in Figure 4.4. The complexity of the decision-making environment, in turn, means that the information used by the simulation itself is complex.

---

[9] Banerjee develops a model of social structure development rather than foreign policy but its underlying structure and methodological justification is very similar to the other models discussed here.

[10] "Blackboard systems" (see Nii 1989) are another architecture—inspired by bureaucratic problem-solving—that would seem natural for bureaucratic modeling. In such systems, a number of semi-autonomous computational agents work together on a problem, sharing a common memory.

JESSE is based on Chandrasekaran's (1986) theory of "generic tasks" and with frames containing information on various policy options; Table 4.3 gives one of those frames



Source: Sylvan, Goel and Chandraskaran (1990:82)

Figure 4.4.Policy Options in JESSE

Table 4.3. Hypothesis Frame from JESSE

Hypothesis: FlowDueToChangeInExportCapability
Superhypothesis: EnergyFlow
SubHypotheses:
MajorChangeInExportCapability
ImmediateChangeInExportCapability
CostDueToChangeInExportCapability
Ask (Yes,No, Unknown):
Q1: Has there been a decline in the enregy production capability of an energy-exporting country?
Q2: Has there been a decline in the energy transportation capability of an energy-exporting country?
Rule 1:
If (Q1 = Yes) or (Q2 = Yes), then hypothesis very likely;
otherwise Rule 2
Ask (Yes,No, Unknown):
Q3: Is there likely to be a decline in the energy production capability?
Q4: Is there likely to be a deline in the energy transportation capability?
Rule 2:
If (Q3=Yes) or (Q4=Yes), then hypothes is likely;
If (Q3=Unknown) or (Q4=Unknown), then hypothes is uncertain ;
otherwise hypothesis is unlikely.
Actions: If hypothesis is very likely or likely,
then establish hypothesis;
activate subhypotheses;
else reject hypothesis

Source: Sylvan, Goel and Chandrasekaran (1990,92)

While a complex system such as JESSE *could* be implemented as a rule-based system (Sylvan, Goel and Chandrasekaran 1990,104), and the system represents part of its knowledge as production rules, the system goes far beyond framework. As those authors note:

> There are several important differences between [JESSE and production systems]:
>
> First, our analysis of Japanese energy-related decision making is at the level of the generic information processing tasks. … Since the language of production rules is at a lower level of abstraction, it does not offer any constructs for capturing these task-level distinctions.
>
> Second, the knowledge in JESSE is organized around concepts that are represented as frames. … This leads to a more "molecular" form of knowledge representation and information processing. Rule-based systems, in contrast, have a more "atomic" representation.
>
> Third, the semantics of the concepts is different in the various modules of JESSE: the concepts are hypotheses in the first and second classification modules, indexical categories in the third classification module, and plans in the fourth module.

> Fourth, the generic task framework explicitly specifies the abstract control of processing for each generic task.
>
> Finally, the generic task methodology enable more efficient processing, easier knowledge acquisition, and more perspicuous explanation of knowledge and processing.
>
> (Sylvan, Goel and Chandrasekaran 1990,104)

The clear advantage of a system such as JESSE is its internal complexity, which allows for a far richer set of behaviors than one can generate using a simpler model. The obvious disadvantage is the amount of effort involved: the source code for JESSE runs to some sixty pages, and since this code is written in a specialized high-level language designed for the implementation of complex expert systems, the equivalent implementation of the system in LISP or C would be much longer. The "knowledge acquisition" phase of the system's development was also complex and based on substantial earlier interview work (see Bobrow, Sylvan and Ripley 1986). Based on the JESSE experience, the labor required to develop a large multiple agent model is probably of roughly the same magnitude as that required to develop a large numerical global model.

Uncertainty in Rule-Based Systems

Most contemporary rule-based systems have some means of representing uncertainty, but the methods used vary considerably. For historical reasons, the development of expert systems proceeded largely uninformed by statistical decision theory, so the resulting approaches are something of a hodgepodge. In general, three methods have been widely employed—traditional Bayesian logic, Dempster-Shafer logic, and fuzzy sets—with the last emerging as the clear favorite; Giarrantano and Riley (1989) provide a good survey of these methods.

For the mathematically trained, the most straightforward approach to representing and processing uncertainty are the Bayesian methods of formal probability theory: prior beliefs (e.g. Nation X believes that Y is very likely to attack but Z is not very likely) are assigned *a priori* probabilities. Bayesian analysis is very well developed mathematically, yet this approach has proven problematic in rule-based systems. As Rothman notes:

> Implementing a Bayesian uncertainty management system is complicated by problems. First, [it] requires knowledge of all of the *a priori* probabilities. Worse, these probabilities are often inconsistent when elicited from human experts. Specifically, the probability of hypothesis H in the absence of any other evidence E' should be the prior probability P(H|E'); this is often not the case. As a result, updating a node with evidence in support of a hypothesis may actually lower the *a posteriori* probability of H. … The complex interactions between the various probabilities may make modification of the knowledge base unwieldy, since the sum of the probabilities must remain constant. Bayesian reasoning assumes that all possible outcomes are disjoint, an oft-violated assumption. For example, a person may suffer from more than one disease or a machine may suffer multiple component failures. (Rothman 1989,59)

In general, the sheer complexity of rule-based systems tends to make Bayesian methods unwieldy, though the method is still appropriate in situations where the input and output of the system is simple and well defined.

A widely used alternative for dealing with probabilities in a complex domain is Dempster-Shafer logic (Shafer 1976), a probability management technique. The Dempster-Shafer technique maintains many of the concepts and vocabulary of probability theory but relaxes the axiom

Prob(X) + Prob(not X) = 1

and separates the issues of the "belief" that a proposition is true from the "plausibility" that it is true based on a specific line of evidence. Because the Dempster-Shafer method was design from the outset to allow uncertainties to be updated without generating logical inconsistencies, it is easier to manage in large systems than Bayesian logic.

By the early 1990s, the most common system of uncertainty management in use in expert systems was fuzzy logic, a concept originally introduced by Zadeh (1965; Zadeh et al 1975). This idea has already been applied in a number of formal models of international behavior (see for example Cioffi-Revilla 1981, Sanjian 1988a, 1988b) and is now widely used in expert systems (see Negotia 1984, Zimmerman 1991, Kandel 1992, Yager and Zadeh 1992)[11]. The basic idea behind fuzzy logic is the ability to assign a "partial truth" value to an assertion. Taber's POLI model uses fuzzy logic:

> Reasoning in POLI is fuzzy. The conditions in a production rule, and the rule itself, carry partial truth value. In Boolean logic, of course, such conditions must be either true or false, leaving no room for degrees of truth, which would be a serious inflexibility in a belief system such as POLI. Decision makers clearly have uncertain beliefs and use verb hedges to qualify their beliefs; a reasonable model of this process should allow reasoning from partial truth.
>
> Consider the rule "IF a person is tall THEN that person is fat" with a CF (certainty factor) of .3. This rule means that the expert system has a .3 confidence in the conclusion "that person is fat" when it is certainly true "that person is tall". When the system receives information that "Harold is tall" with CF=0.9, we can infer the information that "Harold is fat" with a calculable degree of belief. Through an application of the compositional rule of inference we compare the CF of the new information (0.9) with the CF of the known relationship (0.3), taking the smaller. (Taber 1990,5)

Note that while the certainty factors in fuzzy logic are stated as probabilities, the rules of inference are quite different from probability theory.

While the concept of fuzzy logic was initially somewhat controversial, it seems to be winning in both the intellectual and commercial marketplace. This is probably due to several factors. First, fuzzy logic is a complete logical system that turns out,

---

[11] If Japan is a leading indicator, this term, like "zero-sum", may be on its way to entering the popular vocabulary: In the summer of 1991, "fuzzy logic" was being used as a selling point on consumer products such as video cameras, washing machines and automatic rice cookers in the Akihabara electronics district in Tokyo.

mathematically, to encompass both Aristotelian and Boolean logic as special cases, rather than being *ad hoc* and designed solely for expert systems. Second, the method is quite straightforward to implement computationally. Third, the uncertainty statements used by fuzzy logic correspond more closely to the use of uncertainty in human reasoning than do many alternatives, and translation between human uncertainty judgments and fuzzy statements is straightforward. Fourth, and not least significant, systems built using fuzzy logic produce desired behavior—they work—and do not become bogged down as the system becomes more complex. Unless some superior system comes along, fuzzy logic systems, or variations based on similar principles, will probably become the most common means of handling uncertainty in rule-based systems.

Outcome and Process Validity in Rule-Based Systems

The information and processes used in rule-based models are much closer to the written record of foreign policy decision-making than the processes found in a regression equation, numerical simulation or even most rational choice models. Computational models require fewer "as if"'s, and as a consequence, a number of researchers have suggested that process validity should be a major concern (e.g. Mefford 1991). More generally, the complexity of the information processing in a computational model allows for meaningful comparisons with the internal dynamics of an organization, whereas most statistical and rational choice models fall back on unitary actor assumptions. The rules invoked by the model often have much the same form and vocabulary as those found in the discourse of policy debate, and in some models they have been derived directly from records of that debate (e.g. Taber 1992). Similarly, the ability of an rule-based model to work through chains of inference of considerable complexity and to consider multiple lines of inference has much in common with observed policy processes. Rule-based systems provide a sophisticated form of rationality and are a useful illustration of the fact that a rational model does not have to be an expected utility model.

In some respects, little distinction needs to be made between process validity and outcome validity: process validity is simply outcome validity in the component parts of the model. Any foreign policy decision has multiple layers—levels of analysis as it were—and consequently can be decomposed into subprocesses involving ever-smaller decision-making units. For example, the process of setting the U.S. defense budget can be decomposed into the various stages of budget requests and revisions, and each of those components can be modeled separately, an approach used in Majeski (1989). Similarly, the pattern of information flow within an organization could be modeled and, with a suitable metric, compared to the information flow found in a set of memoranda.

Rule-based models that have been explicitly tested do fairly well in terms of outcome validity; Table 4.4 presents some of these results.

Table 4.4. Outcome Validity in Rule-based Models

Model	Behavior Predicted		N	Accuracy
Tanaka 1984	Chinese verbal behavior in crises 1949-1978		383	82%
	Chinese physical involvement			60%
Hudson 1991	Foreign policy behavior in CREON		6605	
		Affect		49%
		Commitment		24%
		Instrument		62%
		Recipient		63%
Kaw 1989	Soviet military intervention 1950-1987		403	88%
Majeski 1989*	US defense budgeting 1954-1979		26	
		Armed services proposal		65%
		Presidential proposal		54%
		Congressional appropriation		61%
Taber 1990	US policy in Asia 1949-1960		161	86%

* Unlike the event-oriented models, the Majeski model predicts specific budget levels; the "accuracy" figures reported here are the ratio of the mean absolute deviation of the rule-based model to the mean absolute deviation of a linear model, i.e. the rule-based model generates roughly half the error of the linear model on Presidential budgeting, and if the model were perfect, the "accuracy" figure would be 0.

The record on process validity, in contrast, is more mixed.[12] For example, in discussing their assessment of the JESSE model, Sylvan, Goel and Chandrasekaran observe

> Our results [comparing JESSE to two actual energy crises] show that the performance of JESSE is reasonable. By this we mean that the outcomes of JESSE are generally in line with what actually transpired. Nothing resembling "contradictory" circumstances took place. … We have demonstrated the system to a few domain experts. … Their judgment so far has been that the energy policy decision-making process followed by JESSE is plausible.…We have not chosen to undertake any quantitative statistical tests. (Sylvan, Goel and Chandrasekaran 1991,97)

Along a similar vein, Job and Johnson note that in examining

> UNCLESAM's capacity to replicate the four major escalatory sequences of U.S. action during 1959-1965, we find that UNCLSESAM performed quite "satisfactorily". Thus for the crises of 1961 and of 1965, and in the two other situations involving U.S. uses of force, UNCLESAM accurately reproduced a record that mirrored U.S. actions in these circumstances (Job and Johnson 1991,238)

[12] Job and Johnson cite Mallery (1988) to justify not using formal tests; Sylvan, Goel and Chandrasekaran cite Alker's (1975) discussion of computational hermeneutics.

While process validity is an attractive objective, it presents a number of difficulties, and some situations the pursuit of process validity to the detriment of outcome validity may be an instance of the best being the enemy of the good. The problems facing the quest for process validity include the following:

First, reality is a low probability event. The likelihood of correctly duplicating the idiosyncrasies of an individual foreign policy decision, particularly one involving small group decision-making, is exceedingly low. A regularized decision process such as credit approval is quite different than crisis decision-making, where even outcome validity presents a challenge. A foreign policy model of a *regularized* procedure—for example visa approval by the US embassy in Brunei—could probably be achieved, but such processes provide only limited theoretical insights.As noted in Chapter 2, extensive research indicates that human probabilistic reasoning differs significantly, and quite idiosyncratically, from any formal system yet devised for handling uncertainty. The various forms of probabilistic reasoning incorporated into rule-based models only approximate, rather than duplicate, human reasoning in comparable situations. Those approximations are clearly sufficient to produce accurate outcomes but it seems unlikely that they will duplicate process, particularly in situations characterized by a great deal of missing and noisy information.

Second, the sheer complexity of any bureaucratic decision that has not been deliberately and rigorously regularized poses another challenge to process validity. The types of decisions commonly studied in rule-based models—major power decision-making in a regional or topical domain—probably involves hundreds of individuals working with thousands of discrete items of information and communicating with tens of thousands of internal memoranda over the course of several months. Despite all of this flurry of activity, the end result may be relatively simple—so outcome validity can be assessed—but the process itself is not simple. At the present levels of resource investment in political science modeling, it seems unlikely that we will have either the information, time or labor required to duplicate systems at this level of complexity.

Finally, it is not clear that a record even exists to test a model for process validity. Former Johnson aide Dean Rusk comments

> … the written record reflects only a portion of the thoughts in the minds of those who are making decisions and of the content of discussions among themselves which do not appear in the written record. Any foreign policy question of any significance has within it dozens upon dozens of secondary and tertiary questions and the minds of policy officers run through a very extensive checklist of such elements — regardless of what the written record shows. (quoted in Neustadt and May 1986,xvi)

This observation cuts two ways. First, if a model correctly mimics the written record of a decision , this means only that the model reflects accurately how those in authority wanted the decision-making process to appear. The recorded decision-making, almost always written with the benefit of hindsight, is likely to appear more logical, ordered and devoid of dead ends than the actual process, and such a simplified process will be easier to model than the actual decision-making. Alternatively, if the model does not reflect the written record, it might still, in fact, be reflecting the actual process, but there is no way of knowing this.

When a complete record has been preserved, the discourse we observe within an organization is the result of the interaction between the limited bandwidth in the organization and the very broad bandwidth at the subcognitive level in individuals. We observe only the *transmission* of pattern information in political discourse—not the entire knowledge base—and the paper trail of debates, meetings, position papers and memoranda is highly idiosyncratic because of the desire for efficient communication, the joint effects of shared tacit information and the various possible states of the system. It only partially describes how a decision was made because the subcognitive political reasoning in an individual is never expressed in language.

Communication using natural language is serial, relatively slow, and often inefficient, and therefore it will be used as little as possible. An individual skilled in persuasion will attempt to transmit only that information which maximizes the likelihood of activating patterns leading to a desired set of actions by the recipient. This means making efficient use of shared knowledge.

Consider the cases of intervention in Laos in the early 1960s and the re-flagging of Kuwaiti tankers in 1987, the cases studied by Majeski and Sylvan, and Boynton (n.d.). A military briefing on the Loatian situation would not start with the information "The Communists are our enemies" or "A victory for Communism would be bad for the United States". While logically necessary for an argument, this information is already shared by the relevant decision-makers and is therefore never mentioned. In the decision to re-flag Kuwaiti tankers, in contrast, the security of Kuwaiti shipping had not previously been a major United States foreign policy objective and therefore it had to be made explicit.

The pattern being matched is the same in both cases: the difference is whether that information is tacit (shared) or explicit. The full pattern would look something like

General pattern	Laos	Kuwaiti tankers
X threatens the US	tacit	explicit
US military force would reduce this threat	explicit	tacit
We don't want escalation	tacit	tacit
There is little risk of escalation	explicit	explicit

While the general pattern is roughly the same in both cases, the arguments present in the discourse are quite different because of the differences in tacit information.

This situation can be generalized further: if we use the schematic representation of any pattern as a binary vector and assume for the moment that each feature has only two states, then discourse takes place on the exclusive-or (XOR) of the two vectors:

Desired vector	1 1 1 0 0 0 1 0 0 0 0 0 1 1
XOR	
Tacit vector	1 1 0 1 1 1 1 0 0 1 1 0 0 1
Discourse vector	0 0 1 1 1 1 0 0 0 1 1 0 1 0

Skillful rhetoric involves the activation of a series of associated patterns to reinforce the validity of the main message. This leads to considerable redundancy in the message from a strictly logical standpoint, but that addition information serves a clear cognitive purpose.

Consider the problem of establishing the first pattern component above, "X threatens the US". In the case of Laos in 1960, the substitution rule "X = Communists" establishes this immediately. In the case of the Persian Gulf, this first point is decomposed into a series of subpatterns that must be satisfied. In the Kuwaiti tanker case, the subpatterns went something like

Iran is an enemy of the US

Kuwait is a weak country compared to Iran

Kuwait is seeking aid to protect its oil tankers from Iran

The Soviet Union is offering aid

The Soviets will gain influence with Kuwait if the USSR aids Kuwait

The Soviet Union is an enemy of the US

Kuwait is an ally of Saudi Arabia

Saudi Arabia is an important ally of the US

Saudi Arabia is an enemy of Iran

The US will gain influence with Saudi Arabia if the US aids Kuwait

I have deliberately not presented a deductive argument here . Aspects of these patterns are redundant and there are multiple paths that could trigger the subpattern as being equivalent to a "threat". In fact, even the Reagan administration itself was never completely clear as to whether re-flagging was designed to improve US relations with the Gulf States, stop the Soviets or give the United States a chance to get back at Iran. These potential logical contradictions actually become one focus of the discourse: Boynton observed that Senator Nancy Kassebaum would query the administration, "The Soviet Union is an ally of Iraq, correct? … And Kuwait is also an ally of Iraq". In this exchange, Kassebaum is trying to invalidate the pattern being constructed by the administration and replace it with the new pattern, "Kuwait is already an ally of the Soviet Union" (ergo: they are the Soviet's problem, not our problem).

The extent to which information is explicit rather than tacit is an aspect of culture. While North Americans and northern European tend to rely on explicit communications, many cultures rely on what Hall (1976) calls "high context" communication:

> A high context communication is one in which most of the information is already in the person, while very little is in the coded, explicit, transmitted part of the message. A low context communication is just the opposite; i.e. the mass of the information is vested in the explicit code.
>
> Japanese, Arab and Mediterranean people who have extensive information networks … are high context. As a result, for most normal transactions in daily life they do not require, nor do they expect, much in-depth background information. [They] keep themselves informed about everything having to do with the people who are important in their lives. Low-context people include the Americans and

the Germans, Swiss, Scandinavians, and other northern Europeans. (Hall and Hall 1987, 8)

The cultural norms of organizations likewise differ considerably. Northern European and most North American bureaucracies compartmentalize information, a sort of an information Fordism. This deals with the communications bandwidth constraint by reducing the information flow to a minimum and formalizing as much decision-making as possible into simple rules. Compartmentalization places few demands on individuals, who as a consequence can be easily trained and replaced. The drawback to the approach is that it is not very adaptive and it is brittle with respect to organizational behavior.

Hall's high context cultures provide an alternative to this approach by placing a great emphasis on socialization and shared values, as well as sharing information very widely. Japanese management is known for this style, though some US companies such as Xerox, Procter and Gamble, and Walmart also use it.[13] A high context management system extracts a very high upfront cost in terms of socialization but the high level of shared knowledge reduces the stress on the communications bandwidth. This results in an organization that can quickly transmit new information and therefore adapt to a changing environment, but which is brittle with respect to individuals because it takes a long time to socialize a new member.

As a consequence of these characteristics of organizational decision-making, the extent to which the written records of an organization reflect its actual decision-making will vary widely. The record is likely to be most complete when a low context organization confronts a novel situation: in that instance, almost all of the information will be explicit. It will be least complete when a high context organization deals with a routine situation: in that instance most of the problem-solving will be done using shared information and the written record will probably be both incomplete and very idiosyncratic.

One should not, however, despair if rule-based models "only" have outcome validity in some circumstances. The expert systems literature provides a useful illustration of this the contrast between human intelligence and machine intelligence with respect to outcome and process validity.

---

[13] Military organizations are an amalgam of the two styles. The stereotype of military information processing is that it is highly rule-oriented and low context. In some aspects military decision-making this is true, for at least two reasons. First, military basic training is designed to replace rational, serial processing with rapid, pattern-based reactions. This reduces reaction time, which in combat can mean the difference between life and death. In addition, combat losses can create a high turnover, and individuals need to be able to take over a job at a moments notice without having to learn to any details of the unit they will manage. These pressures for the rule-based, low context environment that is advantageous on the battlefield can lead to ludicrous results in peacetime, as with the late 19th century infatuation with Napoleonic drill in the age of the machine gun, or the 20th century rules on weapons acquisition.

At the same time, however, the social isolation of the military provides the basis for a high context culture within the officer corps; this has been further reinforced since the mid-19th century by an extensive system of military education. Ironically, this high context military culture often conflicts with the low context civilian bureaucratic infrastructure supporting it; the civilians insisting on "following the rules" while the military want to simply "get the job done", where "the job" is implicitly understood.

Based on about fifteen years of experience, it appears that expert systems can match or exceed human performance in two general areas: classification problems (e.g. diagnosis of disease, repair and configuration of equipment) and the information processing of large bureaucracies (e.g. insurance underwriting or credit approval).[14] Computers clearly do not solve classification problems in a human fashion. When a rule-based classification system is developed, its logical simplicity is usually of considerable (and sometimes demoralizing) surprise to the human experts who worked in the domain. The human expert solves classification problems using associative recall over a large, unsystematic but representative set of examples; the machine uses the logical processing of a relatively small number of rules. These machine intelligence systems duplicate the *performance* of the human expert without duplicating the *process*.

Expert systems that replace bureaucratic functions such as credit approval, in contrast, probably exhibit a great deal of process validity. A bureaucracy has a relatively low amount of associative memory due to high turnover, the expense of training, and physical information files that are generally accessed linearly rather than associatively. As *The Economist* notes

> the range of expertise that can be captured by an expert system is limited to simple, self-contained jobs which require no commonsense reasoning. But in part because the trend in management over the past century has been to break jobs into small, and supposedly more manageable, pieces, there turn out to be many things that expert systems can do without stretching (*Economist* 1992,12)

Low context bureaucratic decision-making reduces human information processing to the level of a machine and because of this artificially imposed procedural regularity, the process by which the expert system reasons is not dramatically different than that used by the individuals in the bureaucracy it replaced.

Learning in Rule-Based Models

While the information contained in most rule-based models of foreign policy is fixed, some of these systems exhibit simple learning. The numerical model in Alker and Christensen (1972) embodies some fairly sophisticated learning concepts based primarily in the psychological literature of the 1960s. Tanaka's CHINA_WATCHER (1984), which is conceptually based on Alker's earlier work on precedent, has the most explicit learning component of any of the models discussed in this chapter, continually assessing whether other nations in the system are friendly or unfriendly. In the PWORLD simulation (Schrodt 1988a), also based on the Alker work, simulated states copy from the successful strategies of other states in the system, though their repertoire of behaviors is very limited. EVIN (Taber and Rona 1995), an extension of the POLI model, will interpret new political information in light of the information it already knows.

A model relying *solely* on the precedent-based learning of behavior would require a large amount of background history. For example, if a system simulating the Cuban Missile Crisis was to "learn" the dangers of missile attacks and naval blockades before

---

[14] Computerized systems are also increasingly used in situations where statistical regularities can be exploited; often these are hybrids of rule-based systems and classical statistical techniques such as regression and discriminant analysis.

deciding on the appropriate policy response, it would need information going back to at least Korea and World War II, if not John Paul Jones. As discussed in Chapter 7, data sets providing information on precedents in the domain of militarized disputes are available (e.g. CASCON and SHERFACS), but to date no rule-based model has incorporated these.

Problem solving by the use of successful precedents is similar to the tactic of "chunking" that is central to the SOAR paradigm of Laird, Rosenbloom and Newell (1986; also see Waldrop 1988a, 1988b). Chunking is a generalization of the earlier Newell and Simon observation that, through experience, humans become more adept at solving problems because they remember the appropriate rules for solving a problem and re-apply those automatically, rather than working deductively through each discrete step of the problem.[15] While this concept is very basic, it can dramatically increase the problem-solving efficiency of expert systems: Waldrop (1988a) cites an example where chunking reduced the number of steps required to solve a problem from 1731 to 7. Organizational standard operating procedures are one form of chunked rules; JESSE explicitly employs "stored plans" in its knowledge base, though in the program's current implementation these are set *a priori* rather than learned

Machine learning could potentially provide a means around the knowledge acquisition problem in rule-based models. Model specification is very labor intensive, which limits the complexity of the model. More subtly, models based solely on human perceptions of the problem-solving process are likely to involve a great deal of wasted effort: Most research in machine learning has shown that human experts rely on great deal of redundant information because multiple chains of reasoning provide a check against missing or erroneous data. Because machines have a far greater capacity to cope with individual items of information, but are very weak in their associative abilities, the means by which they compensate for noise are logically quite different from the methods used by humans.

A system with natural language abilities might also be able to derive some rules from the communications within the bureaucracy itself, a process similar to that involved with the human coding of cognitive maps. In democratic systems, foreign policy decisions are usually explained and justified, often in great detail; sources such as the *Congressional Record* are now available in machine-readable form, and it is relatively straightforward to search a text for statements embodying the declarative or procedural knowledge required by a rule-based model. Explanations in the public record are rarely wholly truthful—for example few analysts believed George Bush's assertion that the US response to Iraq's invasion of Kuwait had nothing to do with the control of oil, and even fewer believed it after Bush's failure to respond to Serbia's attack on Croatia and Bosnia a year later—but rarely are the statements wholly fallacious, and in any case expert systems are excellent at detecting logical inconsistencies. As noted above, it is unlikely that one could ascertain the rules of an organization solely from documents but machine-*assisted* knowledge acquisition from documentary sources might enable the efficient

---

[15] For example, in the days before Velcro™, when a child was first learning to tie his or her shoes, the process was exceedingly slow and awkward. With practice, however, it became completely automatic and could be done without conscious thought. The process remained complex, but its solution had been chunked into the single operation "Tie your shoes" rather than the series of operations "First put one string around the others, then make a bow…".

development of models that are more complex, more logically efficient and somewhat more objective than models developed solely through human coding.

Conclusion

Bureaucracies, by definition, are rule-following social organizations, and to the extent that a bureaucracy actually follows a set of known implicit and explicit rules, a computational model based on those rules will have a great deal of validity. Rule based systems provide a much richer modeling environment than rational models, which are typically limited to a very small number of options and formal probabilistic reasoning, or numerical models, which are very awkward when dealing with qualitative behavior.

At the present time, the major problem inhibiting the development of additional rule-based models is the sheer complexity of the task; this is particularly true of the multiple-agent models. An alternative research strategy might be to work on models in more regularized and clearly defined domains, rather than focusing on difficult domains such as militarized disputes; such models would at least provide information on what "normal" foreign policy behavior looks like, and the crisis behavior could be compared against this. Rule-based systems are likely to be most successful in highly regularized and repetitive decisions, though the high level of outcome validity shown in existing models of crisis decision-making may indicate that even crisis behavior is more regular than one would intuitively expect.

It may also be possible to do more than we expect using simple models: there is nothing inherently inferior about a rule-based model involving tens or hundreds of rules compared to a model containing thousands of rules, provided the simple model mirrors behavior. Because of the rule-based nature of bureaucratic behavior, even the simple model is likely to provide far greater detail and predictive accuracy than a regression or game theoretic model, because the knowledge structures and information processing of a rule-based model are far more congruent with bureaucratic processing.

Chapter 5
Machine Learning

Imagine that we are living on an intricately patterned carpet. It may or may not extend to infinity in all directions. Some parts of the pattern appear to be random; other parts are rigidly geometrical. A portion of the carpet may seem totally irregular, but when the same portion is viewed in a larger context, it becomes part of a subtle symmetry.

The task of describing the pattern is made difficult by the fact that the carpet is protected by a thick plastic sheet with a translucence that varies from place to place. In certain places we can see through the sheet and perceive the pattern; in others the sheet is opaque. Light passing through the sheet is often refracted in bizarre ways, so that as more of the sheet is removed the pattern is radically transformed. No one knows how thick the plastic sheet is. At no place has anyone scraped deep enough to reach the carpet's surface, if there is one

Martin Gardner

In the rule-based systems discussed in Chapter 4, and in most of the early expert systems work in AI, the information required to model a process was obtained "by hand". A programmer or "knowledge engineer" first designed a general system appropriate to a particular domain (e.g. medical diagnosis or credit approval), and then worked with domain experts such as doctors or bank officers to determine the rules and other information required. This approach to knowledge acquisition produced functioning systems—no small feat in an era when the abilities of computers to perform at human expert levels was very much in doubt—and had face validity since the program's knowledge was based on a trusted human expert.[1]

As expert systems proliferated, however, it became apparent relying on human expertise had several disadvantages. It was expensive: Extracting information from (well-paid) experts was often time consuming and knowledge engineers (also well paid) were in short supply. In addition, experts were often unable to unambiguously articulate problem-solving rules appropriate for machine implementation because the experts relied on their subcognitive and associative pattern recognition rather than using deductive logic and probability theory. Finally, even a successful system contained only the knowledge of experts who contributed to it, which precluded the possibility of developing systems that performed better than the experts. These obstacles became known as the "knowledge acquisition bottleneck" and were particularly acute in complex domains where case descriptions involved hundreds or thousands of features.

The problems with the knowledge engineering approach revived interest in machine learning methods, which had been developing in parallel with expert systems

[1] This is sometimes called the "<Name>-in-a-box" approach: Charlene is extremely good at evaluating mortgage applications, but Charlene is about to retire. So bring in the knowledge engineer and capture Charlene's expertise in a computer program—which conveniently will not require cost-of-living increases, a corner office or vacation leave—and create "Charlene-in-a-box".

throughout the 1960s and 1970s.[2] Machine learning requires the *a priori* specification of a knowledge representation system and the identification of the variables required to solve the problem, but the program itself figures out how to use the information by learning from examples. The program is given a large set of cases whose solution is known—for example medical case histories containing information on both the patient's symptoms and the diagnosis of their disease, or credit applications with information on whether the loan was repaid. From these "training cases" the program induces a method of solving the general problem; that solution is then validated on additional cases.

While all rule-based systems are based on elaborations of *if...then* logic, the knowledge representation structures and inference techniques used by various machine-learning systems differ dramatically. This chapter will consider four of the most common methods: two of these use structures somewhat like rules (ID3 and genetic algorithms); the remaining two involve arithmetic calculations. The class of objects that an algorithm can discriminate is dependent on the examples on which it has been trained, the consistency of the features that account for discrimination in the training and testing cases, and on the ability of the knowledge structure to capture the type of discrimination that is empirically present. For example, if the classes are actually determined by a multivariate linear model, a machine learning algorithm seeking logical rules to describe them will have a difficult time discriminating between then; the converse is also true.

The distinction between machine learning methods and conventional inferential statistical techniques can also be quite vague: for example the texts of Weiss and Kulikowski (1990) and Schalkoff (1992) cover both statistical and machine learning approaches to classification problems.[3] In a sense, parametric statistics such as regression are a simple form of machine learning. In a regression equation the "knowledge structure" is the linear equation and its coefficients; the least-squares algorithm "learns" the values of the regression coefficients from the values of the dependent and independent variables in a set of cases. The knowledge structures used in machine learning are usually much more complicated than those used in statistics, and machine learning is not based on probability theory.

The choice of the cases used to train the system is critical. Training cases must be representative and must include both examples and counter-examples. Most machine-learning systems trained on a set of cases consisting solely of black crows would not conclude that "All crows are black" but instead that "All objects are black" or "All objects are crows". To impart the knowledge that all crows are black, some chickens are

---

[2] See for example Forsyth and Rada (1986) and Michalski, Carbonell and Mitchell (1983). For the purposes of this discussion, I'm using the term "expert system" to refer to any system where the knowledge is provided by a human rather than induced by the machine; this is a more general use of the term than is normally found in the literature.

The shift from knowledge engineering to machine learning was dramatically reflected in the AI sections in bookstores. When I taught an AI workshop at the ICPSR in 1987, roughly half of the AI books at Border's Books in Ann Arbor dealt with expert systems (the remainder were mostly introductions to AI and LISP-oriented works); by 1994 half of Border's much-diminished AI section was devoted to machine learning, mostly neural networks.

[3] Systems used in applied settings also combine different machine-learning methods: *Byte* (July 1993, 107-115) discussed several of these.

black and some chickens are white, the training cases must include black crows, black chickens and white chickens.

Machine learning systems are not purely inductive: human expertise is required to define the problem, identify the appropriate features, and develop a representative set of training and validation cases. Judicious human intervention can also be used to fine-tune an otherwise awkward knowledge structure generated by a machine.

However, once an appropriate set of variables has been identified, machines are often superior to human experts in generating systems that completely describe the regularities in the set of training cases. Human experts tend to forget about small exceptional subsets in the domain until reminded of them by failures in the validation set; consequently the development of the system is iterative and slow. Due to the differences in human and machine memory, machine-learning systems usually make more efficient use of the information available in the data . Because human associative perception has a high internal bandwidth—we can deal with a very large number features in a case so long as we are only doing pattern recognition—there is little cost to considering redundant information, and redundant features may in fact be useful when compensating for noise. With a computer, in contrast, adding features increases the processing required, and because the machine has no intrinsic error-correction methods[4], there is no advantage to retaining redundant information. As a consequence the classification systems produced by machine learning systems often require fewer variables to achieve a given level of accuracy than a human expert would use.

Finally, machine learning does not require a human expert to "think like a machine", artificially translating his or her expertise into a novel form such as rules. In addition, the knowledge structures of the machine are not restricted to a level of complexity that humans can logically cope with. For example, in Chapter 6, I discuss a sequence comparison technique whose knowledge structure is a matrix containing about 30,000 cells. The values in a structure of this size are far too numerous for a human to deal with consciously—even though the matrix is simpler than the structures used subcognitively by the brain—but the machine can determine those values using a set of examples.

This chapter covers four machine learning methods that have been widely used in machine learning: ID3, genetic algorithms, neural networks and nearest neighbor (clustering) methods. Each method is illustrated with an international relations application; the neural network and ID3 methods are compared with each other and with some statistical techniques. While much of the chapter involves the technical discussion of the algorithms and their properties, I also make a number of general observations about parallels between aspects of machine learning, organizational learning, and knowledge structures that might be relevant to the development of foreign policy.

Each of these techniques was implemented in Pascal on a personal computer. Contemporary machines are at least an order of magnitude faster than those I was working with, and it would now be possible to use these methods on problems of considerably greater complexity than those demonstrated here. Commercial software is available implementing each of the methods, but the core algorithms for ID3, genetic

---

[4]Neural networks do this as part of their classification process but not in their inputs.

algorithms and neural networks are quite simple—a couple hundred lines of Pascal or C—so writing custom programs is not overly challenging.

ID3[5]

As discussed in Chapter 4, most rule-based systems encountered in international relations use forward-chaining *if...then* rules. To date, almost all of these rules have been coded by hand, but such systems are actually very straightforward to construct by machine. This section discusses a bootstrapped version of the most common inductive learning algorithm, ID3, which is used in most commercially-available inductive expert systems.[6] ID3 is a very robust technique—in a sense it is the linear regression of machine learning—and it is often used as the basis of more elaborate methods.[7]

The core of the ID3 algorithm is the tree-building CLS ("Concept Learning System") technique developed by Hunt (Hunt, Marin and Stone 1966). Hunt's CLS created a classification tree by successively determining which variable out of a given set should be used to make the next node of the tree. Quinlan (1979, 1983) made a key modification to CLS by employing the information-theoretic concept of *entropy* to determine the choice of variable; his overall system was called ID3. Quinlan was working on the problem of chess endgames and the full ID3 algorithm involved a scheme for iteratively generating rule sets out of very large sets of data by looking at subsets of data and then determining exceptions in the full set (see Cohen and Feigenbaum 1982,406). In most contemporary applications, the number of cases is sufficiently small that the entropy-based CLS is computationally efficient, so only the CLS part of the algorithm is used. ID3 can be modified to work with interval data but the method used here assumes the data are nominal.[8]

The basic mechanism of ID3 is simple. The data are a set of cases, each containing a number of nominal features and a single nominal dependent variable to be predicted. ID3 starts building the classification tree by choosing the independent variable that minimizes the total entropy of the dependent variable after the cases have been split on that variable. Entropy is defined using the standard information theory definition (Pierce 1980)

$$H = \sum_{i} p_i \log_2(p_i)$$

where p_i is the proportion of dependent variable values in category i.

---

[5] Parts of this section appeared earlier in Schrodt (1991c).

[6] Thompson and Thompson (1986) provide a very readable description of the basic ID3 technique; Garson (1987) describes ID3 in the context of a political science modeling problem.

[7] See for example the I^2D method developed by Unseld and Mallery and applied to the analysis of the SHERFACS data set (Unseld and Mallery 1993; Mallery and Sherman 1993).

[8] According to Garson (1991,402), "For interval data, the ID3 method treats the mean of each pair of adjacent ordered values as a potential threshold (cutting point), then uses [ID3] to select the threshold point expected to contribute the most information on the dependent variable."

ID3 successively partitions the data set based on the values of features in the set of cases. For any subset of the data C, the entropy given a specific value of a dependent variable a_j is computed as

$$H(C|a_j) = -\sum_i p(c_i|a_j) \log_2 (p(c_i|a_j))$$

where

$p(c_i|a_j)$ = proportion of the cases in the subset C where the dependent variable has the value i and the feature A has the value a_j

The total entropy of the subsets generated by a feature A is the sum of the entropies of each category of that feature times their probabilities:

$$H(C|A) = \sum_j p(a_j) H(C|a_j)$$

Because ID3 works by successive partition, the classification tree can be constructed recursively. In pseudocode, the ID3 algorithm is:

```
Procedure Split(C,V)
{Find the optimal partition of the data set C based on the features
in V}
begin
Find A in V that minimizes H(C|A)
for each value aj in A do
 if H(C|aj)>0
  then
      Define the subset C*={C|A=aj}
      Define  V* = V-A              {remove A from V}
      Split(C*,V*)             {partition the subset}
  else
    Finished       {all cases in C have the same value}
end
```

In words, ID3 starts by choosing the feature that maximizes the probability of correctly classifying the cases. It then splits the cases into M subgroups based on their value on that feature. For each of those subgroups, the *Split* procedure is repeated to find which of the remaining features provides the best classification within the subgroup. This choice can be—and usually is—dependent on the variable choices already made. The chosen feature is used to generate sub-sub-groups, and the *Split* procedure is repeated on those groups. The procedure is repeated until all of the cases in a subgroup have the same value for the dependent variable, or until one runs out of features.

Because ID3 uses different features depending on the subset being classified (equivalently, depending on where it is in the tree) a classification tree is quite different from a contingency table. In a contingency table, the nesting of variables is identical for all cases, whereas in a classification tree, a feature necessary to classify one set of cases may be irrelevant for others. For example, in building a classification tree that identified birds and mammals, the branches involving "birds" would presumably ask questions about beaks and feathers; those questions would not be asked about mammals.

On the surface the asymmetry of the tree leads to a more complex knowledge representation structure than those traditionally used in social science research. In terms of the information demands, however, this approach is more robust against extraneous data, and makes more efficient use of the available information. If a feature contributes nothing to the classification of a subset, it can be ignored and have no effect on the classification within that subset. This is particularly useful in historical and cross-national data sets where there is likely to be a great deal of missing data. In addition, if two features carry redundant information, one of them will simply be ignored in ID3, instead of creating the collinearity problems one would find in a comparable situation in a linear model.

A rather quirky but useful aspect of ID3 is that a missing value can be treated the same as any other value and under some circumstances the missing value will contain information. For example, as any user of survey research information knows, people who are never home, who terminate interviews early or who refuse to answer certain questions are not randomly distributed: their missing answers may reveal something about the case. Infant mortality statistics are missing for rural areas in developing countries because those areas have inadequate education, communication and medical personnel, but those characteristics are also usually associated with high infant mortality. The presence of a missing value can sometimes be used to infer additional characteristics of a case.

In a data set with a large number of features, ID3 will usually classify 100% of the cases correctly. The only situation where this fails is when two cases are identical on all features and have different values for the dependent variable. This indicates that the set of variables being considered is not sufficient to correctly classify all of the cases in the sample. While 100% accuracy sounds impressive, its practical value is limited because many roots of the tree terminate in single cases. To ascertain the predictive value of the variables, one must use a split-sample test, building the tree on part of the data (training cases) and then testing it on the remaining part (test cases).

This study uses "bootstrapping", a method used in statistics to empirically ascertain the stochastic structure of a non-repeatable set of data (see Mooney and Duval 1993). Instead of doing a single training and validation test, I analyze a large number (50 to 200) of random split samples. The basic bootstrap involves:

```
for i:=1 to Number_of_Experiments
   begin
     Randomly select half of the cases as a training set
     Compute a tree describing that sample using ID3
     Classify the remaining cases using the tree
     Compute statistics on the accuracy of the classification
   end
  Compute summary statistics
Construct trees using only the most frequently used variables
```

This method in effect does a series of "what-if" experiments to determine how accurately ID3 would predict the unknown cases if only half of the cases were known.

Bootstrapping can also be used to determine which variables are most commonly used in building the trees, then a simplified tree can be constructed using only this subset. This has two potential advantages. First, it provides parsimony with respect to the information required to classify a case, eliminating redundant features. Identifying the subset of features with the highest level of information is analogous to the use of stepwise regression (in the absence of collinearity) in statistical modeling. Second, there is a potential trade-off between accuracy and generality. Because ID3 usually classifies cases with 100% accuracy, it has no generality; it uses every bit of information required to classify the case. The resulting tree may over fit the training set and then fail on many cases in validation set because of its specificity. By forcing ID3 to use a smaller number of features, rules can be kept more general to avoid over-fitting.

Application

Data in this example are from the Butterworth "Interstate Security Conflicts, 1945-1974" data set (Butterworth 1976; ICPSR 7586). This is described in Appendix A: the set covers 310 cases of interstate conflict and codes 47 variables dealing with conflict characteristics, actions to manage the conflict and the outcome of that management.[9]

Two measures of fit were computed. The first was simple accuracy

$$\text{Accuracy} = \frac{\text{correct predictions}}{\text{total cases}}$$

The problem with the accuracy measure—in Butterworth and most other international relations data—is that the data are heavily modal. International affairs are boring: the same thing happens most of the time. Because the modal category accounts for about 55% to 75% of the cases, the simple rule "predict the mode" has an accuracy of 55% to 75% and is a tough horse to beat.

The mode, however, is an unsatisfactory null model because it has little practical utility. A model should be able to predict non-modal cases as well as modal cases. In particular, foreign policy analysts can be excused—up to a point—for inaccurately predicting cases where a crisis does not in fact escalate provided they manage to correctly predict the cases that do escalate. In other words, there is more value to predicting rare events than common events.

There are a variety of measures of rarity but since ID3 uses an entropy framework, the obvious weight is the entropy measure $\ln(p_i)$. The second measure used in this analysis measures "entropy explained": Let

---

[9]Variables dealing with characteristics and management actions were treated as independent variables. Variable 23 ("Specific agent") was not included in the analysis: this contains 45 categories and in many cases there is only one conflict per category due to the specificity of the agent (e.g. "Pope Paul VI","US/Argentina/ Brazil/Chile"). In ID3, this variable would have a large amount of discriminatory power since it predicts many single cases; equally obvious is that it would have no generalizability. The general character of the management agent (i.e. states, type of international organization, individuals) is coded in variable 22, "Management Agent". Since this test deals only with nominal variables, "Number of Agents" is not included.

$$\text{ME} = \sum_i \left(\frac{\text{correct predictions}_i}{\text{cases}}\right) \ln(p_i) \qquad \text{(Model entropy)}$$

$$\text{DE} = \sum_i p_i \ln(p_i) \qquad \text{(Dependent variable entropy)}$$

then

$$\text{Entropy Ratio} = \text{ER} = \frac{\text{ME}}{\text{DE}}$$

The index i is over the values of the dependent variable; p_i in both equations is the proportion of each value in the observed cases of the dependent variable. ER will vary between 0.0 (no correct predictions) and 1.0 (perfect prediction). ER can be computed for the modal prediction as well as the ID3 predictions and the two compared. Because the natural logarithm of a number is equal to a constant times the logarithm base 2, there is no loss of generality in using ln(p) rather than $\log_2(p)$.

Results

The results reported here are based on two series of computer runs that occurred in two stages. First, two sets of 200 Monte-Carlo split-sample tests were run and aggregate statistics were recorded on how frequently each independent variable was used in the classification tree and on the fit of the trees using the validation set. Based on these results, a series of 50 split-samples were done using the N most highly ranked variables where $3 \leq N \leq 10$. This provides an indication of the accuracy of trees using a small number of variables.

This design was done with two different sets of features. The first set included all 38 of the features; second deleted the four “Technique of Management” features. The second set was considered because the “Technique of Management” features had a comparatively large number of categories and were therefore used quite frequently. The specificity of these features also made them suspect as predictors, since the management technique is presumably closely related to the outcome expected by the human conflict managers. Most of the remaining features, in contrast, refer to objective characteristics of the conflict situation.

Classification Trees

Classification trees were initially constructed using the complete set of features. Unsurprisingly, given the large number of features, these trees correctly classified 100% of the cases. The trees were fairly parsimonious, only rarely going deeper than five levels, though this parsimony was substantially enhanced when the “Technique of Management” variable was used to initially partition the set.

A simplified (as these things go…) classification tree for the *Stopping Hostilities* variable is presented in Table 5.1. This tree uses only the best five features identified from the bootstrapped ID3 on the 34-feature set: leadership, type of issue, fatalities, action and strongest antagonist. The 89.8% of the 98 cases are classified correctly; erroneously classified cases are in italics.[10]

At a number of points in the tree, only some of the values of the feature are used as branches; this occurs when the other values did not occur in the subsets examined at that node. If the system encounters an independent variable value that is not present at a node, it creates a new "unrecognized value" node that predicts the modal value of the dependent variable in the subset of cases being considered at that point.[11] This doesn't happen in the tree displayed here because the entire set of cases was used for training but it occurs fairly often (10%-20% of the cases) in the Monte Carlo process.

In examining the tree in Table 5.1 and similar trees for the remaining four dependent variables (see Schrodt 1987b), several things are apparent. First, despite the fact that only five of the original 40 features in the Butterworth set are used, the classification accuracy is very high: between 95% and 100% as measured by either simple accuracy or entropy-explained (ER). The accuracy and ER measures have roughly the same values, and the tree does a better job of classification than the single rule "predict the mode".

Second, the trees are fairly complex, though not so complex as to always terminate in single cases. The *Stopping Hostilities* tree in Table 5.1 uses 26 roots to classify 98 cases with 90% accuracy and in general the number of roots in the ID3 trees were proportional to the number of cases classified. The tree on which Table 5.1 is based originally used 69 branches to classify those cases with 99% accuracy, so the number of roots of the tree can be reduced by 62% while giving up only 9% in accuracy.

---

[10] Classification trees for the remaining variables are provided in Schrodt (1987b). Table 5.1 is based on the tree presented in Schrodt (1991c) but in the interests of space, it has been substantially simplified by combining similar branches. The order in which the classification features are used has not been changed.

[11] The classification program processed these situations automatically and the nodes are not shown in the tree.

Table 5.1. Classification tree for “ Stopping Hostilities” using five features

```
Leadership
       2 superpowers | 10 15 304  <1>
       Middle powers | 92 178 260 273  <0>
        1 superpower | Type of Issue
                           Cold War | 67 83 89 95 263 274 <0>  21 [1]
                    Internal, gener | 158 250 291  <0>
                           Colonial | Fatalities
                                            26-100 | 69  <0>
                                              >101 | 16 167 <2> 63 [0]
                    Interstate,gene | Fatalities
                                              1-25 |Strongest Antagon
                                                   Smallest Power|99<2>
                                                     Small Power |25<0>
                                            26-100 | 125  <2>
                                             Other | 53 52 47 <1>
        Large powers | Strongest Antagonist
                       Middle Power | 264  <1>
                              Other | 100 122 228 162 <0>
        Small powers | 108 229 4 62 96 216 97 <0> 282 [2]
     Smallest powers | Fatalities
                             0 - 25 |  192 221 258  <0>
                             26-100 | Strongest Antagonist
                                      Middle Power |  161  <0>
                                        Other      |  110 259 <1>
                               >101 | 202 223 257 269 143 213 253 86
                                      103 227
                                    | 43 65 68 154 179 239 217 265
                                      310 <0>
                                    | 114 128 196 [1]
         Sec-general | 159 194 277 302 308 <0>  280 [1]
        Inapplicable | Fatalities
                              0-100 |  292 57 <0>
                             101-1K | Action
                                      Coercive ops | 175 231  <2>
                                             Other | 30 132 190 244
                                                     281 <1>
                                                   | 75 243 [2]
                              1K-2K |  119 130  <1>
                             2K-10K | Strongest Antagonist
                                      Middle Power |   29  <1>
                                        Superpower |  139  <0>
                               >10K | Strongest Antagonist
                                       Large Power | 245 11 <1>
                                             Other | 118 211 218 285
                                                    207 301 <0>
                                              | 298 [1]
```

Key: the features are in **bold face,** the values are in plain text followed by |. Each level of indentation represents another level of the tree; at the root of the tree the set of numbers identifies the case numbers (Butterworth 1976) that were classified at that root, and the number in **‘< >’** is the value of the dependent variable for those cases. Incorrectly classified cases are indicated by italics.

Third, most of the cases are classified without going very deeply into the tree. Usually cases begin to be classified at the third level of the tree; in other words, knowing the values of only three variables allows some correct classification. While the *global* structure of the tree is relatively complex, the local structure required to classify a particular case is simple, and the individual rules are certainly well within the information processing capabilities of humans.

Rank-Order Results

Tables 5.2 and 5.3 give the aggregate results for the *Stopping Hostility* variable for both the 38- and 34-feature sets. Several statistics are presented:

- *Modal prediction Accuracy and ER***:** The values of these statistics that would be obtained following the rule "predict the mode".
- *"Accuracy", "Average ER"*: The average values for each statistic across 200 split-sample tests
- *Level where used*: This shows the frequency with which each feature was selected for use in the trees. Level 1 is the top of the tree, the first feature used to partition the cases. Levels below 5 are not recorded: use of features at the lower levels are reflected in a difference between the total of the levels and the Total frequency.
- *Total*: The total number of times a feature was used in the tree. If a feature is used more than once in the same tree, it is counted multiple times. The rank-order of the feature in the table is based on this value.

Examining Tables 5.2 and 5.3 and comparable tables for the remaining dependent variables in Schrodt (1987b) revealed several general characteristics of the trees generated in the bootstrap samples. First, the trees rarely go very deep: level 5 is close to the limit. Despite the availability of more than 30 features, five or fewer features seem to almost always suffice for classification. This may be due in part to the sample size since the split-samples have only fifty to one-hundred cases and $\log_2 64 = 6$.

Second, there are definite patterns in the choice of features used. The frequencies of the use of a feature generally seem to follow a rank-size law [12] and there is a great deal of consistency in the choice of features used to classify the different dependent variables. Features are clearly not being chosen at random, though all of the features are used at least once. The rank-orders produced by the two independent Monte-Carlo runs on each of the data sets were also quite similar, particularly for the most commonly used features. As expected, the "Technique of Management" features were used frequently when available, but these had little effect on the relative rank of the remaining features except for "Action", which apparently carries much the same information as the "Technique" features.

---

[12] The frequencies are proportional to the rank-order of the variable; the second-most-common variable has roughly half the frequency of the most common, the third-most-common has one-third the frequency and so forth. This phenomenon, known generically as Zipf's Law, is found in an assortment of empirical distributions such as word counts in manuscripts and the populations of cities.

Table 5.2. Summary of B/ID3 experiments for *Stopping Hostilities* with All Management Action features included

```
Modal prediction ER = 0.329
Modal prediction Accuracy = 0.653

Accuracy = 0.583
Average ER= 0.407
```

Level where Used					**Total**	**Feature**
1	**2**	**3**	**4**	**5**		
66	97	16	1	0	180	Tech of Management Action(1st)
21	92	40	1	0	154	Leadership
69	60	22	0	0	151	Tech of Management Action(2nd)
0	90	46	12	0	148	Fatalities
0	67	26	1	0	94	Strongest Antagonist
0	47	38	4	0	89	Likelihood Disappear.
0	45	31	5	1	82	Likelihood Abatement
0	38	39	3	0	80	Duration
38	23	3	0	0	64	Tech of Management Action(3rd)
0	41	13	1	0	55	Type of Warfare
0	31	20	0	0	51	Type of Issue
0	32	16	3	0	51	System Period
6	26	12	2	0	46	Action
0	23	14	4	0	41	Level of Agreement
0	23	14	2	0	39	Alignment of Parties
0	23	16	0	0	39	Other Managers
0	22	13	2	0	37	Agent's Autonomy
0	20	15	1	0	36	Strategic Category
0	20	9	3	0	32	Likely Degree of Spread
0	7	19	3	0	29	Initiative for Intervention
0	5	20	4	0	29	Agents Relative Power
0	8	15	5	0	28	Agents Previous Role
0	11	14	0	0	25	Agent's Bias
0	11	10	1	0	22	Degree of Spread
0	15	5	1	0	21	Previous Involvement
0	15	4	1	0	20	Phase of Agent's Intervtn
0	9	9	2	0	20	Management Agent
0	9	9	2	0	20	Past Relationship
0	9	8	1	0	18	Ethnic Conflict
0	12	3	0	0	15	Phase of Agent's First Action
0	11	2	1	0	14	Likhd Superpower War
0	6	5	1	0	12	Joint Leadership
0	7	3	2	0	12	Great Power Interests
0	1	8	2	0	11	Phase of Agent's Strongest Act
0	6	3	1	0	10	Ideological Conflict
0	2	5	1	0	8	Agents Primary Role
0	6	1	0	0	7	Tech of Management Action(4th)
0	5	1	0	0	6	Power Disparity

Table 5.3. Summary of B/ID3 experiments for *Stopping Hostilities* with no Management Action features

```
Modal prediction ER = 0.329
Modal prediction Accuracy = 0.653

Accuracy = 0.581
Average ER = 0.430
```

Level where Used 1	2	3	4	5	Total	Feature
19	209	67	0	1	296	Fatalities
115	44	43	5	0	207	Leadership
54	51	33	3	0	141	Action
6	79	25	0	0	110	Strongest Antagonist
1	78	24	1	0	104	Type of Issue
1	48	34	5	0	88	Likelihood Disappear.
0	52	33	2	1	88	Alignment of Parties
0	44	35	2	0	81	Likely Degree of Spread
0	34	43	2	0	79	Duration
0	24	47	5	0	76	Likelihood Abatement
0	24	48	3	0	75	Level of Agreement
0	13	33	1	0	47	Degree of Spread
0	15	30	0	0	45	Agent's Bias
0	17	22	4	0	43	Other Managers
0	14	27	1	0	42	Strategic Category
0	6	34	1	0	41	Previous Involvement
0	14	21	2	0	37	System Period
2	7	22	1	0	32	Type of Warfare
0	8	21	2	0	31	Agent's Autonomy
0	12	12	2	0	26	Ethnic Conflict
0	6	16	0	0	22	Power Disparity
0	2	19	1	0	22	Agents Relative Power
0	9	10	1	0	20	Past Relationship
1	7	12	0	0	20	Management Agent
0	6	12	0	0	18	Agents Previous Role
0	7	9	0	0	16	Initiative for Intervention
0	10	4	0	0	14	Phase of Agent's First Action
1	1	12	0	0	14	Phase of Agent's Intervtn
0	3	9	1	0	13	Ideological Conflict
0	3	7	1	0	11	Phase of Agent's Strongest Act
0	4	6	0	0	10	Joint Leadership
0	5	5	0	0	10	Great Power Interests
0	4	5	0	0	9	Agents Primary Role
0	4	1	0	0	5	Likhd Superpower War

Distribution of Accuracy and ER

Both the accuracy and ER measures are roughly normally distributed; Figure 5.1 shows the distributions of ER. In none of the cases examined did either accuracy or ER exhibit any noticeable departure from a bell-shaped curve. While I did not compute standard deviations on these distributions, they appear to be about 0.07 for Accuracy and 0.1 for ER. Because the distributions are essentially normal, the discussion of these summary statistics will be done with respect to their means, reported in Table 5.4.

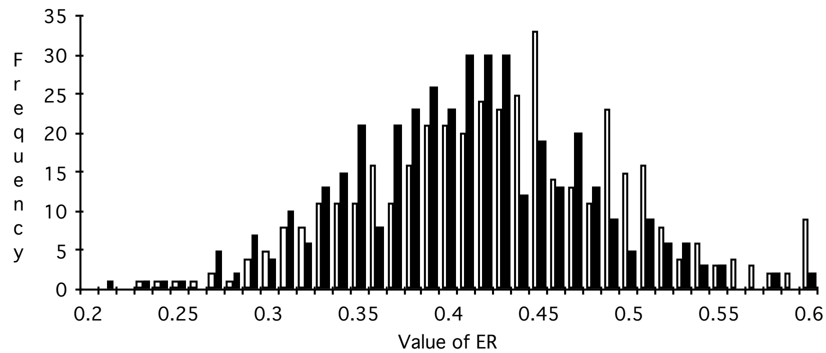


Figure 5.1. Stopping Hostilities

As expected, the mean accuracy of the bootstrapped models was well below that of the modal prediction. If simple predictive accuracy is the only criterion, just use the mode. In contrast, the ID3 trees consistently did about 20% to 30% better than the mode on the ER measure. The 34-feature models have a slight advantage over the 38-feature models on the ER measure: while numerically this is small it is probably statistically significant given the size of the samples. While the accuracy and ER statistics are not overly impressive, note that these measures are computed on *split-sample* tests—the training set contained none of the cases from the validation sets. Since the classification trees are generally 100% accurate, the accuracy and ER on the *total* sample would be around 0.7 to 0.8 for the total set when only the training set was used to construct the tree.[13]

Table 5.4. Fit of the Model

Variable	**38 Model***		**34 Model***		**Mode**	
	Acc	**ER**	**Acc**	**ER**	**Acc**	**ER**
Stopping Hostilities	.583	.407	.581	.430	.653	.329
Abating the Conflict	.503	.401	.495	.407	.568	.342
Isolating the Conflict	.668	.397	.663	.414	.745	.303
Restraining the Conflict	.484	.407	.484	.416	.545	.344
Settling the Conflict	.619	.370	.603	.372	.734	.300

*Statistics for ID3 are mean values over 200 bootstrap experiments

[13] 0.50 + 0.5*ER (or Accuracy).

Number of Features

Some additional experiments were done by varying the number of features allowed in the classification tree to see how much information was gained or lost as a function of the depth of the tree. When features were added according to their rank order in the initial experiments, the number of features used in the tree can be reduced to about 4 or 5 without much loss in either ER or Accuracy. In some cases trees based on the reduced set of features classified the validation cases slightly more accurately than trees based on the full set of features.

This has two implications. First, the features coded in the Butterworth data are quite redundant with respect to predicting the dependent variables studied here—a small subset have much the same classification information as found in the full set. However, not just *any* subset will work: In separate set of experiments I progressively deleted the most frequently used features, and as the features that were the best classifiers were eliminated, the average values of Accuracy and ER declined. Second, based on these experiments, the bootstrapped ID3 can be used as a technique to identify the sub set of features with the most general classification power.

Discussion

The first conclusion is simply that the bootstrapped ID3 technique works for the purposes of both classification and identifying a core set of features. Applied to a set of data with between 100 and 200 cases and 38 features, the method is able to identify a small set of features that achieve between 95% and 100% accuracy when used to construct a classification tree for the entire data set, and can do better than the mode in an entropy-based measure in split-sample tests. The technique appears to be quite robust and did not exhibit any unusual behaviors in any of the various tests. While the technique is computationally intensive, most of this comes in doing the bootstrapped tests; the ID3 algorithm itself is reasonably efficient.

A more sophisticated application of these same methods is found in Fürnkranz et al (1994), which applies an ID3 method (Quinlan's C4.5 algorithm) to the analysis of Bercovitch's international mediation data set (Bercovitch and Langley 1993), which is structurally similar, though larger and more detailed, than the Butterworth data.

The general results of the Fürnkranz et al work are similar to those found here: a full-sample analysis of the 718 cases in the data, using 52 attributes, results in nearly perfect (99.7%) accuracy, but this drops to only 40% accuracy in cross-validation testing with split samples.[14] The initial tree, however, has some 547 nodes, and most of these terminate in single cases.

Using C4.5, the researchers go on to apply several tree-pruning methods with considerable success. Two methods of pruning are used. First, complex rules are collapsed into simple rules even if this results in a small number of erroneous predictions. Second, a minimum cluster size limit is imposed to eliminate the single-case branches.

---

[14] The dependent variable is the success or failure of the mediation attempt; unlike the Butterworth example, the cases are fairly evenly split between the two values 57% failures and 43% success. The cross-validation was done by dividing the sample into ten subsets, fitting on the data in nine of those, and then testing on the remaining 10th; the out-of-sample accuracy is the average accuracy across these ten tests.

These experiments typically result in a classification tree with about 20 modes, and *increase* the cross-validation accuracy to 45%.

Decreasing the number of variables also is effective at improving the accuracy, and this method has more dramatic results in the Bercovitch data than in the Butterworth data. Limiting the depth of the tree the four variables found most salient by the procedure results in only 76% percent accuracy on the training data, but in cross-validation tests the accuracy remains at 68%.

Given the effectiveness of simple machine-induction methods such as ID3 there seems to be little point in constructing rule-based expert systems classification trees solely by hand. Human expertise is still necessary to identify a set of relevant features, but if a classification tree is the desired knowledge representation structure, then ID3 will probably produce a far better global structure than a human expert can devise. Machine induction is superior to human induction in constructing classification rules just as statistical regression is superior to eyeballing when constructing a least-squares line.

From the standpoint of human cognition about international affairs, these experiments may partially explain why very different models of international conflict seem to have similar levels of empirical accuracy. When measured in split-sample tests, in regularity in the conflict data isn't very great, so any model making predictions outside the set of cases used in its construction will contain considerable error. Models predicting future behavior on the basis of behavior observed in the past have this characteristic.

While human analysts do not look at *all* historical data in the systematic fashion of ID3, a sophisticated analyst may go through a mental procedure not unlike a bootstrap in examining a number of cases—usually the "classical" cases identified by an organization or a paradigm rather than a random sample—and sequentially adjust his or her mental classification model to correctly predict those cases. As these ID3 experiments indicate, the redundancy in the data would allow two analysts or organizations to construct very different, but perfectly-fitting models from their *selected* cases, and still find that those models fit just about as well on *future* cases. In other words, two quite different models, based on quite different interpretations of history, may still predict at about the same level of accuracy.

Somewhere in the Butterworth data set we encounter a limit on the predictability of international events, at least so far as international conflict is concerned. ID3 in effect eliminates parsimony as a source of model error—the model is allowed to reach 100% predictive accuracy on the training set, which a linear model, in general, would not do. Despite this, the trees are relatively parsimonious—they do not nest very deeply and a small number of features will correctly classify most of the training cases. But 100% accuracy on a training set does not lead to 100% accuracy on the validation set: 50% accuracy is more common. In other words, on average about 50% of the cases in the validation set are not described by information from the training set when the two sets are chosen randomly. This gives some idea of the degree of unpredictability likely to be encountered in forecasting future crises, at least when using the variables of the Butterworth data set.

Consider two extreme cases of a classification algorithm attempting to create rules to identify birds. At one extreme we could have a data set consisting 1,000 separate bird species, i.e. the dependent variable has 1,000 different values. Here, the accuracy of ID3 in split-sample tests is zero—none of the validation cases overlap with the training set and therefore none can be predicted. At the other extreme, suppose the data set contained only 10 very distinct species. In this instance, the split-sample accuracy would be close to 100%, because a case would be misclassified is if it had no representatives in the training set. That would occur with a probability of roughly 0.99^{500}, a very small number.

The Butterworth conflict data are somewhere between these two extremes. It has sufficient replication that about half of the validation cases are classified correctly. In contrast to most statistical treatments, the errors are not due to limitations of the learning method or the knowledge representation scheme, since the method correctly classifies 100% of the training cases when using the full set of features. That accuracy says little, however, about the chances of classifying an unknown case correctly, and only the bootstrap tests provide a measure of how likely the variables will predict unknown cases.

Genetic Algorithms[15]

Genetic algorithms (GAs) are a general form of machine learning based on the evolutionary processes of selection, mutation and recombination. This approach was originally proposed in John Holland's *Adaptation in Natural and Artificial Systems* (1975). Holland's focus was the study of general techniques by which a problem-solving device could improve its performance in an evolutionary fashion. The problems of interest to Holland were characterized by:

- The impossibility of enumerating all possible devices for solving the problem.
- The performance of a device had a large number of local minima: i.e. the initial introduction of a component might initially degrade performance even if it ultimately improved performance.
- The structure of the problem-solving device was sufficiently complicated that it was not always obvious which components were responsible for improvements in performance

Complexity of this sort constrains the type of optimization techniques that can be used to develop problem-solving devices, and in particular classical techniques for optimization are of very little value in these circumstances.

The international environment has many of these characteristics. The problems confronting a foreign policy organization have high dimensionality, noise, slow and imperfect feedback, and a number of locally-optimal points, so developing a workable foreign policy is quite different from designing a bridge or calculating the optimal price for colored contact lenses. If Holland's theory is correct, the evolutionary approach to formulating policy may not be merely a fallback position imposed by information processing constraints, but might actually be one of the few effective problem-solving methods available.

Holland used biological genetic evolution as a metaphor for problem solving because evolution has produced billions of devices capable of functioning in a

[15] Parts of this section appeared earlier in Schrodt (1986b) and Schrodt (1989a).

continually changing, nonlinear, stochastic environment. Genetic evolution uses a fairly simple chemical device, DNA, to construct a cascading series of other chemicals that ultimately produce an organism that will produce more DNA.

Holland's theoretical contribution was the development of a *general* scheme for describing how such adaptive mechanisms might function, and showing how they could be simulated on computers.[16] The general Holland system uses a gene-like system, applying selective recombination and mutation to develop new problem-solving structures, and makes extensive use of feedback through a performance function to select of ever-improving devices. New rules are produced from existing rules by "mating" rules that have strong performance using genetic "cross-over": the first k bits of one rule are combined with the last N-k bits of the other to produce a new rule. A critical feature of mating is that the probability of reproduction is proportional to the relative strength of the rules—only the strong rules survive and reproductive success is directly proportional to strength.[17] Mutation is also applied to the new rules, which allows the possibility of producing new rules with components that were not were in the original system.

The example I will use to demonstrate genetic algorithms involves prediction, but the GA method is particularly well suited for the development of complex strategies. For example, Axelrod (1987) uses GAs to find strategies for playing Prisoners' Dilemma games; on a purely self-organizing basis, the system constructs a number of strategies that do as well as tit-for-tat, but are usually somewhat distinct from tit-for-tat. Similarly, I've experimented with determining whether a GA can develop strategies for playing zero-sum games with mixed strategy solutions; the system almost always finds the mathematically optimal mixed strategy for one of the players.[18] While the strategies derived by the GA are known for the zero-sum games and Prisoners' Dilemma, GAs might be effective in discovering strategies for games where analytical solutions are not known; Kollman, Miller and Page (1992) provide an example in the context of the adaptation of political party platforms.

Genetic Algorithms and Bureaucratic Learning

GAs are attractive as models of political learning because the evolutionary development of organizational rules is often quite explicit, and much of this development is based on selection. The GA's mechanisms of mutation and recombination capture in an abstract fashion the experimentation and imitation of bureaucratic and individual learning. The system creates new problem solving strategies on the basis of old strategies that were partially successful, while maintaining successful old strategies. While GAs

---

[16] Grefenstette 1987; Davis 1987; Rawlins 1991; Koza 1992; and Michalewicz 1992 are several general introductions to this approach.

[17] Without this, evolution occurs only randomly. The difference between the results reported here and the less successful system described in Schrodt (1986a,1986b) are due to a change from random mating to strength-based mating.

[18] In these experiments the sets of rules played against each other. The optimization halts when one player has found the optimal mixed strategy because once that strategy is played, all mixed strategies used by the other player produce the same results—the defining characteristic of the optimal mixed strategy in zero-sum games—and consequently there is no selective pressure on the second player's strategy.

make no attempt to actually simulate human learning, they show the classical learning curve of human learning, and exhibit some learning dilemmas such as the tradeoff between optimally adapting to a fixed environment (high performance/low flexibility) versus maintaining the ability to adapt to new environments (lowered performance/high flexibility). GAs can also exhibit elements of "artificial stupidity", where an organization becomes caught in a locally maximal behavior pattern that has low overall performance.

The evolutionary aspect of rule development was discussed in Chapter 3 and is widely evident is most organizational behavior. In the its written rules and procedures of an organization this evolution is explicit and could, with sufficient effort, actually be traced over time, but evolution also applies to the informal rules and norms of behavior. The population of rules governing an organization is under selective pressure—intense in a competitive business environment or a rapidly changing international situation, weak in the day-to-day workings of foreign policy—and a rule that causes a major problem will be discarded.

Adaptation proceeds using both imitation and experimentation. Even a "new" organization or policy—for example the human rights initiatives introduced by the Carter administration—is likely to contain large elements of policy and practice adopted with little or no modification from earlier policies. For example the Carter human rights policy drew on earlier United Nations initiatives as well as programs and criteria developed by private organizations such as Amnesty International, and used a combination of diplomatic and public relations techniques previously developed in other contexts.

The genetic algorithm, with its random mutations and recombinations, is probably much less efficient than the experimentation and imitation of a human organization. Human organizations can use foresight and planning to consciously seek out solutions that are likely to work and discard without trying many of those that are likely to fail. When the international environment presents novel situations where those consequences are not known—for example, new phenomena such as Nazi Germany, nuclear weapons, guerrilla movements with access to the international arms market or Shi'a fundamentalism in Iran—then even a search for policy options assisted by deductive foresight may be little different than a random evolutionary search. In other words, once a general policy has been found by adaptation, deduction allows the strategy to be fine-tuned in a relatively systematic fashion, but this does not work for the big picture.[19]

The evolutionary approach is not a sure path to an optimal solution, nor even a satisficing solution. One risk, found in biological populations as well as artificial ones, is "genetic drift": the population of rules may become sufficiently uniform that mutation and recombination no longer significantly change them. This is unproblematic in a fixed environment but causes major difficulties when the environment changes. For example, corporate giants such as IBM, Sears and General Motors and the Communist parties of Eastern Europe evolved bloated bureaucracies and volumes of rules to insure that policies developed by managers in the 1950s were still adhered to by their grandsons and granddaughters. These organizations, well adapted to the 1960s and 1970s, found themselves at a substantial disadvantage in comparison to smaller organizations with very

---

[19] The contemporary debate on the implications of the end of the Cold War is an excellent example of this: it shows little consensus, even among individuals sharing similar theoretical perspectives.

high experimentation/mutation rates—less well adapted but substantially more flexible—when their environments changed in the 1980s. This drift that can ultimately lead to the catastrophic failure of an organization is a direct consequence of the mechanisms that allowed the organization to initially succeed.

Application

To illustrate the use of a genetic algorithm in the context of international behavior, I will focus on a very simple problem of short-term prediction using a Holland classifier, one of the most widely studied GAs.[20] The Holland classifier is based on a production system-like scheme with three components: messages, a rule-base, and a "bulletin board". The overall scheme works as follows.

1. **Messages** are posted on the bulletin board to describe the situation.
2. The bulletin board is scanned by the rules for messages that match the **classifier** of the rule. If a match is found, the rule "bids" for the right to replace that message with the rule's **result**. Bids are based on the **strength** of the rule and the **specificity** of the match.
3. The highest bidding rule gets to replace the message. It then **pays** the bid to the rule that posted the message it responded to.
4. This process continues until a **terminal** message is posted, which becomes the classification. The rule posting the terminal message is rewarded with a payoff. This payoff is positive if the terminal message is the correct classification (based on some evaluation function); it is negative if the classification is incorrect.

I have called this process of posting a single message (i.e. data point), bidding, and classifying (i.e. predicting) an "experiment".

Rules in the Holland system consist of three parts: a "classifier", a "result" and "strength". Strength is a number; the classifier, result and messages are all strings of symbols of the same length. Rules in effect are *if...then* statements like those used in expert systems: the classifier is the antecedent and the result is the consequent. Holland's scheme uses an alphabet with three values:

0	feature is absent
*	pass-through: feature may be present or absent
1	feature is present

---

[20] Holland's description of the classifier is found in Holland (1986); Goldberg (1989; chapters 6 and 7) provides a very readable description as well as the discussion of a number of applications.

My version of the classifier differs from Holland's in several details, though none of these are critical to the underlying logic of the process. The major difference is my initializing the system via historical learning and my use of a loose rule-matching criterion. In Holland's system, a rule can bid for a message only if it matches the posted message exactly except for pass-through characters. My system instead makes bids proportional to the strength of the rule and the *degree of fit*, rather than insisting on exact fits. A rule and message must meet at least a minimum level of fit controlled by the parameter *Match_Minimum*. If two rules make identical bids, the first bidder will win; rules are periodically shuffled in order to randomize this.

The same alphabet is used in the result, with the provision that the value encountered in the message matched is substituted for the pass-through character, so messages contain only [0,1] values.

The process of competitive bidding and payoffs for posting terminal messages insures that the rules that have been the most successful in posting correct messages in the past will be most likely to do so in the future. Rules posting meaningless or incorrect messages are reduced in strength so that they are less likely to bid successfully. Similarly, the requirement that rules pay the rules that posted the messages to which they responded means that rules can gain strength by setting up situations that lead to the successful posting of a terminal message, a process Holland calls the "bucket brigade".

Main Classifier Loop

```
Loop
With each message do
 With each rule do
    1. Compute Match(Rule, Message)
    2. Compute
            Bid=Match*Strength*Bid_Weight
    3. If Bid>Highest Bid for Message then
       Replace Bid_Winner with Rule
 End_With {rule}
 With Bid_Winner do
    1. Subtract Bid from Strength
    2. Add Bid to Strength of Rule which posted Message
    3. If Bid_Winner's Result is a Prediction
         Then
       1. Compute Payoff by comparing prediction to actual
          events.
       2. Add Payoff to Strength
       3. Remove Message from Board
         Else
       1. Replace Message with Result on Board
 End_With { Bid_Winner}
End_With {message}
End_Loop

 The loop is continued until one of three conditions has been met:

 1. There are no messages left on the board
OR
 2. All bids for rules are below a set threshold (i.e. none of
    the rules match the messages particularly well)
OR
 3. The loop has been repeated more than a set number of times.
    This rule prevents infinite looping due to recursive
    parasitic rules whose classifiers match their own output
    messages.
```

Implementation

Converting prediction to a classification problem is simple. A set of N events can be coded in the Holland scheme as a binary string of length N, where a 1 in position k indicates that event k is in the set; a 0 means the event is not in the set. The system was tested using COPDAB event data on three North Atlantic dyads for 1948-78; the details of the coding are described in Appendix A. The COPDAB coding scheme uses 15 categories of events, ranging from 01 for highly cooperative events such as two nations merging to 08 for demonstrations of indifference to 15 for major violent hostilities.

The COPDAB data were used to generate event sets I've called "archives"; this is the set of events in a time period, coded for occurrence without regard to frequency. A prediction problem is created by choosing a random date T and then coding the following archives (the metric is days):

Input messages:	[T-1 to T-10]
	[T-11 to T-20]
	[T-21 to T-30]
	[T-31 to T-40]
Output occurrence:	[T to T+20]

In other words, the system attempts to predict the types of events in the dyadic relationship for a twenty-day period based on the events in four previous ten-day periods. The coding allocated 16 bits to the events and 4 for the lag. This gives a surplus of at least two bits (and actually more, as event types 01, 14 and 15 aren't found in the data) that the system could adapt to identify intermediate rules.

The initial rules in the system were generated by reading a random set of data records, randomly choosing one of the lagged archives as the classifier and the output archive as the message. Both parts of the rule were then randomly mutated by changing some of the bits to a pass-through character. This captures, in a primitive fashion, the notion of precedent: the initial rules are based on earlier observed sequences. Data archives from each of the COPDAB sets were sampled in random order rather than sequential order in order to avoid biasing the system with the lower-density data that occurs in the early years of the data set. In order to limit the amount of running time to something reasonably finite—these experiments were done on an Apple II!—a set of only 32 rules was used. A couple of non-evolutionary runs were made with 128 rules and no major differences were observed.

The input archives, with information identifying their lag (0, 10, 20 or 30 days) were posted on the bulletin board of the classifier and a measure of the difference between the prediction produced by the classifier and the actual outcome occurrence was used to compute the payoff. The evaluation measure used the difference between the number of correct and incorrect predictions to avoid the system simply predicting all events.

The function *Match* is a simple comparison of the features of the classifier and the message. Match was initially defined simply as

$$\text{Match} = 40 - \sum_{i} |\text{Classifier}_i - \text{Message}_i|$$

with the summation over the features using the numerical values

'0' = 0   '*' = 1   '1' = 2

A perfect fit would have a value of 40; a perfect mismatch would have a value of 0, and a match against a classifier consisting solely of * would have a value of 20.[21]

A prediction is identified by the code '1111' in the final four bits of the message. The *Payoff* function compares the output predicted in the message to the observed outcome: this function is similar to the *Match* function except that it can take negative values:

$$\text{Payoff} = 20 - \sum_{i} |\text{Outcome}_i - \text{Message}_i|$$

If the prediction is particularly inaccurate, the payoff will be negative and decrease the strength of the rule that posted it. This punishes rules for posting incorrect predictions.

To allow for the possibility of the system recognizing that it has not seen a situation before and responding by not making a prediction, there is a minimum threshold for acceptable bids. If no rule can make a bid above the threshold, then the system creates a new rule by randomly choosing one of the messages on the board (i.e. one of the actual antecedent archives), mutating this with a few pass-through characters, then taking the actual outcome as a result. This rule then replaces the weakest rule in the classifier. Thus if the classifier encounters behavior which it has not seen before, it should incorporate that new behavior into its rule base.

The genetic learning of the system follows the Holland scheme described above. Evolution occurs at the end of each 50 or 100 experiments. As in most Holland classifiers, the system relies mostly on recombination rather than mutation for evolution.

Genetic Evolution Algorithm

```
1. Sort rules by Strength
2. Discard any rules which have no made any successful bids in
   the previous eon.
3. Discard weakest rules until only Extinct_Probability * N_Rules
   remain
4. For k:= Extinct_Probability *N_Rules +1 to N_Rules
     with rule k do
     1. Randomly choose two "parent" rules from the surviving
         rules, with the probability of being chosen being
         proportional to the strength of the rule.
     2. Choose a random cross-over point p between 1 and
         2*N_Feature
     3. Create a new rule from the first p bits of the first
         parent and the last N_Feature-p bits of the second parent.
     4. Mutate each bit in the new rule with probability
         Mutation_Prob:
            [0,1] mutate to *
            [*] randomly mutates to 1 or 0 according to
               their frequencies in the surviving rules
     End_With
```

---

21 In other experiments I replaced this metric with one that did not penalize the matching against pass-through codes—this follows Holland more closely and slightly improves the performance of the system.

Results

Benchmark estimators

The classifier is a stochastic process and its performance must be measured statistically. Random behavior enters at the selection of the initial rule set, the random sampling of the database, and the random mutations of the rules. In general the variation in the performance of the system is quite bounded but unusually high and low performance occurs occasionally even in a mature system.

A measure of prediction success, S, was computed for several estimators:

$$S = \frac{\text{(correct predictions - incorrect predictions)}}{\text{total predictions}}$$

This produces values between -1.0 (all incorrect) and 1.0 (all correct).

In order to ascertain whether the Holland classifier was doing anything other than simply making random predictions, comparative performance measures from alternative estimators were computed. Let P_k=Probability of occurrence of event k (i.e. the probability of a "1" is position k of the archive) based on the samples observed and E_k be the k^{th} element of the archive, then the estimators are:

Random: $E_k = \begin{cases} 1 & \text{with probability } P_k > 0.5 \\ 0 & \text{otherwise} \end{cases}$

Previous: E_k = value of E_k in previous observed output archive

Best: $E_k = \begin{cases} 1 & \text{if } P_k > 0.5 \\ 0 & \text{otherwise} \end{cases}$

The *Random* estimator makes a random guess on the basis of the observed frequency of E_k. *Previous* just uses the previous observed archive to predict the next archive; because the data are sampled randomly this is not the temporally previous archive, but rather another archive taken randomly from the population. As a bit of calculus will show, *Best* is the estimator that maximizes the value of S based on P_k as an estimate of the population value of P_k.

For the US/European data set, *Random* =.683, *Previous*=.661 and *Best* =.773, based on 10 random sets of 100 samples. The striking feature here is the fact that there is a great deal of regularity in the COPDAB-based data sets, in large part based on the lack of data density, so that even a poor estimator such as *Previous* manages a score around 0.66.

Overall Performance

The program has several free parameters and due to the time required to run the program, I had the opportunity to experiment with only a few combinations. In the initial experiments on the parameters, the program was initially run in a non-evolutionary mode, with an eon-length of anywhere from 120 to 200 samples. After a variety of runs, it became clear that the predictive success of the system stabilized after about 50 samples:

On average the success ratio changes only about 0.005 to 0.01 from experiment 50 to experiment 100. Statistics were collected on the Holland success and the Best statistic at every ten samples; the analysis reported below is done on the ratio of these two. The "ratio" statistic is the ratio

$$R_t = \frac{\text{average } S_{\text{Holland}}}{\text{average } S_{\text{Best}}}$$

The *Best* statistic was given all of the information that the Holland classifier had, so the prediction made on the first sample includes the 32 samples used to produce the rules in the classifier. The average S of the Holland classifier is based only on the experiments that have occurred since the previous evolutionary phase, in other words, it measures the performance of the current set of rules.

The typical performance of R_t is shown in Figure 5.2, which used the parameter values

Mutation Probability	.20	Eon_Length	50
Reward_Adjust	1.25	Match_Minimum	38
Extinct_Probability	.50	N_Rules	32

Performance shows a very gradual increase in performance asymptotic to 1.0. This compares to an asymptotic R_t ratio of 0.884 for the *Random* estimator and 0.855 for the *Previous* estimator.

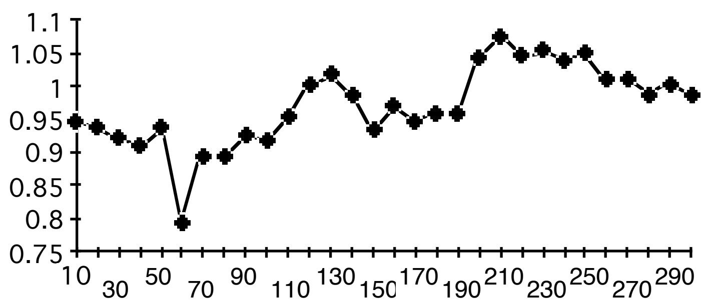


Figure 5.2. Performance as a Function of Number of Experiments

The random fluctuations in the performance are largely due to the irregularities in the data set; when the classifier hits a series of atypical cases, its performance drops. Figure 5.3 shows the mean values of R_t sampled at five 100-experiment intervals. Comparing the mean performance during the first five-hundred and second five-hundred experiments, one can see a consistent pattern of a slight improvement when the mutation probability is less than 0.35.

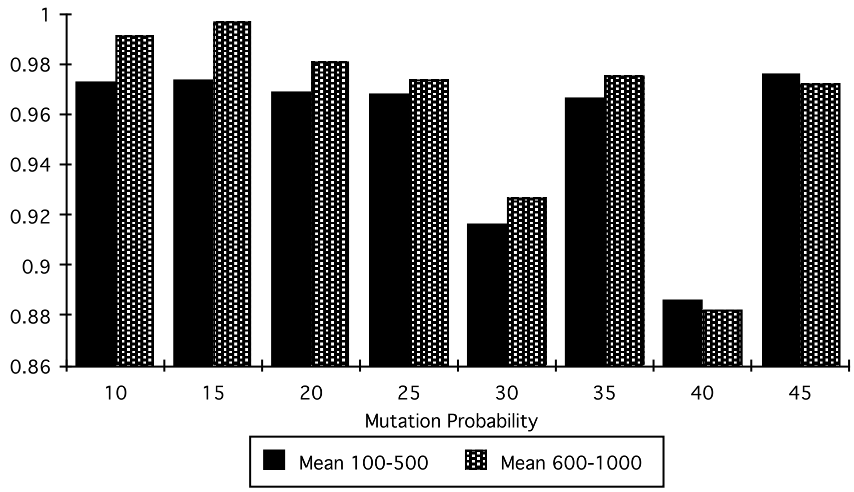


Figure 5.3. Mean Performance as a Function of Mutation Probability and Experiments

Effects of parameters

The Holland classifier has several free parameters and I did a variety of experiments with these. The results of these are reported in detail, albeit for a somewhat different program, in Schrodt (1986a); those results will be briefly summarized here.

The *Mutation_Probability* is the probability of a spot mutation, which can occur under two conditions. First, mutation occurs when a rule is introduced from the data set during rule initialization or due to the lack of a good match. In such cases the only mutations occurring are 0 or 1 changing to *. Second, mutations occur in rules created during evolution. These mutations can involve * changing to 0 or 1 as well as 0 or 1 changing to *.

Figures 5.3 and 5.4 show the effect on performance of changing the mutation probability in the range 0.10 to 0.45. Figure 5.3 shows this in terms of mean performance; Figure 5.4 provides a more fine-grained track of the performance over a number of experiments. In general, lower mutation rates—those in the 0.10 to 0.20 range—seem to do better than higher mutation rates; this is particularly apparent when one looks the effects of mutation across a variety of parameter combinations not reported here. As would be expected, the variance of performance is also higher with the higher mutation rates.

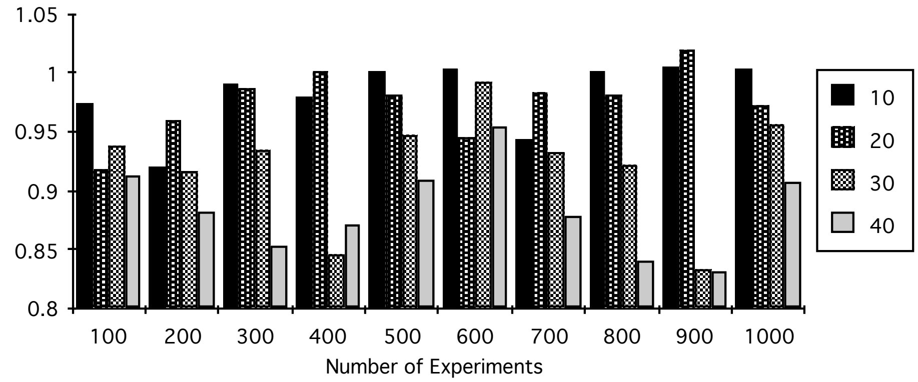


Figure 5.4. Performance as a Function of Mutation Probability

The deleterious effects of high mutation rates are important for understanding the causes of organizational rigidity. In general, for any environment there is an optimal rate of experimentation (mutation) for an organization. If experimentation exceeds this rate, the organization is wasting efforts trying to improve its performance: most changes will be deleterious. As noted in Chapter 3, the longer the organization has been in a stable environment, the greater the likelihood that it has found a position in the space of rules that is not only locally optimal, but is also superior to any other configurations that can be reached by incremental changes in its rules.

Such was the situation of IBM, Sears and the Polish Communist Party until the late 1980s. However, in the 1980s, the positions of these organizations were challenged not by slow, incremental changes in their environment—changes compatible with their low mutation rate—but rather by challenger organizations with radically different sets of rules: Apple, Compaq and Microsoft; Walmart; and Solidarity and Gorbachev. These new organizations effectively changed the overall environment, and the previously well-established organizations found themselves trapped in the pincers of a rapidly changing environment and a bureaucratic structure that had evolved to resist change.

The parameter *Reward_Adjust* is used to determine the payoff to a rule for making a prediction. There is no discernible relationship between this parameter and performance: it can be set to most any reasonable value and about the same results obtain. Even if rules, on average, lose strength on a bid the accurate rules will still lose relatively less strength than less accurate rules, and hence remain competitive.

I allowed a rule to bid even when there was an imperfect match between the message and rule. This was controlled through the *Match_Minimum* parameter. The program was run with Match_Minimum set at 40, 38,36, and 34, which corresponds to 0,1,2, and 3 incorrect matches (i.e. a 0 matched against a 1 or vice-versa) respectively. The 36 level had a very slight advantage over 38—about a 0.01 difference in mean R_t. When the strict matching criterion 40 is used, rule replacement occurs with a high frequency: few rules get an opportunity to build up any strength and good rules are vulnerable to being wiped out should they lose strength in a series of atypical cases. Conversely, loosening the match criterion to 34 allows too much flexibility and encourages random bidding. While I have not done sufficient experimenting to draw a firm conclusion on this issue, slight looseness on matching (1 or 2 mismatches) seems best.

Finally, I experimented with resetting and not resetting the strength of each rule at the end of each eon. Resetting strength means that all rules start out on an equal footing at the beginning of each eon; not resetting means that rules "inherit" their strength across the evolutionary phase of the simulation. There is a clear advantage to not resetting, and the results reported here are based on runs using inherited strength.

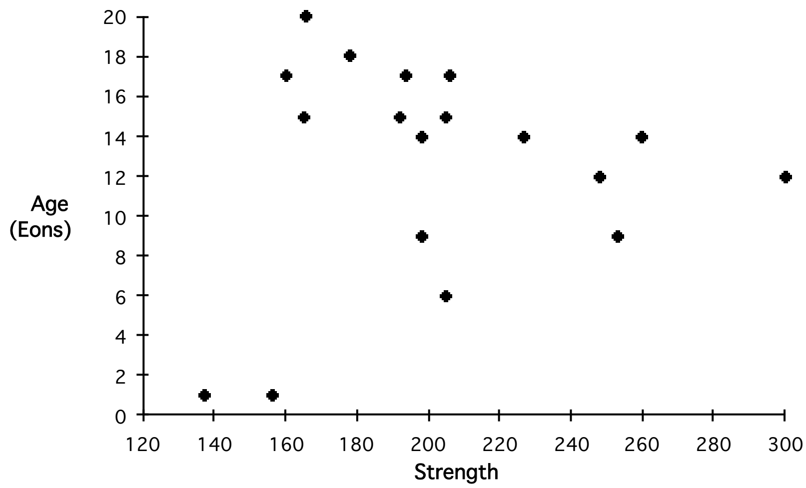


Figure 5.5. Rule Age versus Strength at Eon 20

Figure 5.5 shows a scattergram of the strength of rules versus the age of rules at the end of Eon 20[22] . Only the rules that made a successful bid are displayed; the remaining rules had an age of 100 and the initialized strength of 128 and were wiped out in the evolutionary phase. Two characteristics are clear. First, a preponderance of the rules are old—72% of the rules displayed are 10 eons old or older. Second, there is an apparent negative relationship between age and strength—the eight rules aged 15 and older have a mean strength of 183; the six rules aged 6 to 14 have a mean strength of 249. When one examines the characteristics of rules surviving the evolutionary process, they tend, unsurprisingly, to approximate the *Best* estimator (i.e. they predict modal behavior), though they have somewhat greater variety than in the statistical estimator.

Discussion

The Holland estimators managed to achieve virtually the same accuracy as the statistically optimal *Best* estimator is impressive when one considers the following factors. The Holland classifier's success is far better than what would be produced by chance, never dropping to the level of the *Random* and *Previous* estimators and it closely approximates the statistically optimal *Best* estimator despite the fact that the classifier has no information about the nature of the optimal predictor. I had to use calculus to figure out *Best*, but the Holland system was able to do as well on its own.

At the same time, the classifier didn't do any better than the statistical estimator, and potentially it should be able to. One problem in getting high average performance

---

22 Eon length was 100.

with the classifier is that inaccurate predictions are required to clear out unsuccessful rules and these errors affect the overall performance. An alternative approach would be to evaluate the system by distinguishing between two types of predictions: low-match predictions used only for learning—the system is saying "I'm going to try this out but don't hold it against me if I'm wrong…"—and high-match predictions where the system is confident, based on the strength and match of the rule, that the prediction will be accurate. This would be similar to how human analysts operate: one learns in private and only makes public predictions when one is confident that those predictions will be correct.[23]

The modal character of the COPDAB data also made it difficult for the classifier to improve on the modal predictor. However, some indication of an environment where the Holland system would show superior performance occurred in an inadvertent but enlightening test of the system in a rapidly changing environment. While debugging the program, I cleverly turned off array bounds checking to save 10% on run times, and then forgot to reinitialize a variable, resulting in two days worth of runs where the classifier roamed through the computer's random-access memory reading "data" until some fatality occurred with the operating system. In this situation, the classifier performed considerably better than the Best estimator because, with rule-replacement, it adapted fairly quickly to the fact that the data had changed; the *Best* estimator was slower to respond. In this artificially complex environment, R_t values of 2.0 and greater were not uncommon. In a denser, more varied international events data set than these COPDAB dyads, the Holland classifier would probably be superior to a statistical estimator.

Neural Networks[24]

Neural networks have attracted a great deal of attention in the past five years as a solution to complex classification problems with ambiguous and noisy inputs.[25] The term "neural network" applies to any algorithm using a data structure composed of "weights" and "neurons". Computationally, "weights" are real numbers (either positive or negative), and neurons are functions. The value of a neuron—its "state" or "output"—is determined by the weighted sum of the values of the neurons to which it is connected. If N_i is connected to a set of neurons N_j, $j \varepsilon J$, through a set of weights w_{ji} then the state of N_i is determined by

---

23 Or when one is a well-paid media pundit whose future employment is unrelated to the accuracy of the predictions.

24 Much of this material originally appeared in Schrodt (1991d).

25 Hecht-Nielsen (1990), Wasserman (1989), Caudill (1989), Freeman and Skapura (1991), Schalkoff (1992), Michie, Spiegelhalter and Taylor (1994) and Hertz, Krough and Palmer (1990) provide introductions to this method at various levels of formality; Rumelhart et al (1986) present a detailed discussion of various techniques along with considerable detail on the physiology of biological nervous systems. Garson (1991) and Kimber (1991) provide two additional applications of neural networks in the social sciences. A great deal of commercial (and shareware) software is now available for implementing neural networks, so this is one method where it is probably not worthwhile to write your own code, particularly if you have access to a package that allows experimenting with a variety of network configurations and training methods.

$$N_i = F(\sum_j N_j w_{ji}) \quad (5.1)$$

where F is some function, usually nonlinear and monotonic.

This mathematical structure is a simple approximation to the mechanism through which biological neurons—the basic cells of the brain and the nervous system—operate. Biological neurons are interconnected by long branching dendrites; the activity of one neuron may either excite another neuron or inhibit it. In a neural network, excitatory connections correspond to positive weights; inhibitory connections to negative weights. While mental activity is more complex than the simple activity or inactivity of neurons—for example it is mediated by a vast array of neurochemicals that influence the transmission of signals—the process of neural stimulation and inhibition is clearly an important aspect of the brain physiology.

Figure 5.6 shows a schematic of a typical neural network. The network has multiple "layers" of neurons. The input layer is activated by the values of the independent variables of the problem to be solved. The input layer values in turn activate the hidden layer according to equation 5.1, and the hidden layer activates the output layer. For example, a neural network designed to identify letters of the alphabet might contain an input layer determined by a grid of nine boxes placed over the letter. Each box would control four input neurons—giving a total of 36 neurons in the input layer—that would be activated if the box contained a horizontal, vertical, right diagonal or left diagonal line, as illustrated in Figure 5.7. The output layer would consist of 26 neurons, each corresponding to a letter of the alphabet.

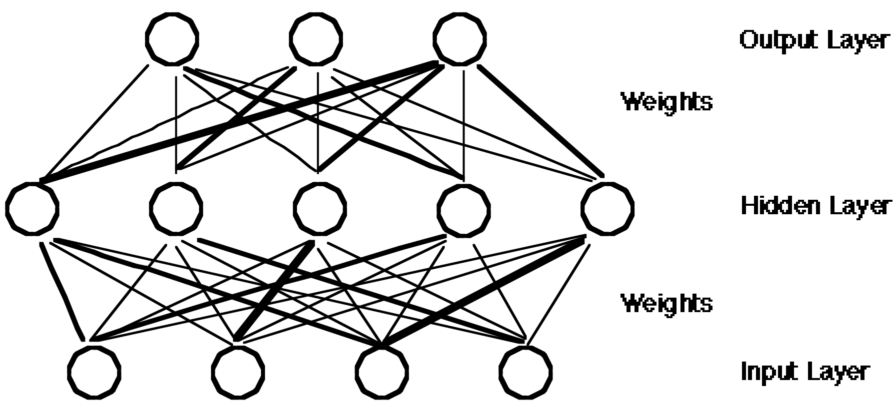
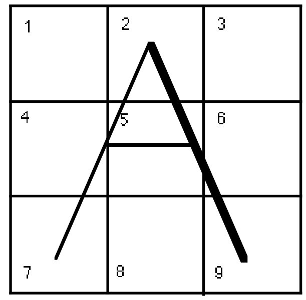


Figure 5.6. Neural Network Schematic

In the example in Figure 5.7, the inputs are binary. This configuration is useful when dealing with nominal or categorical data, since their values can always be expressed as a binary string. If the network is working with interval-level measures—for example economic data or a physical process such as a control of a pipeline—neurons could be set to real numbers and Equation 5.1 applied, or the binary and real-valued inputs could be mixed in the same network.



Input Vector:	1	2	3	4	5	6	7	8	9
	0000	0011	0000	0010	1011	0010	0010	0000	0001

Figure 5.7. Input Coding for a Neural Network

The values of all of the neurons except those in the input layer are determined by summing weights and neuron values according to Equation 5.1. This is called a "forward propagation" process: the values of each layer of neurons are determined by the values of the neurons below them and the intervening weights. The performance of the network is determined entirely by the values of the weights, so the weights are the knowledge representation structure of the neural network. In the letter-recognition problem discussed above, the weights would need to be set so that when the input layer was set to a vector describing the letter "A", only the A neuron in the output layer would be activated; when set to a vector corresponding to a "B" only the B neuron would be activated and so forth.

Because a neural network contains a large number of weights, their values cannot be determined analytically except in the most trivial problems. Instead, the weights in a network are randomly initialized and the neural network is trained by examples. Cases are presented for which the desired state of the output layer is known, and the difference between the actual state of the output layer and the desired state is used to adjust the weights. The presentation of training cases and weight adjustment continues until the weights in the network stabilize.

Research on neural networks dates back to Hebb (1949) who proposed the simple learning rule for adjusting weights that forms the basis of many algorithms still in use. As the basic physiology of neurons became better understood and with the advent of digital computers, the prospect of simulating realistic neural activity using computer-implemented mathematical models was explored. This led to a flurry of research in the 1960s on "perceptrons" (Rosenblatt 1962) that had an input layer connected directly to an output layer—without the hidden layer common in contemporary neural network research—and the linear activation function

$$N_j = \begin{cases} 1 & \text{if } \sum_i w_{ji} N_i > t \\ 0 & \text{otherwise} \end{cases}$$

While simple by contemporary standards, perceptrons had many of the characteristics of neural networks, including parallels with biological neural behavior, training by example and some aspects of associative memory and error correction.

Perceptron research was brought to an abrupt halt in the late 1960s by an important but thoroughly misinterpreted analysis by Minsky and Papert (1969), which proved mathematically that a perceptron could not solve a very basic class of discrimination problems, the exclusive-OR (XOR) problem. Since XOR is a fundamental logical operation that could be expected to be part of any serious classification problem, this was quite a damning limitation of the perceptron concept.

Minsky and Papert's paper marked the virtual end of neural network research for about a decade and a half. Unfortunately, the computer science community did not properly understand the limitations of Minsky and Papert's work—a mathematical proof—. Minsky and Papert had shown that a *perceptron*—a neural network with single layer of weights and a linear threshold response function—was unable to solve the XOR problem. Their proof said nothing about whether multiple-layer networks could solve the XOR problem, nor did it deal with non-linear response functions. While Minsky and Papert made no mathematical claims beyond the analysis of perceptrons, their work was popularly interpreted as applying to neural networks generally.

Despite the demise of the perceptron, some research continued on more general neural networks, most notably by the Carnegie "Parallel Distributed Processes" group (Rumelhart et al, 1986). From a mathematical analysis, these researchers knew that *multi*-layer, *nonlinear* networks could solve the XOR problem (see Wasserman 1989, chapter 2); the problem was finding an appropriate training algorithm for such networks. The breakthrough came with the "backpropagation" algorithm developed by PDP research group.[26] The backpropagation method has now been used in a variety of successful applications of neural networks to real-world problems (see Caudill 1990) and has formed the basis of a great deal of experimentation with various neural network configurations and training techniques since the mid-1980s. A wide variety of training techniques are now available in addition to backpropagation; to date none of these have provided dramatic breakthroughs in accuracy or training efficiency, though they may be preferable to backpropagation in some problems.

The current research on neural networks uses two modifications of the perceptron. First, one or more hidden layers are used between the input and output instead of the direct connections of the perceptron. Second, a nonlinear response function is used, typically the "sigmoid function"

$$s = \frac{1}{(1 + e^{-r})} \qquad \text{where } r = \sum_i N_i w_{ji}$$

This function is the familiar "S-shaped" curve used in statistical logit analysis. While the sigmoid function has some desirable mathematical attributes, the exact form of the response function is not critical and a variety of other bounded, nonlinear monotonic functions have also been used. The key to the sigmoid response is that it compresses high values of $\sum_i N_i w_{ji}$ to the limited range [0,1] and therefore no single value in the input layer can propagate through the hidden layer to dominate the response of the output

---

[26] As Wasserman (1989,43) notes, the backpropagation training method had actually been anticipated in articles published in 1982 and 1974, but none of these earlier results were noticed at the time.

layer. This ability to filter the effects of outlying input values contrasts with linear models, though it is also found in nonlinear statistical methods such as logit analysis.

While neural networks were initially attractive as a simulation of a biological information processing, the method has proven to have a number of more general advantages. First, neural networks provide a form of knowledge representation distinct from either the *if...then* logic of rule-based systems or the linear models of statistical studies. Second, because the information processing of the network is diffused among hundreds or thousands of weights, neural networks tend to be robust against noise. A single input value rarely determines the behavior of the network, so missing values or erroneous values are less likely to result in incorrect classifications. Finally, in common with ID3 and genetic algorithms, neural networks are trained by example, so they may be able to solve problems that human experts have not solved.

Neural Networks as an Organizational Metaphor

Neural networks may provide a useful structure for predicting human political behavior due to their parallels to *organizational structure* rather than because of analogies with human brain.[27] A hierarchical organizational structure would be capable of decision-making and some associative recall despite noise and the failure of some of its component parts. Such a structure also emphasizes the perceptual role of the individual while minimizing the information that must be transmitted between individuals.

A foreign policy organization, interpreted as a neural network, would be seen as having the following components: The input layer consists of various intelligence gathering offices that are monitoring specific types of behavior. These offices report to a "middle management" that integrates the various reports, weighing the reported information positively, negatively or ignoring it altogether. Each of these middle level offices has access to all of the collected information, but they use it differently. For example an office interested in recommending military activity in a conflict might positively weight information on power disparities, negatively weight the presence of ethnic conflict, and ignore information on mediating agent biases. These middle-level offices then make positive or negative recommendations to the highest, policy-making level, which integrates all of their recommendations into a final decision. The "output layer" is probably not a set of individuals so much as a small set of standard operating procedures from which the penultimate layer of managers chooses—for example the {negotiate, invade, bomb, blockade, do nothing} options of the Cuban Missile Crisis.

Based on what is now known about the properties of neural networks, this approach provides at least three advantages in decision-making; these might also be instructive in explaining the ubiquitous nature of hierarchical organizations. First, in contrast to a rule-based system, a network has some associative recall, error correction, insensitivity to missing information, and resistance to systemic failure due to the failure of individual components. Because organizational components do fail randomly—inexperienced or incompetent personnel dispersed throughout the organization as well as the effect of retirements, illness and other distractions—insensitivity to component failure is

---

27 See discussion in Chapter 7.

particularly important. Those capabilities derive from the networked structure of the organization, not from the cognitive capabilities of its individual members. We expect error correction and robustness in organizational decision-making because we experience those features in individual cognition, but a purely rule-based organization would not exhibit them. Furthermore, as Minsky and Papert pointed out, a simple two-layer network, devoid of the much-maligned "middle management", cannot solve many simple problems.

Second, a network structure is an effective means of dealing with organization bandwidth limitations. Individuals at the input level act as feature detectors: In deciding whether to raise concerns about an issue, they can use their high bandwidth associative and subcognitive capabilities to infer complex motives, draw historical analogies, and deal with multiple counterfactuals and contingencies.

Individuals in the middle and output layers must weigh the information coming in from the levels below them and continually adjust those weights, but these are memory-intensive processes.[28] In contrast, the information that must be *transmitted* through the limited organizational bandwidth is simple: one needs only to know whether individuals are sending "on" or "off" signals, and the only decision one must transmit is one's own resulting state. Information processing tasks requiring high bandwidth are placed in individuals, not in the links between them.

In human organizations, the information transmitted in memos is frequently more complicated than a yes/no decision, but the neural network model indicates that the transmission bandwidth could actually be as narrow as a single bit without impeding robust decision-making. Analyses, position papers and policy recommendations tend to be simplified notoriously as they are passed up a hierarchy, a behavior consistent with each decision-making level being primarily concerned with the simple issue of approval or disapproval. Many of the queries in the final processing of a complex decision are simply, "Has Smith signed off on this?" If the answer is affirmative, the discussion proceeds; the top layer does not care *why* Smith signed off, and Smith may have signed off for entirely inappropriate reasons. Only if something goes wrong are the rules governing Smith's lower-level behaviors explored and possibly modified; this process is comparable to backpropagation in a formal network.

The neural network is only an analogy: middle and upper managers clearly do more than sum the recommendations of lower level offices and feed those into a non-linear function, just as the human brain contains components other than neurons and synapses. But the neural and organizational networks are structurally similar and one can certainly argue that the organizational structures of foreign policy decision-making resemble neural networks more than they resemble regression or expected utility equations.

---

[28] I am not suggesting that a manager literally does the mathematical calculations of a network, only that a manager learns who to ignore and who to pay attention to on various problems. Some of these weights are formally embedded in organization procedures, such as the formal organization chart and rules dealing with who must "sign off" on a decision. Other weights develop informally; for example most organizations are very efficient at bypassing individuals who are temporarily incapacitated by illness, and are often almost equally efficient at permanently bypassing individuals whose judgment is distrusted. Because humans have much higher information processing capabilities than neurons have, an individual will typically serve as both an integrator and a feature detector. Good desk officers read newspapers as well as their cable traffic.

Application

The Butterworth data set used in the experiments with ID3 was analyzed using a neural network; as with ID3 the tests were done using split-sample training. As would be expected for a neural network of the size required to study the Butterworth data, the program is relatively slow: Experiments with ten split-samples and forty training cycles for the interval-level model required about 24-hours to run on an 8Mhz Macintosh SE. While this processing speed would be increased using a faster machine, the training of neural networks is a notoriously slow process in general.

Neural Network Configurations

Two different configurations of the neural network were studied. In the interval configuration, each independent variable was treated as an integer, with the input neuron set to that value. These values are multiplied by the network weights as they propagate forward to the hidden layer, with the sigmoid function forcing the neuron values to remain in the range [0,1]. There were 24 input neurons corresponding to the 24 independent variables. Experiments were done with hidden layers of 6, 9, 12, 24 and 48 neurons; the output layer had 3 neurons corresponding to the three possible values of the dependent variable.

In the binary configuration, a separate input neuron was used for each possible *value* of each independent variable. For example, the "Duration" variable can take the values 1,2,3 and 4, and therefore has four input neurons assigned to it: the neuron corresponding to a value of "Duration" in the case being evaluated was set to 1; the remaining neurons for that variable were set to 0. This configuration produces an input layer with 107 neurons; given the disparity between size of the input layer and 3-neuron output layer, the hidden layer was set somewhat arbitrarily at 30 neurons. With a total of 3300 weights, these networks were exceedingly slow to train and I did only limited experimentation with them.

Alternative Predictive Techniques

The critical issue concerning the utility of the neural network is whether it is an improvement over existing techniques. The performance of the neural network was compared to the results of the ID3 analysis discussed earlier, and three other methods:

Mode

As noted earlier, the Butterworth data are heavily modal so the mode provides a good "null model" for assessing accuracy.

Discriminant Analysis

The linear model appropriate for a nominal dependent variable and ordinal independent variables treated as if they were interval is discriminant analysis (Klecka 1980). Accordingly, five discriminant analyses were run using the same split-sample protocol used for the neural networks and ID3. In each case, the Butterworth cases were randomly divided into subsamples using a random number generator, a set of discriminant functions was estimated on half of the set, the other half of the cases were classified using the Fisher discriminant functions, and the accuracy of the resulting classification was tabulated. *SYSTAT 5.0* was used for the discriminant analysis.

Logit

The most commonly used nonlinear maximum likelihood statistical technique used for nominal dependent variables is multinomial logit (see Aldrich and Nelson 1984). The same split-sample samples described above for discriminant were analyzed using logit, again treating the ordinal variables as if they were interval. *SST* was used to generate the logit coefficients; computing the logistic probability functions in a separate program generated the predictions in the validation set.

Because logit is a nonlinear method, the number of independent variables that *SST* was able to handle was restricted to 13 or fewer. These were chosen from the variables determined to be most important by the discriminant and neural network analyses.[29]

Results

Accuracy in Split-Sample Testing

The same measures used in the study of ID3—accuracy and the entropy ratio—were used in this study. Table 5.5 and Figure 5.8 report the overall results of the experiments. Table 5.5 reports the average value of the accuracy and ER statistics[30] across 10 experiments in the neural networks, 5 experiments in the discriminant and logit analyses and 200 experiments in ID3; Figure 5.8 shows the results on the validation set graphically.

In the validation tests of the model—classifying the cases not originally used to estimate the model—the neural network is consistently more accurate than the ID3 model; it is roughly equal to the discriminant analysis on two of the five variables and substantially more accurate on the remaining three; and it has about the same accuracy as the logit model. The differences between the accuracy of the neural network and discriminant are statistically significant at the 0.01 level in a t-test for the Stop, Isolate and Settle variables. Information was not available to do a t-test comparing the neural network and ID3 but the network has a consistent advantage over ID3 of about 4% on all of the dependent variables.

As Figure 5.9 demonstrates, the neural network provides superior predictions only in the split-sample tests: if the training cases alone are tested, the discriminant and logit analysis appear to have a decided advantage.[31] Discriminant, however, tends to over-fit the training set, and consequently has less accuracy on the validation set; the neural network, while less accurate in the initial fit, is more robust than discriminant when dealing with new cases. None of the analytical techniques are more accurate than simply

---

29 See Table 4 in Schrodt (1990b): the variables were number of agents, likelihood of abatement, likelihood of spread, strongest antagonist, duration, degree of spread, ethnic conflict, system period, agent's bias, agent's autonomy, level of agreement, action and relative power. In 6 of the 25 variable/sample combinations two additional variables had to be deleted: likelihood of abatement and strongest antagonist. Anthony Nownes did the logit analysis.

30 The standard deviations vary with the dependent variables and techniques but are generally in the 0.03 to 0.06 range. Due to a programming error, the ER measure is not available for the binary neural network experiments but it is probably similar to the ER for the interval network; ER was also not computed for the training sets of the logit model.

31 ID3 will fit almost any training set with close to 100% accuracy so it is not included in the comparison.

predicting the mode. Since the neural network is biased towards correctly predicting the modal case, this might account for some of its greater accuracy for this data set, as might the fact that neural networks—and logit—have a nonlinear response to the input and are therefore less sensitive to outliers.

Table 5.5. Measures of Fit with Overtraining

Statistic	Stop	Abate	Isolate	Restrain	Settle
	Dependent Variable				
Accuracy					
Interval NN training	.647	.703	.839	.711	.752
Discriminant training	.918	.760	.931	.767	.779
Overtrained NN training	.805	.769	.859	.806	.836
Interval NN validation	.609	.536	.694	.523	.640
Discriminant validation	.469	.543	.547	.523	.549
Overtrained NN validation	.552	.479	.648	.497	.598
Entropy Ratio					
Interval NN training	.371	.589	.527	.597	.464
Discriminant training	.921	.766	.937	.766	.792
Overtrained NN training	.756	.736	.739	.812	.772
Interval NN validation	.320	.403	.374	.373	.302
Discriminant validation	.364	.499	.469	.490	.487
Overtrained NN validation	.387	.394	.394	.443	.366

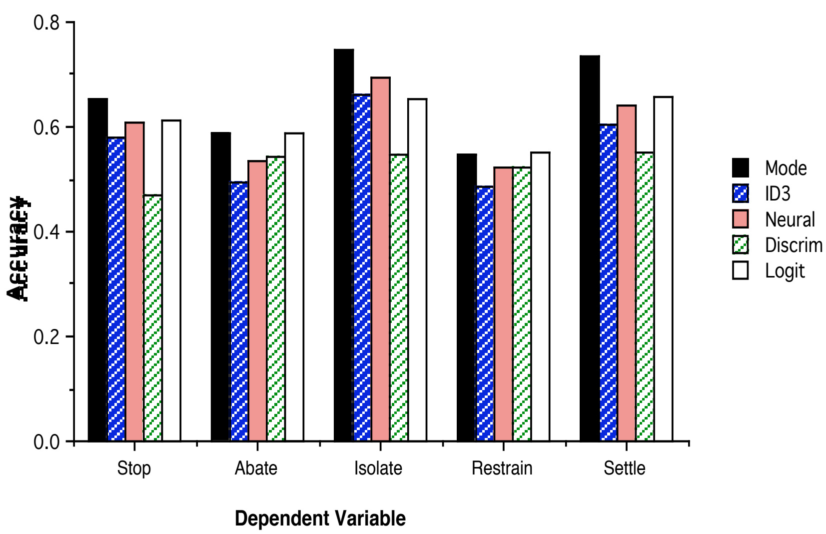


Figure 5.8. Accuracy by Method in Validation Sets

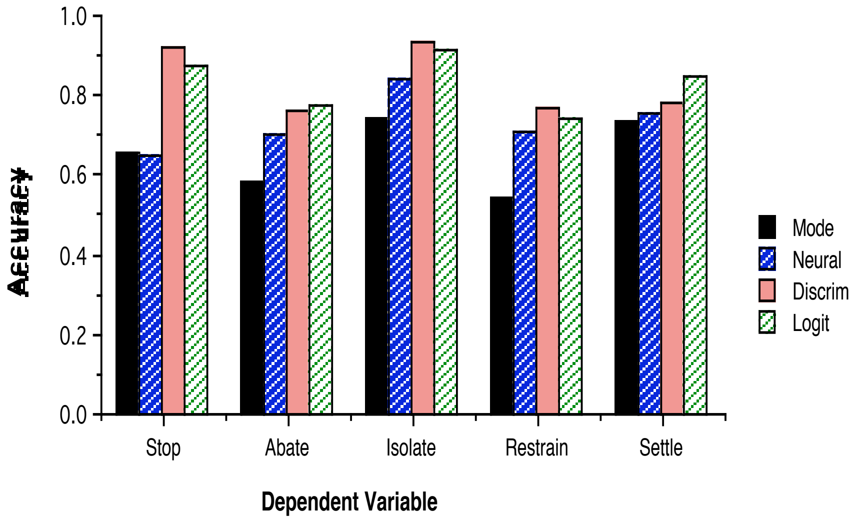


Figure 5.9. Accuracy by Method in Training Sets

The dependence of the neural network on modal predictions can be seen by looking at the entropy ratio (ER) measures. As Figure 5.10 shows, the discriminant and logit analyses do substantially better than the neural network or ID3 on this measure, and the neural network performance also tends to be slightly worse than ID3. The neural network only marginally improves on the predictive abilities of the mode, particularly on the *Stop* variable (where it is actually worse than the mode) and on *Restrain*. Interestingly, the neural network also does poorly on the ER measure even in the training sets: it never exceeds an ER of about 0.60, whereas discriminant and logit are in the 0.75 to 0.95 range during training. As noted below, these figures might be improved by very long training times for the neural network.

These results are generally consistent with those of Garson (1991), who also compared neural networks with ID3, discriminant, logit and regression, using a variety of simulated data sets. Garson found that in the simpler of his simulated data sets, neural networks provided only a modest improvement—around 5% to 10%—in the number of correct predictions, but on more complex and noisy data, neural networks had substantially higher accuracy, with an improvement of 25% in one set and 50% in another. Garson's study also found a version of ID3 that works on interval-level data to have a small but consistent advantage over regression, discriminate and logit. The data used in the Garson study, unlike those found in Butterworth, are not highly modal, and his results may be more indicative of the likely effectiveness of neural networks on typical social science data.

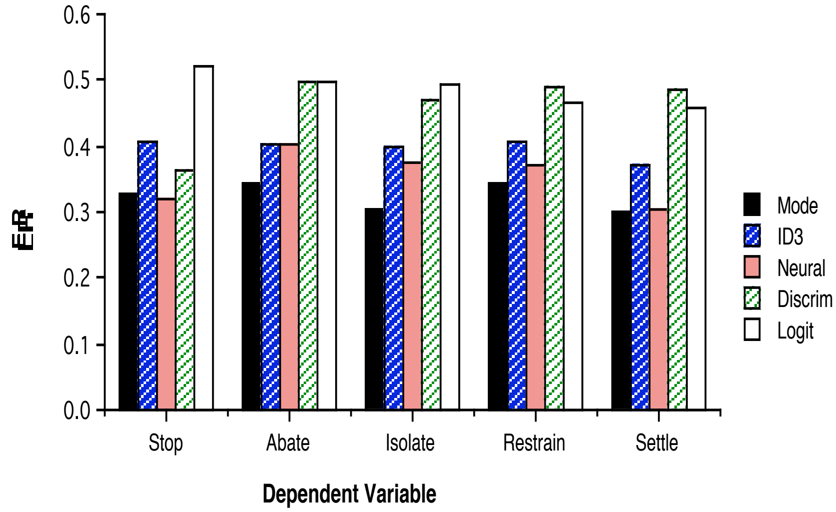


Figure 5.10. ER by Method in Validation Sets

Alternative Neural Network Configurations

I did experiments with several different configurations of the neural network. Three variations will be discussed here: modifying the size of the hidden layer; using binary coding rather than interval coding of the dependent variable; and "overtraining" for the less frequent values of the dependent variable.

Hidden layer sizes of 6, 9, 12, 24 and 48 were tested but the size of the hidden layer makes no consistent difference in accuracy or in the rate of learning. On the *Isolate* variable the accuracy actually dropped off slightly as the size of the hidden layer increased, but this did not hold for the other variables. Since the training time of the network is proportional to the size of the hidden layer, most work was done using a hidden layer of 12 neurons.

As noted earlier, the use of the ordinal independent variables as interval inputs to the input layer is metrically dubious, though no worse than comparable infractions chronic in applications of regression analysis. Binary coding treats these variables in a strictly nominal fashion. This provides additional information by eliminating the artificial magnitudes of the ordinal codes, but does so at the substantial computational cost of an input layer containing 107 neurons.

Figure 5.11 shows the performance differences between the binary and interval configurations: these are minimal. Unsurprisingly, the binary configuration does consistently, though not dramatically, better than the interval configuration on the training set. The two configurations are indistinguishable on the validation sets: the binary configuration does slightly better on three of the five variables but these differences would clearly not be significant in a t-test. In view of the substantially faster training rate for the interval configuration, it was used in the remaining experiments.

Figure 5.11

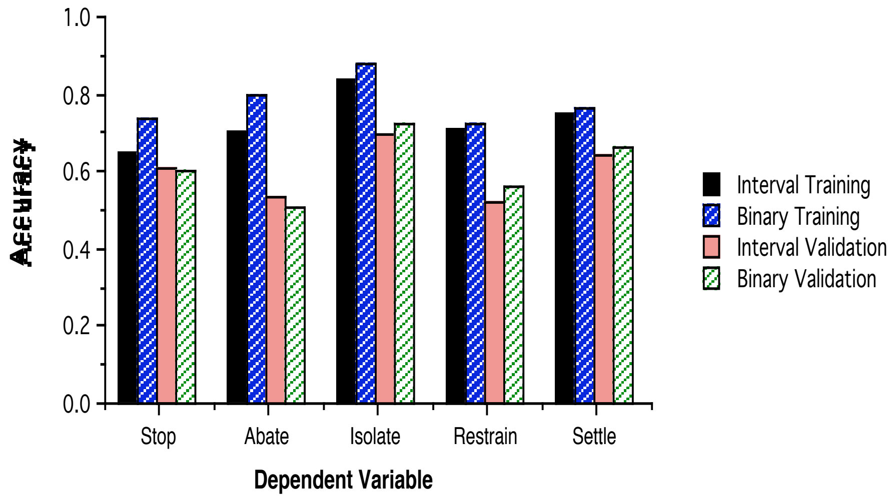


Figure 5.11. Comparison of Interval and Binary Configurations

My final experiment dealt with "overtraining". Because the Butterworth data set is heavily modal, the weights are seldom modified to account for less frequent values of the dependent variable and this results in a system biased towards predicting the modal value. This can be offset by "overtraining" the system on the less frequent values, so that the system is exposed to roughly equal numbers of cases having each value of the dependent variable. This equal distribution was approximated by running one training cycle when the dependent variables was equal is "0", two cycles when it was "1", and three when it was "2".

This approach has mixed results, as indicated in Table 5.6, which also shows the measures on the conventionally-trained network and discriminant analysis for purposes of comparison. As expected, the overtraining substantially improves the accuracy in the training set: on the *Abate*, *Restrain* and *Settle* variables, the over-trained network does as well or better than discriminant analysis on both the accuracy and ER measures. The improvement over the conventional network training is quite dramatic on the ER measure—for example it is almost twice as high for *Stop*—and performance is consistently higher on the accuracy measure for all variables.

Table 5.6. Measures of Fit with Overtraining

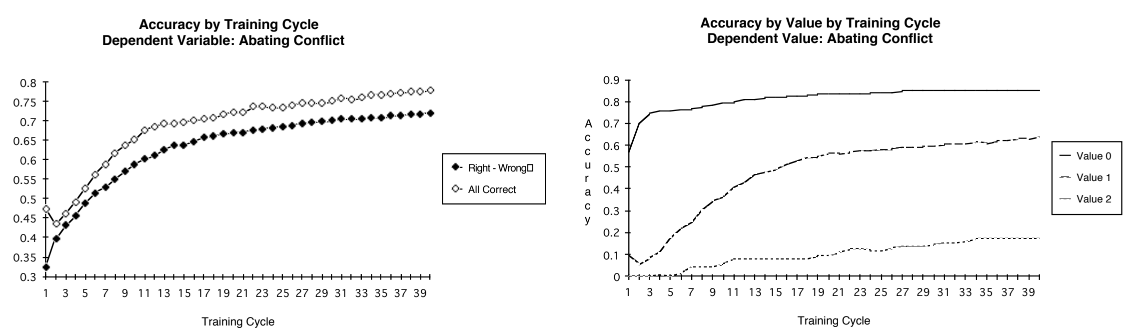
	Dependent Variable				
Statistic	Stop	Abate	Isolate	Restrain	Settle
Accuracy					
Interval NN training	.647	.703	.839	.711	.752
Discriminant training	.918	.760	.931	.767	.779
Overtrained NN training	.805	.769	.859	.806	.836
Interval NN validation	.609	.536	.694	.523	.640
Discriminant validation	.469	.543	.547	.523	.549
Overtrained NN validation	.552	.479	.648	.497	.598
Entropy Ratio					
Interval NN training	.371	.589	.527	.597	.464

Discriminant training	.921	.766	.937	.766	.792
Overtrained NN training	.756	.736	.739	.812	.772
Interval NN validation	.320	.403	.374	.373	.302
Discriminant validation	.364	.499	.469	.490	.487
Overtrained NN validation	.387	.394	.394	.443	.366

The consistently increased accuracy and ER on the training set does not extend to the validation set: the over-trained network has slightly higher ER measures than the conventionally trained network—except for the *Abate* variable, where the measures are about equal—but it is consistently worse in terms of accuracy. The over-trained network is also substantially weaker than discriminant analysis on four out of five of the dependent variables as measured by ER, and worse than discriminant on two out of five variables on accuracy. The overtraining is a two-edged sword: the improvements on the training set do not consistently translate into comparable improvements in the validation set, probably due in part to the over-trained network being less reliant on modal predictions.

Training Patterns

Figures 5.12 and 5.13 show the general pattern of training as a function of training cycles. Figure 5.N7 exhibits a classical learning curve[32]: the only point of interest is that even after 40 training cycles the learning curve has not entirely leveled off, so it might be possible to improve the performance somewhat by substantially longer training times. While neural networks can exhibit an assortment of learning pathologies—notably getting stuck in weight configurations that are local maxima but provide low accuracy—these problems were not encountered in the Butterworth set, and improvements in the performance of the network almost always followed a classical learning curve.



Figures 5.12 and 5.13. Accuracy by Training Cycles

32 These figures give the accuracy on the training set at the *end* of each cycle of training on all cases. The accuracy at cycle 1 is considerably better than chance because weights have been modified after examining each case and the network adjusts to the modal cases even before the first training cycle is completed.

The "Right minus Wrong" statistic is the number of correct output layer values minus the incorrect values; this increases monotonically during the training. "All Correct" is number of cases where all three output neurons had the correct value: 1 for the neuron corresponding to the observed value of the dependent variable and 0 for the other two neurons.

The pattern of learning illustrated in Figure 5.13 is interesting. The system first learns the modal value ("0"), rapidly reaching an asymptotic accuracy. It then begins to learn the second most frequent value ("1"), and as the accuracy of that increases, it very gradually picks up the least frequent value. This is generally characteristic of neural networks—as well as human beings—and would argue against using a neural network to predict unusual cases of political behavior.

Weights and the Significance of Independent Variables

The fact that a neural network stores its knowledge in a diffuse set of weights makes it robust against erroneous input and component failures, but poses problems when one is trying to ascertain the variables most important in determining the behavior of the network. One plausible, if imprecise, measure of variable importance are the weights themselves: *ceteris paribus*, weights that have a large absolute magnitude should have greater influence on the output of the network than the weights with a small magnitude.

The key variables of the Butterworth data set for the neural network were identified by examining the combined weights linking input and output neurons for each of the five dependent variables. The total weight is computed as

$$wi,o = wi,h + wh,o$$

In other words, the sum of the input to hidden layer weight and the hidden to output layer weight. This is an imperfect measure since a high positive weight in one layer could be canceled out by a high negative weight in the other, and the effect of any input to a neuron is moderated by the sigmoid function, but it is a reasonable first approximation.

The choice of heavily-weighted inputs was clearly non-random.[33] Certain variables—for example *Action*, *Agent's Bias*, *Ethnic Conflict* and *Relative Power*—showed up quite frequently across the different dependent variables. Many of these high weights were concentrated on a single neuron in the hidden layer.

Table 5.7 is a comparison of the list of input variables corresponding to the highest ranked weights in the neural network compared to the sets of important variables identified by the discriminant analysis and bootstrapped ID3 models. The variables from the discriminant analysis were those significant at the 0.10 level or higher for each of the dependent variables. The ID3 comparison uses the five variables most frequently used in the bootstrap experiments.

There is a fair amount of commonalty between the important variables identified in the discriminant analysis and neural networks: the percentage of the high weight variables that also have significant discriminant coefficients is consistently about twice the percentage of variables with significant discriminant coefficients as a percentage of the number of independent variables. In other words, the variables with high neural network weights are about twice as likely to be from the list of variables with significant discriminant coefficients than one would expect by chance. The discriminant

[33] Table 4 in Schrodt (1990b) identifies these variables by name. Because the logit analysis used a restricted set of variables determined by the neural network and discriminant results, that technique is not included in this comparison.

Table 5.7. Proportion of High Weight Variables in Common with Significant Independent Variables using Other Techniques

	Dependent Variable				
Statistic	Stop	Abate	Isolate	Restrain	Settle
Signif Discriminant Coefficients	.533	.666	.733	.733	.333
Signif Discriminant Coeffs. as a Proportion of Independent Variables	.250	.333	.417	.333	.450
B/ID3 Variables	.200	.267	.466	.266	.250
B/ID3 Variables without "Action" .000	.000	.266	.066	.200	
B/ID3 Variables as a Proportion of Independent Variables	.208	.208	.208	.208	.208

Discriminant analysis and neural network appear to be using much the same information from the independent variables, particularly when one considers the imprecise character of the summed weights as a measure of the importance of the variable.

This commonalty does not hold true for the ID3 variables. Except for the *Isolate* variable, the level of agreement is only slightly higher than one would expect by chance. Even that level of agreement is due almost entirely to a single variable—*Action*—and when it is removed, there is almost no agreement for three of the five dependent variables, and agreement at only about the level of chance for the remaining two. ID3 is using different information, which is probably unsurprising given the totally different structure of a ID3 classification tree compared to the neural network.

Discussion

Over the past decade neural networks have moved from the periphery of machine learning methods to the mainstream. In 1985, neural networks were too-frequently dismissed with a brief "Minsky and Papert proved they don't work." In 1995, the symbolic methods advocated by Minsky are in eclipse, and neural networks are an option in SPSS. A large amount of research is now being done with neural networks—alone or in combination with other methods such as genetic algorithms or clustering techniques—and the application I have presented here is quite basic in its structure and objectives compared to what could be done with contemporary software.

The neural network demonstrated here was a classifier trained in a "supervised" mode. Neural networks can also be trained in an "unsupervised" mode where the network itself decides what categories are present in the data: for example the SPSS neural network package offers two different methods for doing this. Such clustering could be used, for example, on a set of event sequences to find common patterns—much as I will do in the next chapter using other algorithms—and the neural network would provide the advantage of some associative and error-correction capabilities.

When used to model a set of international political behaviors, a neural network is significantly more accurate than a discriminant model, slightly more accurate than an ID3 model, and comparable to a multinomial logit when studied in split-sample tests. Obviously this does not mean that a neural network will always work better than linear or ID3 models—in particular the heavily modal character of the Butterworth data plays to a strength of neural networks—but neural networks are certainly a contender. Based on the commonalty of variables, I would guess that the neural network, logit and discriminant analysis probably make similar classification errors; ID3 probably makes quite different ones.

Clustering and nearest neighbor methods

A fourth method of machine learning is clustering or nearest neighbor evaluation. To classify using clusters, one takes a large number of cases, places them in a very high dimensional space on the basis of their feature vectors, and then classifies new cases by associating a case with the cluster to which it is closest according to some metric. Discriminant analysis is one such method in common use in political science, but a variety of other clustering methods can also be used.[34]

The clustering and nearest neighbor methods used in machine learning typically use a far greater number of features than comparable approaches in statistical analysis; for example, in a text retrieval system, hundreds or even thousands of features might be used. This dimensionality restricts the techniques that can be employed to create the clusters: Matrix inversion is usually not an option when 2000 variables are under consideration, and iterative methods are used instead. The high number of dimensions also means that a large number of cases are needed to define the clusters, though unlike regression-based approaches, the number of cases need not be larger than the number of variables.

The full dimensionality of the data is usually considered only in the machine representation of the problem. Due to the inability of humans to perceive more than a few dimensions in data, nearest neighbor systems are often presented as if they involved only a small number of dimensions.[35] This simplification is augmented by the frequent empirical tendency for most of the classification power in a large set of features to be found in two or three appropriately-chosen dimensions.

The efficiency of any nearest neighbor method depends heavily on the metric used to determine the distance between cases. It is unusual to find that all of the features are equally important, so an effective classification metric will emphasize some features more than others. For example, discriminant analysis is based a least-squares distance between cases, and "learning" in discriminant analysis involves weighting the features (independent variables) describing a case to maximize the difference between cases in different categories.

---

[34] Aldenderfer and Blashfield 1984; Bailey 1995; Everitt 1980; Michie, Spiegelharter and Taylor 1994 and Fayyad et al 1995 review a variety of clustering methods. Schalkoff (1992,chapters 2-5) is a good technical introduction to the statistical pattern recognition literature and Salton (1989,chapter 10) describes a variety of nearest neighbor methods in the context of automated document retrieval.

[35] In addition to the three spatial dimensions, additional nominal and ordinal dimensions can be displayed using the color, shape and labels of data points; the color and size of points can also express a limited range of interval values. It is difficult to push even these methods beyond about a half-dozen dimensions, however.

Finding a metric is more complicated when nominal features—for example type of government, dominant religion, and membership in specific alliances—are considered. Frequently the features describing a case involve nominal, ordinal and interval variables, so the distance metric must be able to deal with all of these levels of measurement. When a classification problem involves hundreds of features and thousands of cases, computational efficiency also becomes an issue. As a consequence a wide variety of metrics, sometimes customized for specific problem domains, are used in nearest neighbor systems—for example the SPSS hierarchical clustering routine offers almost two-dozen options—although simplifications such as treating nominal and ordinal variables as if they were interval are often disgustingly effective.[36]

While there is often a great deal of overlap between the mathematical techniques used in inferential statistics and those used in nearest neighbor systems, the two approaches are almost opposite in their philosophy. Inferential statistics use a small sample of cases to construct a general model of the population from which those cases were drawn. Clustering, in contrast, employs a large number of training cases to classify a small number of new cases. Nearest neighbor systems are generally descriptive rather than confirmatory—there is no null hypothesis and no significance test.

Properties of Nearest Neighbor Methods

An attractive aspect of nearest neighbor techniques is their conceptual simplicity. Devijver and Kittler (1982,69) note "Roughly speaking, nearest neighbor methods exchange the need to know the underlying distribution for that of knowing a large number of correctly classified patterns."[37] The "training" of a nearest neighbor system consists simply of placing a large number of cases in the multi-dimensional space. New cases are classified by determining their distance to these points according to some metric. The metric can either be decided *a priori* or, more commonly, it is chosen to optimize some characteristic of the space, such as minimizing the distance between cases known to be in the same category.

For example, one could classify a country as friendly or hostile by looking at its political interactions in a high-dimensional space such as that defined by the WEIS or COPDAB event coding schemes. New clusters of interactions over time would reflect changes in the relations between two states. In 1979, the United States found that the behaviors of the Islamic Republic of Iran were located in quite a different part of the space of diplomatic behaviors than those Iran had shown during the Pahlavi regime. The United States rapidly learned to anticipate conflictual rather than cooperative behaviors from Iran. Egypt's actions following the assassination of Anwar Sadat in 1981, in contrast, were generally located in the same regions of the behavioral space as they had been prior to the assassination.

---

[36] This is similar to the situation of using binary dependent variables in regression: while there are lots of theoretical reasons why logit, probit or discriminant analysis should work better than OLS, regression tends to do quite well in actual problems.

[37] Clusters are particularly useful when the phenomenon being studied derives from an evolutionary mechanism where new cases are adapted from old cases (Aldenderfer and Blashfield 1984). For this reason, two of the most fruitful sources of nearest neighbor techniques are biology—e.g. genetics; paleontology—and linguistics.

Petrak, Trappl and Fürnkranz (1994) use clustering to construct a very effective system for identifying sets of similar crises coded in the KOSIMO database. One of the interesting characteristics of this system was that crises tended to cluster geographically—European crises tended to be associated with other European crises, Asian with Asian, African with African and so forth—despite the fact that "region" was only one of about a dozen features used for the classification. The regional clustering occurred because multiple characteristics of the crises—for example superpower involvement and level of economic development—were coassociated with geographical location, and this was reflected in the clustering in the vector space.

If the population of cases being analyzed is changing, the metric used to classify them may need to be modified periodically, particularly when the existing model is producing a number of erroneous or ambiguous classifications. This new metric may in turn establish new clusters. The revised metric constitutes a "re-evaluation of policy" that can change not only the evaluation of future interactions, but also a reassessment of past interactions. The US re-evaluation of Japan as an economic rival rather than a client provides an example of this: Some past Japanese actions that had earlier been interpreted as supportive of US policy—for example Japanese encouragement of market systems in southeast Asia and Africa—were re-interpreted as evidence of Japanese competitiveness.

Nearest neighbor methods also allow for the assessment of certainty, just as humans often assess the confidence with which they know something. A new case being classified can fall into any of the four situations illustrated in Figure 5.14.

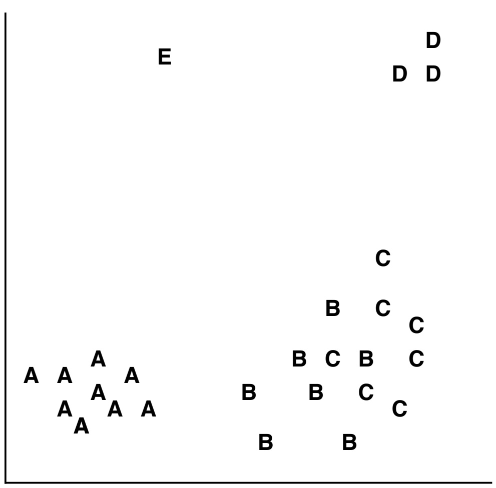


Figure 5.14. Nearest Neighbor Configurations

Cluster A: The case is near a large cluster of cases of a single type.

Interpretation: Classify the case using the type found in this cluster. The classification can be done with a high degree of confidence.

Cluster B&C: The case is near a cluster containing cases of multiple types.

Interpretation: The case is ambiguous and might be either type B or C. The classification could either be a "best guess"—classify according to the type of the nearest or most common points in the cluster—or else a fuzzy classification into more than one category.[38] This is comparable to how a human would evaluate a similarly ambiguous case. A significant number of overlapping clusters indicate that the set of independent variables being used for classification is not sufficient to unambiguously solve the problem.

Cluster D: The case is near a small cluster of cases of a single type.
Interpretation: Classify the case using the type of the nearby cases, but do so with a low degree of confidence.

Point E: The case is distant from all other cases.
Interpretation: This case is distinct from any others in the data set, and therefore cannot be classified with any confidence.

With this approach, classification, ambiguity and confidence are all handled in a single structure. The disadvantage of the method is that a large and representative number of cases are needed before classification can be done with confidence. ID3, in contrast, could correctly classify a large number of cases based on a single example, provided the characteristics of the feature vector uniquely defined the case.

Missing values can be dealt with in nearest neighbor systems in two ways, both of which basically ignore them. Missing values may occur on a dimension of little importance, so they can simply be dropped from the distance calculations without affecting the classification. Second, and more systematically, a case with missing values can be projected onto the space formed by those dimensions where information *is* available and the distance to clusters or centroids computed in that reduced space. The usual set of statistical approximations, such as replacing the missing value with its mean, also frequently work well.

Centroids

As noted in Chapter 3, humans do pattern matching against archetypal, and often hypothetical, cases rather than with actual cases. When one speaks of a “revolution”, the reference is to the general concept of revolution, not a specific crisis. In the nearest neighbor system, these templates or ideals have a straightforward analog in the *centroid* of a cluster of points. The centroid summarizes the information of the cluster, even though it may not correspond to an actual point. For most purposes of classification, a cluster can be replaced by its centroid, which provides considerable economy in the storage and transmission of information.

Consider the problem of defining a “revolution”. This could either be learned by studying a large number of actual revolutions and inducing the general concept (the centroid), or one could be given the location of the centroid that had been determined by

---

[38] In cluster analysis, a numerical measure of the degree of category membership can be easily computed using the distances to nearby points or by tabulating the points within a certain distance of the case. For example, if an overlapping cluster contains 6 cases of type B and 4 of type C, a reasonable fuzzy classification would be 60% probability of category A and 40% probability of category B.

someone else's earlier inductive work. This location could be either an actual revolution close to the center of the cluster or a hypothetical case.

This parallels the technique of teaching political behavior by using a combination of archetypal examples and general principles. In addition to locating the concept in a feature space, the examples provide information on the key dimensions used for classification. Usually such instruction provides explicit counter-examples in order to distinguish the concept being taught from others with which it might be easily confused; this information provides the locations of nearby clusters that might in fact be closer to an ambiguous case. For example, when teaching the elementary concept of revolution, most instructors combine also discuss the counter-examples of *coups d'état* and civil wars.

Constructing a centroid is not trivial, particularly if a complicated non-Euclidean metric is involved. A centroid can always be approximated, however, with an actual case: this is computationally straightforward because one can simply compute the mean distance of a point from all other points of its type and choose the case with the smallest mean distance as the representative. In many metrics, cluster centroids are sensitive to outlying values, so the best centroid may be one constructed from "typical" cases rather than from the complete population of cases.

If a reliable set of centroids can be established, then the computational burden of the nearest neighbor method drops considerably: a new case need by compared only with the centroids rather than with all of the cases. The disadvantage of this approach is that centroids fix the classification knowledge and the emergence of a new cluster at the boundary of two existing clusters might go unnoticed. In a situation of on-going training or prediction, a more conservation classification system might employ a learning scheme such as the following:

```
1. Establish the clusters and metric using the training cases
2. Replace the training cases with their centroids; weight the
   confidence of the centroid by the number of training cases in
   the original cluster.
3. Classify new cases using only centroids.
4. Add new cases to the case list and repeat steps [1] and
   [2]whenever a. Some change occurs in the external environment
   OR b. The system begins to produce a large number of errors
```

Correspondence Analysis

A method variously known as "correspondence analysis" (CA) or "dual scaling analysis" provides a simple illustration of the use of clustering.[39] Briefly, a CA map is an orthogonal projection of a set of points in N-dimensional space to a 2-dimensional plane chosen to minimize the distance between the original points and the projected points as measured with a chi-square metric; Figure 5.15 illustrates this process schematically. The method is quite similar to the singular value decomposition (SVD) methods used in principal components analysis and canonical correlation analysis, and uses much of the same mathematics. CA projects both the cases and the features onto planes that are

---

[39] Greenacre (1984) provides a thorough, if less than totally transparent, discussion of this technique.

scaled so that features tend to be plotted in the same general vicinity as the cases with which they are most strongly associated[40].

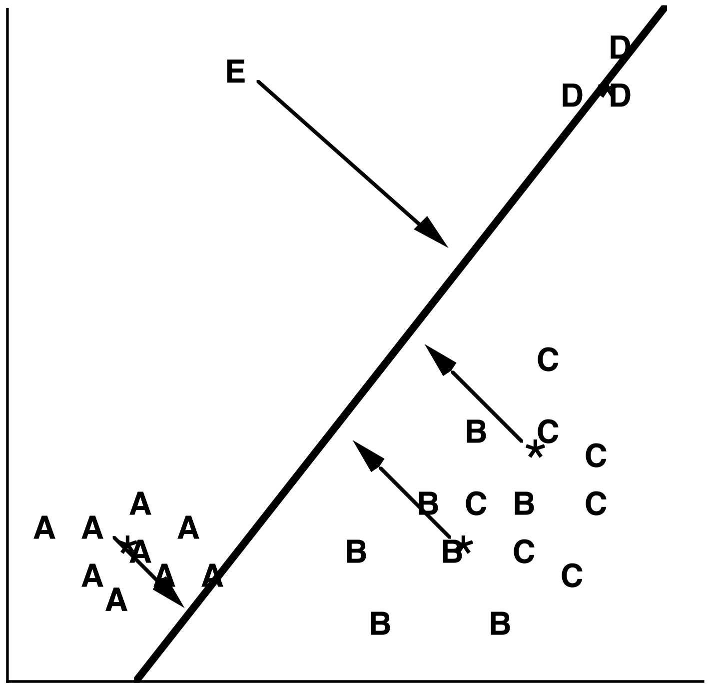


Figure 5.15. Projecting Group Centroids to a Line

The projection of the original set of clusters in Figure 5.15 onto a linear subspace would produce a new clustering similar to that in Figure 5.16. In a typical correspondence analysis, the projection is from a space of much higher dimension into a two-dimensional space, but the geometrical principles are the same.

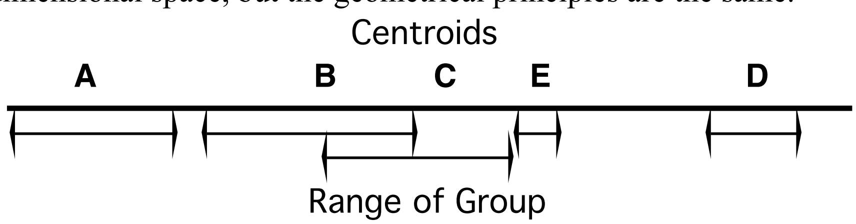


Figure 5.16. Clusters in the Subspace

Correspondence analysis, following the SVD technique, associates an eigenvalue with each dimension. The cumulative proportion of the eigenvalues indicates how much of the total distance between the points in the original space is retained when those points are projected into the subspace, and therefore provides a measure of the "variance explained" by the subspace. The highest eigenvalue is associated with a line (1 dimension) of maximum variance; the two highest eigenvalues are associated with a plane (2 dimensions) and so forth. While most CA maps give only the highest two

---

[40] This spatial association of features and cases is one of the distinguishing features of correspondence analysis but it will not be exploited in the analysis here; I'm only interested in the chi-square metric and the projections.

dimensions, higher dimensional maps are appropriate when the first two dimensions leave much of the variance unexplained.

Much of the early development of correspondence analysis was done by the French linguist, Jean-Pierre Benzécri, who wished to compare languages on the basis of allowable vowel-consonant combinations. Each language was described by a large number of features based on whether a vowel-consonant combination was allowed. Benzécri assumed that similar languages would be located close to each other in this high dimensional space while languages that were different would be located far apart. Because the main role of CA is the reduction of the dimension required to describe the data, the approach is *geometrical*, rather than statistical.[41] In social science terminology, CA is a descriptive, or data reduction technique, though it borrows concepts from statistics and the use of the chi-square metric is obviously motivated by statistical considerations.

Application

Nearest neighbor methods are typically used in classification problems, and the data they use are the familiar rectangular matrix of cases and variables. The example I will use to illustrate the approach is somewhat different: the analysis of two sets of *rules* used to define the characteristics of international systems. These come from Rosecrance's analysis of the European system in *Actions and Reaction in World Politics* (1963) and Kaplan's theoretical systems in *System and Process in International Politics* (1957), which he extends to a total of nine systems in Kaplan (1969). This exercise will demonstrate both the extent to which geometrical clustering can be used to find underlying regularities in a complex set of data, and, in passing, demonstrate an approach to the formal analysis of qualitative sets of rules.

Rosecrance and Kaplan both describe a number of different historical or hypothetical international systems using sets of rules and attributes. Each system can be associated with a feature vector based on those rules and attributes. In the simplest approach—which I apply to the Rosecrance systems—one can use a binary feature vector whose dimension corresponds to the total number of rules in all of the systems. The feature vector for each system will contain a one if the rule is present and a zero otherwise. Alternatively, the vector can describe the degree to which a rule is present in a system; this is used for Kaplan.

With a sufficiently large number of systems, the full dimensionality of the vector space could be used for clustering the systems but because Rosecrance and Kaplan describe only a small number of cases, CA is used instead to reduce the dimensionality to two continuous dimensions. By reducing the dimensionality of the case descriptions to a plane, it is possible to visually examine the spatial relationships among the various systems.

Rosecrance identified nine different European international political systems for the past 250 years.

---

[41] The rallying cry of the French CA approach is "statistics is not probability"; this has a clear intellectual affinity with the exploratory data analysis approach in the United States.

I	Eighteenth Century	1715-1788
II	Revolutionary/Napoleonic	1789-1813
III	Concert of Europe	1814-1822
IV	Truncated Concert	1823-1847
V	German Unification	1848-1871
VI	Bismarckian Concert	1872-1889
VII	Imperialism	1890-1917
VIII	Inter-War	1918-1942
IX	Cold War	1948-1989

I coded each of these for the presence or absence of eleven features identified explicitly by Rosecrance. These features deal primarily with the political objectives of the actors (e.g. status quo, revolutionary, revisionist, conservative) and the presence of dynastic or nationalist actors. The full list of features and the coding of each system are given in Appendix A.

The feature vector places each case in an 11-dimension binary space. Figure 5.17 show the CA projection of this space onto a real-valued plane.[42] By convention, the horizontal dimension corresponds to the highest eigenvalue; the vertical dimension to the next highest. While the clustering is not counter-intuitive and shows a clearly defined cluster in the lower-left corner of the map, there is no immediately obvious interpretation to the two dimensions[43]. The significance of the cluster becomes apparent when one examines the movement of the system over *time*, as illustrated in Figure 5.18: The cluster is a “core regime” that the European system returns to after temporary departures. In Figure 5.18, these departures from the core are indicated with dotted lines; the returns with solid lines. The one exception to this pattern is the transition from the Revolutionary/Napoleonic system to the Concert of Europe, which is not in the core. However, the next transition, to the Truncated Concert, returns to the core.[44]

---

42 The correspondence analysis algorithm described in Greenacre (1984) was implemented using eigenvectors were extracted with the Weilandt routine from Borland International's *Turbo Pascal Numerical Methods Toolbox* software library, Macintosh version 1.0.

43 The plane explains only about 60% of the variance of the system. The third largest dimension—which adds another 15% to the variance—primarily differentiates the Concert of Europe and Cold War systems from the remaining systems. This makes political sense as both were very stable systems established after large wars.

44 I experimented with coding a "post-Cold War" system by setting the Communist feature to zero; this in fact creates a system that is closer to the core, but not dramatically. If Rosecrance's "Secure" feature is also set to one, the movement is greater, with the new system located on a line between the Cold War and Inter-War systems, about two-thirds of the way towards the latter.

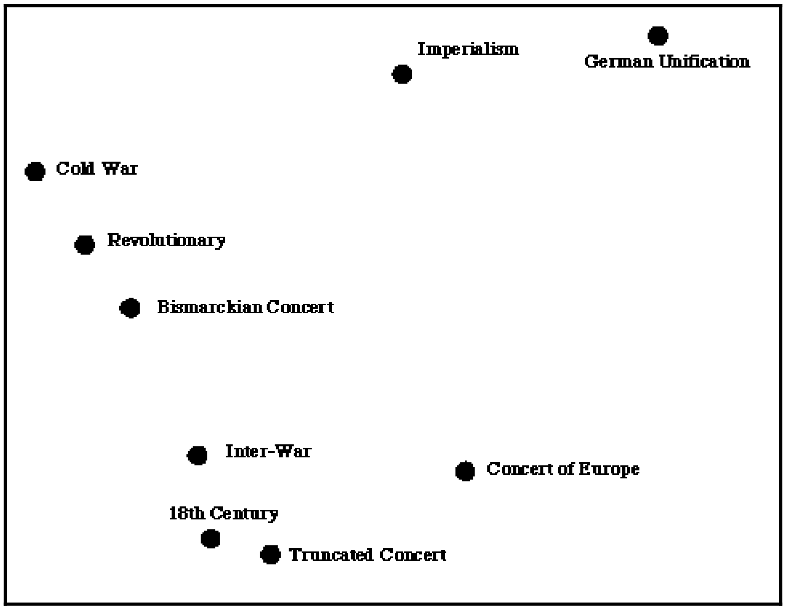


Figure 5.17. Map of the Rosecrance Systems

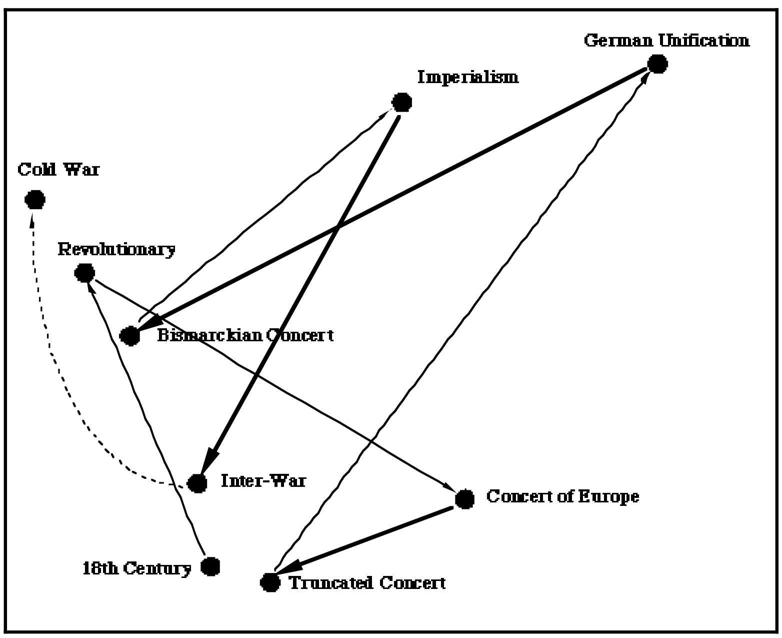


Figure 5.18. Rosecrance System Transitions

This analysis does not necessarily prove the existence of a "core regime" in Europe; it merely shows the presence of such a concept in Rosecrance's presentation of those systems. Because Rosecrance was trying to develop a coherent theory, this finding is perhaps not surprising, but the pattern is certainly less than obvious when one examines only the 11-dimensional vectors.

Kaplan's systems provide another demonstration of this technique. Kaplan's hypothetical systems were coded using a 1 to 9 ordinal scheme (see Appendix A) reflecting the degree of the presence of a feature. The correspondence analysis map of the complete set of Kaplan systems is shown in Figure 5.19. The horizontal dimension is essentially the "tightness" of the system, ranging from the most structured systems on the left to the least structured on the right. The vertical dimension, in contrast, serves only to differentiate the "Balance of Power" system from the remainder of the systems. This configuration should come as little surprise to anyone who has ever studied (or taught) Kaplan, who describes the balance of power system in substantially greater detail than any of the remaining systems. These two dimensions account for about 80% of the variance, substantially more than explained by the first two dimensions in the Rosecrance systems.

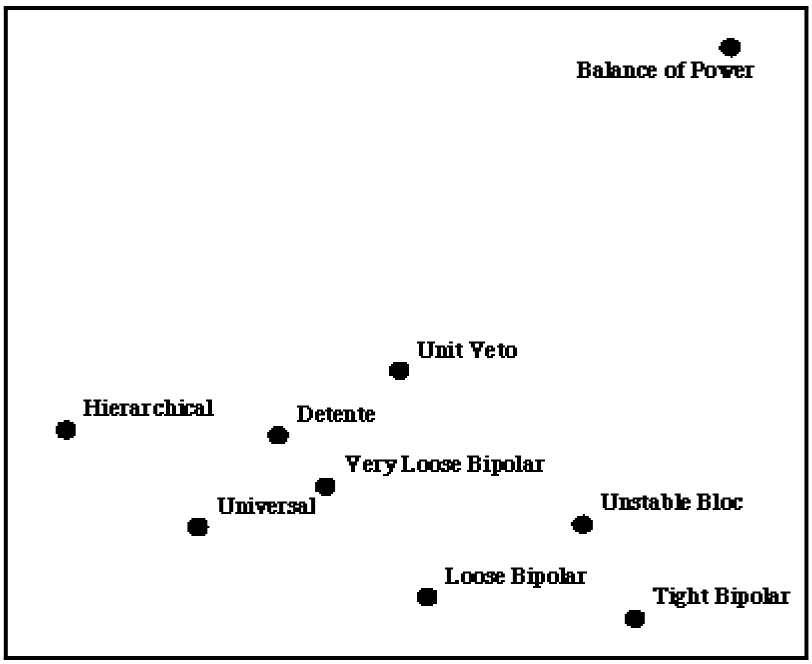


Figure 5.19. Map of the Kaplan Systems

Correspondence analysis, like any linear technique, is affected by outliers. If balance of power system is removed from the analysis, the horizontal dimension stays the same, but the vertical dimension emphasizes a new outlier: the unit veto system. As before, a conventional analysis of Kaplan's models would find the unit veto system—where all states have nuclear weapons and can potentially destroy each other—to be atypical. Removing the unit veto system produces the map in Figure 5.20. The "tightness" of the system is now on the vertical dimension; the interpretation of horizontal dimension is less obvious but is probably related to the level of conflict in the system.

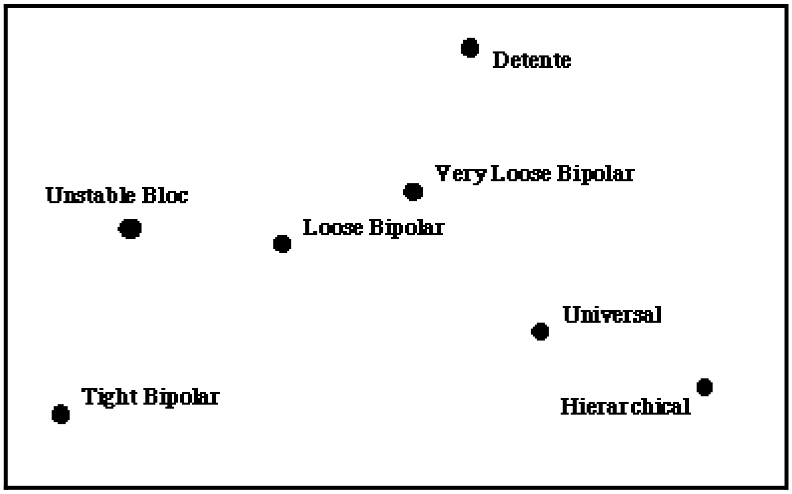


Figure 5.20. Reduced Map of the Kaplan Systems

Discussion

I argued in Chapters 2 and 3 that one of the problems with the behavioral approach was its focus on very simple data structures, typically small numbers of interval-level variables. The classical approach, and human analysts more generally, tend to employ much more complicated data structures, and sets of rules are one particularly salient example of such structures in the study of political behavior. Nearest neighbor methods are dependent on numerical computation and as such probably have little in common with the actual mechanisms of human associative memory. Humans are poor at numerical computation, and they are almost hopeless at the systematic analysis of high-

dimensional vector spaces. However, in the grand tradition of Wigner's observations about the "unreasonable effectiveness of mathematics" (Wigner 1960), nearest neighbor methods often provide a surprisingly effective quick-and-dirty means of exhibiting human-level expertise on many classification problems.

This analysis of the Rosecrance and Kaplan systems is only illustrative, but it demonstrates the possibility of formally analyzing set of rules. Even the fairly simple coding schemes used here produce results consistent with human understanding for Kaplan's systems, and for the Rosecrance system provide insights not immediately obvious from the raw data.

Compared to many nearest neighbor analyses, the Kaplan and Rosecrance systems have very low dimensionality; I suspect that this approach would be even more interesting in systems with a much larger number of rules. To take a hypothetical example, it might be interesting to cluster nation-states on the basis of their rules for dealing with multinational corporations or with environmental protection, as well as looking at the change in those rules over time. This would presumably produce some of the same dimensions a human analyst would see—for example one would expect to find the most important regarding MNCs to be found on dimensions involving market versus planned economies and the level of economic development—but the analysis might also identify clusters and dynamics not evident from the statement of the rules alone.

While my discussion has emphasized the importance of clusters of data points, the regions of the vector space that are *empty* can also provide important information. In an otherwise densely-populated vector space, voids probably indicate feature combinations that are inherently unstable, such as centrally-planned economies with independent labor unions, or military conflict between democracies. In most actual analyses, it may be difficult to distinguish these situations from voids that are just an artifact of a small sample size, but, for example, if points in a time series were observed to occasionally enter a void but never linger in it, one could conclude that the feature combinations within the void were probably socially unstable.

Cluster analysis—alone or in combination with other machine learning and statistical techniques—seems to be emerging as the techniques of choice for knowledge discovery in large data bases (Piatetsky-Shapiro and Frawley 1989; Fayyad et al 1995), also known as “data-mining”. Many clustering methods are memory-intensive and have only recently become practical as computers with tens of megabytes of RAM and gigabyte-sized hard drives have become widely available. While clustering methods have been something of a curiosity in the past, at least in the social sciences, they are likely to see increasing use in the future. To the extent that classification by clustering has characteristics in common with human expert classification—particularly the increment incorporation of knowledge from new cases, the use of centroids to create archetypal cases and the ability to compute measures of confidence—these methods may be useful models for individual and organizational learning and pattern recognition.

APPENDIX: Data Sets

Butterworth Data

The data are from Butterworth's "Interstate Security Conflicts, 1945-1974" study (Butterworth 1976; ICPSR 7586). This data set was originally collected for the purpose of studying the effects of international mediation on political conflict. Briefly, the set consists of

> …postwar conflicts that centrally involved specific power-political aims and demands having direct impacts on national behavior, and that were perceived internationally as being focused on political and security affairs. … Cases are characterized in terms of the goals of parties and management agents … based upon collective statements by international actors, statements by the parties and assessments by expert observers. … As efforts to *manage* these conflicts we included initiatives from any actor (except the parties [to the conflict]) who achieved access to the parties and issues and whose manifest intention was to prevent the conflict from escalating. (Butterworth 1976:3-4)

The data set covers 310 cases of interstate conflict and codes 47 variables dealing with conflict characteristics, actions to manage the conflict and the outcome of that management. The variables used are: number of agents, fatalities, likelihood of abatement, likelihood conflict will disappear, likelihood of spread, likelihood of superpower war, type of war, strategic category, strongest antagonist, power disparity, duration, degree of spread, ethnic conflict, system period, previous involvement, agent's bias, agent's autonomy, phase of intervention, phase of first action, phase of strongest action, leadership, level of agreement, action and relative power.

The Butterworth data reports the effect of the management effort on five outcomes—stopping hostilities, abating the conflict, isolating the conflict, restraining the conflict and settling the conflict. These are treated as dependent variables in the ID3 and neural network analyses. All five variables have the same codes: in the absence of activity by the agent, the outcome of the conflict with respect to the behavior would be

0	No different
1	Somewhat different
2	Very different
9	Variable is inapplicable to this situation

The five dependent variables, the number of valid cases and the distribution of their values is given in the Table 5A.1.

Table 5A.1. Distribution of values by dependent variable

Variable	N	Values		
		< 0 >	< 1 >	< 2 >
Stopping Hostilities	98	65%	26%	9%
Abating the Conflict	192	57%	31%	12%

Isolating the Conflict	106	74%	18%	8%
Restraining the Conflict	191	54%	32%	13%
Settling the Conflict	192	73%	16%	10%

The cases used in the analysis were selected according to two criteria: First, only those cases identified by Butterworth as "Core Cases" (variable 3) were used. This means that each conflict was considered only once; the full data set has multiple entries on some conflicts because multiple agents attempted to manage it. Second, if a dependent variable was coded '9' ("Inapplicable"), the case was not included. Application of these two criteria left between about 100 and 200 cases depending on the dependent variable.

COPDAB

The data used in the test are from a COPDAB daily events set for the period 1948-1978 (see Azar 1980, 1982). The version used was a set made available by Azar to the Merriam Laboratory for Analytical Political Research at the University of Illinois and this is presumably similar to the early versions of ICPSR 7767 COPDAB data set.[45]

Only three dyads are used: USA/UK, USA/France and USA/West Germany. This is the densest part of the COPDAB data and was the focus of the GA analysis because variance in the archives depended on a high frequency of events. Events were coded non-directionally—in other words the USA/UK set included both (USA as actor)/(UK as target) and (UK as actor)/(USA as target) events.

The event sets are coded only for occurrence, not frequency: five instances of an event of code 7 are coded the same as one instance. Since the COPDAB coding scheme has 15 categories, 15 bits can represent the occurrence of events in any period.[46]

The Rosecrance Systems

The Rosecrance systems were coded by using the "Direction" and "Control" variables in Rosecrance's systems diagrams. These are quite unambiguous and therefore a binary coding system can be used. System IX—the Cold War system—is a bit of a problem because it contains several categories not found in the other systems: the new "revisionist" category is coded as the same as "revolutionary"; the system has both

---

[45] The original ICPSR COPDAB, however, was withdrawn for several years for correcting by Azar, who died before completing the work; the corrections were eventually finished by his students and the data used here may or may not resemble those found in the COPDAB set currently at the ICPSR. And meanwhile my copy of the data used in this study are on 160KB, 5-1/4" floppy disks formatted for an Apple II. *This* is why we need a replication archive... (see discussion in *PS* September 1995)

[46]Coding only for occurrence is a first approximation that was used because of the low data density is low; it would be simple to expand the coding scheme to record frequency information by allocating additional bits to each code. For example, if one allocated four bits per code and had a situation where there were three events of type 05, eight of type 06 and twelve of type 07, that part of the string would look like

…… 0 0 1 1 1 0 0 0 1 1 0 0 ……
……[5 5 5 5 6 6 6 6 7 7 7 7]……

"secure" and "insecure" codes so it is coded "insecure"[47]; the "liberal-democratic" and "communist" codes are used.

Table 5A.2. Coding for the Rosecrance Systems

	A	B	C	D	E	F	G	H	I	J	K
18th Cent	1	0	0	0	0	1	0	0	1	0	0
Revolution	1	1	0	0	0	0	1	0	0	0	0
Concert	0	0	1	0	0	0	0	1	1	0	0
Trunc Conc	0	0	1	1	0	1	1	0	1	0	0
German Unif	0	0	0	0	1	0	0	1	0	0	0
Bismarckian	1	0	0	0	0	0	1	0	1	0	0
Imperialism	0	0	0	0	1	0	1	0	0	0	0
Inter-War	1	0	0	1	0	0	1	0	1	0	0
Cold War	1	1	0	0	0	0	1	0	0	1	1

Features:

A	Status Quo	B	Revolutionary	C	Conservative
D	Reform	E	Self-Preservation	F	Dynastic
G	National	H	Dynastic-national	I	Secure
J	Liberal-Democratic	K	Communist		

The Kaplan Systems

Kaplan's system is more ambiguous than those of Rosecrance; as numerous authors have noted, there is less to Kaplan than meets the eye. Kaplan rules are stated in "fuzzy" terminology—systems are compared using terms such as "more", "much more" and so forth—so I used a code of 1 through 9 to describe the degree to which a feature was present. A code of 9 means the characteristic is strongly present; 1 means the characteristic is absent. The designation of the features is somewhat arbitrary but attempts to identify the main elements emphasized in Kaplan's typology.

Table 5A.3. Coding for the Kaplan Systems

	A	B	C	D	E	F	G	H	I	J
Balance of Power	1	1	1	8	1	1	1	1	1	1
Loose Bipolar	6	4	1	2	5	3	3	7	5	4
Tight Bipolar	8	1	1	5	5	3	3	1	7	8
Universal	6	9	5	1	1	3	3	7	5	4

[47] If "secure" is set to 1, the "Cold War" case moves to a point just down and to the right of the "Revolutionary" case; the remainder of the map changes very little.

Hierarchical	3	9	9	1	1	3	3	7	5	1
Unit Veto	2	1	1	2	1	1	9	5	1	1
Very Loose Biplr	3	5	3	3	1	5	4	7	7	4
Detente	3	7	1	3	1	2	4	7	6	1
Unstable Bloc	5	3	1	7	7	7	8	4	7	8

Features:

A	Fixed alliances	B	Universal actor
C	Political system	D	Mediated rather than war
E	Actors eliminated	F	Conflict between major and minor actors
G	Nuclear weapons	H	Neutral states
I	Equality rather than hegemony	J	Conflict between major actors

Chapter 6
Sequence Analysis

> *International conduct, expressed in terms of event data, is the chief dependent variable of international relations research. ... [The] prime intellectual task in the study of international relations is to account for actions and responses of states in international politics by relating these to the purposes of statecraft ... tracing recurring processes within these components, [and] noting systematically the structure and processes of exchange among the components…. Obviously the classical definition of diplomatic history is less ponderous and more literary than the general system definition of the task but both, as we shall seek to show next, carry about the same information and involve nearly the same range of choices of inquiry and analysis.*
>
> *Charles McClelland*

The machine learning techniques discussed in Chapter 5 are very general: they can be, and have been, applied in domains ranging from medical diagnosis to stock market forecasting. This chapter, in contrast, will focus on a problem more specific to the analysis of international politics: event sequences. As discussed in Chapter 3, sequences are one of the primary means by which analysts solve the fundamental problem of short-term prediction in determining the likely consequences of their own policies and the intentions of their opponents. Political analysts have, in associative memory, a large number of sequences acquired through experience and the study of history, and "understand" observed political events when they can match those events to a sequence stored in memory.

If this process of sequence recognition is as important as I've suggested, computational modeling should place a high priority on mimicking it. This, however, is a more difficult process than rule-based modeling because sequence recognition occurs in long-term memory and is a sub-cognitive process. Organizational rules, in contrast, are explicitly articulated and are limited by the information transmission bandwidth available in the organization. While we know that an organization's behavior cannot be modeling entirely by its explicit rules, those rules provide a solid foundation from which to start a model, particularly when judiciously supplemented by tacit information.

Sequence recognition, in contrast, presents problems at every step. First, the amount of information is very large. Foreign policy analysts have a tremendous store of sequence-based information, including formal political knowledge such as the history of the Cold War and the evolution of the Westphalian nation-state system, current information on past activities of individual actors—Saddam Hussein, Boris Yelstin; the differences in Japanese and British foreign policy bureaucracies—and "common sense" knowledge about human behavior, usually learned informally—"hit someone and they won't be happy, and they may hit you back, or if they are smaller than you they may find someone else to hit you back.". The quantity of such information is unclear, but it probably runs to tens of thousands of sequences of varying lengths.

Second, the human brain accesses this information associatively, on the basis of content. As noted in Chapter 3, a typical undergraduate foreign policy exam question such as "List three successful and three unsuccessful uses of economic embargoes by the

USA in the post-WWII period" can be answered quickly and relatively easily by an expert, and part of the training of an analyst involves learning from practice to deal with these types of questions. A computer, in contrast, does not have associative recall: at the basic level, it can find things only by a literal search of its memory, though some methods of organizing memory are more efficient than others. While an organization can make use of the associative memories of the individuals who comprise it, the explicit long-term memory of an organization—typically its standard operating procedures and its physical files—is more like that of a computer than that of the brain. In order to find something in a file, you either have to know where it is already, find it through an index that necessarily emphasizes some types of information at the expense of others, or else you need to look at everything in the file. This, in turn, is why computers have been much more effective in replacing routine organizational work (for example record-keeping) than routine individual work (for example teaching).

Finally, the process by which the matching of sequences is done is largely sub-cognitive. Not only does the process occur relatively quickly without "stopping to think", but even if one does stop to think, one gets only a partial indication of how the matching was done. For example, the question on economic embargoes appears to involve first matching to the general concept of "economic embargo", then classifying those instances into "successful" and "unsuccessful". That classification, in turn, is based on the retrieval of still more historical cases, since it is answered in the context of a "successful compared to what?" in United States history and other history.

When I try to work out the embargo problem, I find I can articulate *some* of the steps. For example, the two cases that immediately come to mind are Cuba and Vietnam, which have been in the news lately.[1] I then started thinking of US enemies during the past 40 years, and check those for embargoes—Libya, Iran and China come to mind, then US pressure on the USSR over the issue of immigration restrictions. I also recalled, directly from the "embargo" stimulus, teaching an article that pointed out that most of the US embargoes were for economic reasons rather than political, but unlike the students who answered this question on the exam in my foreign policy class, I don't have these cases actively in memory and would have to look them up.

This protocol gives some hints as to how my memory is organized, but it is not sufficient to create an algorithm. My description of the process depends heavily on the fact that examples just pop out of memory, in a somewhat unsystematic fashion—for example the current embargo of Serbia did not spontaneously appear, perhaps because it has not been integrated with the other memories—and based on this information, I can then assemble a fairly logical set of cases. Because of these difficulties, the literature on event sequence recognition—as distinct from the large literature on event data analysis (Peterson 1975; McGowan et al 1988; Schrodt 1994)—is substantially more limited than that on rule-based systems or the general-purpose machine learning methods.

Within the AI literature, most of the work on sequences is based on the Schank and Abelson (1977) concept of scripts.[2] While the script model is very useful, this work has

---

[1] A statement that has remained true through several drafts of this manuscript...

[2] An alternative approach is to use complex knowledge structures, rather than sequences, to describe a situation. Mallery, Duffy and Sherman (Mallery and Sherman 1993; Mallery and Hurwicz 1987; Duffy 1991) have done extensive computational modeling using historical precedents described by the highly

been almost entirely deductive: a script is constructed and then finely tuned to match a very limited domain. The best known script from the Schank-Abelson tradition is the “restaurant script”, which handles events that would occur in the process of a visit to a restaurant. Scripts can, however, to extended to less trivial domains; for example Kolodner’s (1984) CYRUS program deals with the official activities of Secretary of State Cyrus Vance and contains script-like structures reflecting the event sequences that a Secretary of State might encounter. With an appropriately refined script, a system can interpret, in considerable detail, the events in a specific domain. Unfortunately, the sequence recognition needed for the analysis of international affairs requires the opposite approach: relatively shallow knowledge over a very broad domain.

The techniques that I develop in this chapter are based on some generalizations of the machine learning techniques discussed in Chapter 5 and they are broad rather than deep. My objective is to begin to develop some machine learning methods that will allow a program to do three things:

- Recognize that two sequences are similar; in other words, simulate the basic sequence recognition function;
- Break a sequence down into its component parts;
- Use sequence similarity and parts to classify sequences into general categories such as war/nonwar.

Because of the complexity of international behavior, a robust system is ultimately going to need to have the ability to learn from example: we cannot afford to write thousands of political equivalents of the restaurant script. The systems discussed here learn by example from event data, but do so at the expense of detail. Before discussing the basic techniques, I will consider the general structural problems involved in political sequences, and I will end the chapter with a discussion of the possibility of using syntactic approaches.

Components of Sequence Recognition

In the simplified approach that I will be using in this chapter, the ideal sequence recognition system would require three components. First, one needs a knowledge representation structure for the sequences themselves. As discussed in Chapter 3, international event data are sufficient for the task, and event data have the additional advantage that they can be coded directly from machine-readable text (see Gerner et al 1994), though the source I will use in this chapter was human-coded. Human sequence recognition in all likelihood tags event sequences with some additional contextual information concerning the national and international environment—the outbreak of the “Soccer War” between El Salvador and Honduras in 1969 is classified differently than the outbreak of the Russo-Afghan war in 1979—but the isolated sequence provides a starting point and the problem of matching contextual information is not likely to be more complex than matching the sequences. Second, one needs a metric that will indicate the

---

structured SHERFACS data set; as noted in Chapter 5, Petrak, Trappl and Fürnkranz (1994) have done the same with the KOSIMO database. .The case-based reasoning literature (see Kolodner 1988, 1993) can provide additional guidance on this, though case-based reasoning systems tend to be very domain-specific.

degree of similarity between two sequences. Finally, one needs to have a very large number of historical sequences in memory.

Given these three components, the sequence recognition problem can be reduced to a nearest-neighbor problem: Take an observed set of events, compute the distance between that sequence and all of the sequences in memory (or a set of archetypal centroids representing general categories of behavior), and classify the sequence using its nearest neighbor. While this provides a general approach, the practical problem of sequence matching is complicated by three factors: events are substitutable; the stream of observed events is composed from parallel sequences; and sequences are partially ordered rather than strictly ordered.

Substitution

As Most and Starr (1984) point out, one of the major problems in finding law-like statements in international politics is "substitutability": actions that appear on the surface to be very different can be equivalent in a political context. For example, the two sequences

> [Nixon uses the phrase "People's Republic of China" rather than "Red China" in a speech]
> [USA/PRC exchange ping-pong teams]
> [USA/PRC negotiate re-establishing relations]
>
> [Nixon uses the phrase "People's Republic of China" rather than "Red China" in a speech]
> [USA invites PRC scientists to conference on earthquakes]
> [USA/PRC negotiate re-establishing relations]

would be seen by most political analysts as equivalent: the USA and PRC had to do something to signal mutual acceptance in 1970; they did so through a friendly table tennis competition but a scientific conference would have accomplished much the same thing. In this context, the two events are equivalent.

While one could not inductively discover the counterfactual equivalence represented in this example, it would be desirable to determine these equivalence *sets* inductively. For example, states have a variety of diplomatic ways of expressing displeasure, including recalling ambassadors for consultation, formally protesting, suspending talks in other arenas (e.g. trade, scientific cooperation), threatening to suspend aid and so forth. Instead of treating each such instance as a separate event sequence, it would be useful to create a general sequence that looks like

Aa Ab [Ax or Ay or Az] Ag Af Ae Af

The Levenshtein metric discussed below is one means of doing this.

Parallel Structures

The international event stream is typically generated by multiple initiatives that are underway simultaneously but which are temporally independent to a large extent. If a state is pursuing two initiatives described by the sequences **A-B-C-D** and **W-X-Y-Z**, the event sequences **A-B-W-X-C-D-Y-Z**, **A-W-B-X-C-Y-D-Z** and **W-X-A-B-C-D-Y-Z** are equally valid manifestations of those two sequences even though the ordering of the events in the sequences are very different.

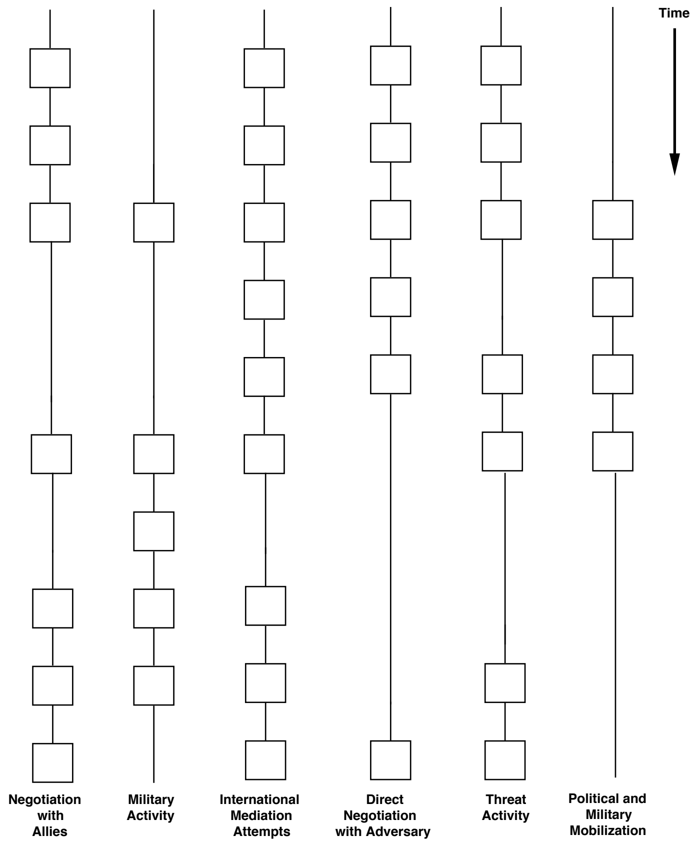


Figure 6.1. Schematic Representation of Parallel Event Sequences

A schematic representation of *parallel* event structures in an international crisis is illustrated in Figure 6.1. The term "parallel" applies in two ways. First, within a crisis there are multiple initiatives underway at any given time. For example, in the months prior to Pearl Harbor, Japan was simultaneously trying to negotiate an accord with the United States while preparing for military actions; had the negotiations succeeded, the military option presumably would not have been pursued on 7 December 1941. Second, if these subsequences are truly characteristic of international behavior, they should occur in multiple historical instances, which means that it should be possible to extract the subsequences from a set of event sequences describing similar classes of behavior.

Partial Ordering

Parts of a sequence often contain pre-conditions: In order for certain events to occur, other events must occur first to prepare the system for the consequent event. However, because an event may have multiple preconditions, and the preconditions can be satisfied in *any* order so long as they occur before the consequent event, the sequence is only partially ordered. This type of structure follows work by Heise (1988a, 1988b) who has developed it for the study of general human interactions, for example ethnographic data, stories or fairy tales.

Figure 6.2 shows a partially-ordered structure for the sequence “Breakfast”:

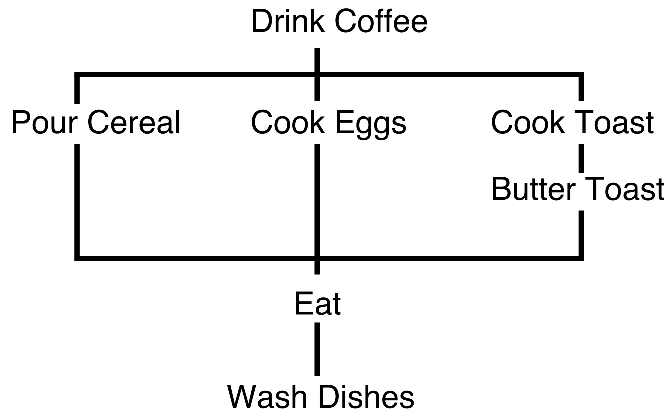


Figure 6.2. A Partially Ordered Event Structure

This would generate any of the following sequences:

```
[Drink coffee, Cook Eggs, Pour Cereal, Cook Toast, Butter Toast,
Eat, Wash Dishes]
```
or

```
[Drink coffee, Cook Toast, Butter Toast, Pour Cereal, Cook Eggs,
Eat, Wash Dishes]
```
or

```
[Drink coffee, Cook Toast, Cook Eggs, Butter Toast, Pour Cereal,
Eat, Wash Dishes]
```

The sequence would not match

```
[Cook Eggs, Pour Cereal, Cook Toast, Butter Toast, Drink coffee,
Eat, Wash Dishes]
```
or

```
[ Drink coffee, Butter Toast, Pour Cereal, Cook Toast, Cook Eggs,
Eat, Wash Dishes]
```

because in each of the latter examples one of the events occurs out of the allowed sequence.

Partial ordering also occurs in international behavior: when states seek to achieve an objective, they initiate an often complex sequence of events that will result, *ceteris paribus*, in the objective. For example, if a state wishes to initiate peace talks, it will contact its allies, the United Nations, initiate trial balloon proposals and so forth. In international relations, these initiatives are often not fully completed because they are interrupted by opposing initiatives by other states or through changes in policy, but the plan was there. Such interruptions also provide a reason to expect that there will be greater short term regularity than long term regularity in international events, since a number of attempted event sequences fail for every sequence that succeeds.

Partial ordering complicates pattern matching against a temporally-ordered event stream such as that found in event data or in newswire reports because it results in permutations being allowed in some events but not others. If one knew the underlying structure, it would be simple to verify that a particular event sequence satisfied that structure, but inductively determining the structure is difficult. Unlike the previous two problems, I will not be demonstrating a solution to partial-ordering, though using a syntactical approach might provide one means of doing this.[3] This asymmetrical situation is similar to that found in computational linguistics: given a grammar and lexicon, it is simple to parse a sentence to determine whether it is grammatically correct but it is exceedingly difficult to inductively extract the rules of a grammar solely from a set of texts.

Levenshtein Learning Algorithm[4]

The Levenshtein metric (Sankoff and Kruskal 1983) is a sequence comparison technique that originated in information theory and is now commonly used to analyze sequences of sound or DNA; Mefford (1984, 1985) proposed using it as a means of sequence comparison in international relations. The Levenshtein metric uses a large matrix of numerical weights to determine the distance between two sequences; these weights can be set, for example, to produce small distances between sequences of the similar type and long distances between sequences of dissimilar type.

I will demonstrate this method using the crises in the Behavioral Correlates of War (BCOW: Leng 1987) event data set; discriminating crises that don't involve war from those that do. The weight will be determined using an example-counterexample machine learning protocol. To learn to discriminate between two classes of objects, the machine is presented with examples from each class and adjusts its knowledge structure—the matrix of Levenshtein weights—on the basis of those examples. Like ID3, the knowledge structure of the Levenshtein metric is sufficiently complex that it can achieve 100% discrimination among the training cases, so it is validated with split-sample testing.

The Levenshtein distance between two sequences is the sum of the weights of the operations needed to convert one sequence into another. If a and b are two sequences $[a_1\ a_2\ a_3\ ...\ a_m]$ and $[b_1\ b_2\ b_3\ ...\ b_n]$, the Levenshtein approach converts one sequence to the other using the operations

Delete an element from a
Insert an element into b
Substitute b_j for a_i

---

[3] I haven't searched for partially-ordered structures for the simple reason that I haven't found an efficient algorithm for doing so. The tricky part is figuring out a way of allowing substitution without the subsequences degenerating to a large number of special cases, as well as limiting any exponentially expanding searches. If this part of the problem could be solved, then adding the more complicated partially ordered structures can probably be done with a few manipulation rules. Algorithms can be found that work on complete data but complications arise when dealing with data that has missing events.

[4] Parts of this section appeared earlier in Schrodt 1991a.

Using the example in Sankoff and Kruskal (1983,11), one could convert the sequence “W A T E R” to “W I N E” by the operations

W A T E R

Substitute I for A

W I T E R

Substitute N for T

W I N E R

Delete R

W I N E

The operations used in computing the Levenshtein distance are those minimizing the sum of the weights. A dynamic programming algorithm for determining this minimum is presented on the next page.

The knowledge structure of a Levenshtein metric lies in the insertion, deletion and substitution weights. Changes in a sequence that reflect important differences should have high weights; those reflecting trivial differences should have low weights. For example, in linguistics, it is clear that as words migrate from language to language, vowels are more likely to change than consonants, and if consonants change, they are likely to change only slightly (an “s” might change to “c” or “z” but probably not “b” or “t”). Thus we see similarities between the English “peace”, French “paix” and Latin “pax”, and similarities between the Hebrew “shalom” and Arabic “salaam”, but see considerable differences between the two groups of words.

The extension of this principle to international event sequences is straightforward (Mefford 1984). Certain international events are quite comparable— for example “mediate” (BCOW code 12142) versus “negotiate” (BCOW 12121)— whereas others are very different—for example “peace settlement” (BCOW 12361) and “continuous military conflict” (BCOW 11533). One would also expect that common events (e.g. the ubiquitous “consult” and “comment” codes in events data) should be inserted and deleted with little cost, whereas rare events such as agreements or the beginning of military conflict would be costly to insert. Two international event sequences would be considered similar if one sequence could be converted to the other using operations that substitute like event for like event; the two sequences would be quite different if they could only be converted by substituting unlike events.

Algorithm for Computing Levenshtein Distances

```
Function Leven_dist(a,b:seq):real;
{ This code follows the Levenshtein distance algorithm described in
Kruskal (1983) }
{a,b are arrays containing the sequences; the 0th element holds the length
of the } {sequence;}
{weight[i,j] gives the insertion, deletion and substitutions weights;}
{dist[i,j] is the matrix used to compute the distance }

var ka,t,r,c        : integer;
    min             : real;
    max_r,max_c     :integer;

begin
  dist[0,0]:=0.0;
  for ka:=1 to a[0] do dist[ka,0]:=dist[ka-1,0] + weight[a[ka],0];
  for ka:=1 to b[0] do dist[0,ka]:=dist[0,ka-1] + weight[0,b[ka]];

{ The code in the "t" loop goes through the matrix starting in the upper
left corner } {then filling by moving down and to the left,ending at the
lower right corner.  r is the } {row, c the column.}

  max_r:=a[0];
  max_c:=b[0];
  ka:=max_r + max_c;
  for t:=2 to ka do begin
    r:=1;
    if t-r<max_c then c:=t-r
                 else begin
                        c:=max_c;
                        r:=t-c;
                      end;
    repeat
     { Determine the operation which adds the minimum to the weight at
each point }
      if dist[r-1,c]<dist[r,c-1] then min:=dist[r-1,c]
                                 else min:=dist[r,c-1];
      if dist[r-1,c-1]<=min then min:=dist[r-1,c-1];
      dist[r,c] := min + weight[a[r],b[c]];
      r:=r+1;
      c:=c-1;
    until (c<1) or (r>max_r);
  end;
  Leven_dist := dist[a[0],b[0]];
end; { Leven_dist }
```

Schrodt (1984, 1985a) reports a feasibility test for using Levenshtein distances to discriminate between general types of dyadic behavior in 1982 using WEIS-coded events (McClelland 1976). Dyads were compared using the distribution of distances from a sample of randomly chosen sequences each containing ten events. The weighting scheme used the fact that two-digit WEIS codes, while technically nominal, are virtually ordinal, so substitution weights were set to the difference between the WEIS codes. Thus the substitution weight of "Force" (WEIS code 22) and "Yield" (WEIS 1) is 21, whereas the substitution of "Force" and "Expel" (WEIS 20) is 3. Insertion and deletion weights were based on the rank order of the frequency of a code: frequent events had low insertion and deletion weights; infrequent events had high weights. While arbitrary and *ad hoc*, this scheme produced plausible differentiation between dyads. For example, the USA/UK and USA/PRC dyads were measured as showing similar behavior; the USA/UK and Iran/Iraq dyads as showing very different behavior.

The clear disadvantage of this approach was the arbitrariness of the weights. Nonetheless, deriving weights for a complex coding scheme on *a priori* theoretical grounds would be difficult: for example, what should be the relationship between BCOW 13551 (Reach Economic Agreement) and BCOW 12641 (Assume Foreign Kingship)? The alternative is induction: weights determined by what one wants to do with the Levenshtein measure itself.

The algorithm demonstrated here is based on the Widrow-Hoff or "delta rule" training method used in training neural networks[5]. The machine is given BCOW cases in two categories: war crises and nonwar crises. The training objective is finding weights that produce small distances between the sequences within each set, and larger distances between sequences in different sets.

To create the weights, the distance between each pair of sequences is computed using the Levenshtein algorithm. Any weights used in computing the distance between a pair of *like* sequences are *decreased* by a small amount. Weights used in computing the distance between *unlike* sequences are *increased* by the same amount. This process is iterated a number of times.

Using this approach, the weights of operations invoked only in the comparison of like sequences are reduced; the weights of operations invoked only in the comparison of unlike sequences are increased; and the weights of operations invoked in comparing both like and unlike sequences remain about the same, since the increase and decrease cancel out. As a consequence, the distances within the groups should decrease, while the distances between the groups should increase. The learning has to be done iteratively since the choice of operations used in computing the Levenshtein distance may change as the weights change, because the Levenshtein algorithm chooses the operations that have the smallest weights.

In the experiments described below, the weights were initialized in two different ways. In frequency-based initialization, I set the insertion and deletion weights to the rank-order of the event frequency in the set of all sequences used to train the system. The most frequent event had a weight of 1, the second most frequent a weight of 2 and so forth. This is consistent with the coding used in Schrodt (1984) and is based on the information theory argument (Pierce 1980) that frequent events have little discriminating value and can be replaced with little cost, whereas rare events should have a higher cost[6]. The substitution cost was initialized as $|r_a - r_b|$, the absolute difference of the ranks of codes a and b. Thus it is less costly to replace a frequent event with another frequent event than it is to replace a frequent event with a less frequent event.

---

[5] The original method was discussed in Widrow and Hoff (1960); Rumelhart *et al* (1986) provide an extensive discussion of variations in the context of neural networks.

[6] No adjustment was made for ties: events tied in frequency were randomly ordered within that tie. The frequency of BCOW events generally follows a rank-size law.

Alternatively, weights were initialized to a constant. These different initializations had no effect on the learning algorithm, though the constant weights proved somewhat less useful for doing discrimination.

The learning scheme produces several peculiarities in terms of the regularities expected of a "distance" (Sankoff and Kruskal 1983:22)—in fact technically speaking, it is not a distance in the mathematical sense—but these cause no interpretive problems in the discrimination test. First, the metric is completely arbitrary, without a zero point, and the "distance" may be negative since many weights become less than zero by progressive subtraction as their operations are repeatedly used in comparing like sequences. This is completely consistent with the substantive interpretation of the weights, since it allows the matching of elements that are important in determining similarity to cancel out mismatches of less important elements. Second, the distance between two identical sequences is not necessarily zero: in fact this tends to be negative because the training algorithm sets the weights for exact matches to negative values.

Finally, unless the weight matrix is symmetric, the distance between A and B is not necessarily the same as the distance between B and A. Because the changes in weights are done while the sequences are compared, a different weight matrix is used to compare A to B than comparing B to A during the training. As a consequence, the weight matrix is not symmetric. There are no substantive excuses for this: it is simply a quirk in the algorithm. These differences seem to get reinforced in the training—note the asymmetries in Table 2—though not so badly as to keep the technique from functioning as intended.

Discriminating War and Nonwar Crises

This system was tested using the BCOW sequences described in the chapter appendix; the short names of the crises (e.g. *pastry*) correspond to the BCOW file identifiers. The training sequences were used to determine the weights; the system was tested with the remaining sequences.[7] Events within a single day are sorted by code so that if identical events occur within a single day in two sequences they will be in the same order in both sequences.

The basic protocol was to run the algorithm on the ten training cases, iterating until the two groups were separated. The resulting weight matrix was then applied to the ten test cases by computing the distance between each test case and the ten training cases. The expectation was that the nonwar cases would, on average, be closer to the nonwar cases in the training set, and similarly for the war cases. The training algorithm showed a monotonic increase in the separation of two groups. This increase in separation was mostly linear with a slight leveling-off that would be expected in a classical learning curve.

---

[7] As noted in the chapter appendix, only the physical events reported in BCOW were used; the sequences were also filtered to eliminate common events and a prefix was added to the BCOW event code to indicate which of five types of dyadic relations were involved in the event.

Table 6.1 gives the results of testing the comparisons with the training cases[8]. The expectation that the Levenshtein distance would discriminate between the war and nonwar crises is fully met, and the discrimination is almost perfect. Only one crisis—Schleswig-Holstein—is not strongly classified into the correct group, though even this case errs only in being almost as close to the nonwar crises as the nonwar crises are from each other; it has the expected negative distances to the other war crises. The war crises cluster strongly, with large negative distances within the group and large positive distances between the groups.

Table 6.1. Distances in Training Set

	rhin	1mor	fash	2mor	bosn	schl	spam	cent	ital	chac
rhine:		330	144	139	311	603	1771	1081	1946	1446
1stmor	190.		-148	189	-275	532	1540	1162	1577	1393
fashod	-288	-77		-177	308	437	1496	754	690	803
2ndm	203	142	-168		351	479	1154	628	1298	972
bosnia	258	-155	307	337		773	-467	662	1493	521
schles	358	432	561	389	540		-2874	-2286	-1301	-1080
spam	1413	1432	1521	1407	-533	-3277		-6399	-5747	-6724
centam	1064	1158	773	613	421	-2371	-6277		-3363	-4641
italet	1903	1781	1586	1648	1594	-1357	-4941	-3145		-3216
chaco:	1184	1154	1028	715	474	-1346	-5076	-3161	-2510	

```
Average within-group distance    =  -1729.4
Average between-group distance   =    987.6
Separation                       =   2717.0
```

Table 6.2 reports the split-sample test using the difference in distances:

Difference = (Average distance to war crises) — (Average distance to nonwar crises)

Results for both frequency-based and constant initializations are reported. As Table 6.2 indicates, the discrimination using the frequency-based weights is perfect: all of the war crises in the target set are closer to the war crises in the training set than to the nonwar crises in the training set; the reverse is true for the nonwar crises. In terms of the rank-order of the distances, the same is true when constant initial weights are used but two of the nonwar crises are actually closer to the war group than to the nonwar group. Despite this, there is still a good differentiation using constant weights: the closest nonwar crisis has a distance difference of -821 whereas the furthest war crisis has a difference of -2932. In this test, frequency based initialization seems to produce a better discrimination matrix than does constant initialization.

---

[8] Table 6.1 was generated with 20 iterations of the training set, constant initial weights set at 10.0, and the weight increment of 1.1.

Table 6.2. Distances in Test Sets

Crisis	Type	War distance minus nonwar distance frequency	constant
palest	War	-2253	-4494
balkan	War	-1264	-4189
bangla	War	-994	-2932
kash1	War	-624	-3685
kash2	War	-557	-3226
munich	Nonwar	21	-708
berair	Nonwar	210	291
pastry	Nonwar	436	-821
anschl	Nonwar	713	316
brtprt	Nonwar	1645	552

Table 6.3 shows the event pairs that had the maximum and minimum changes from the initial values of their substitution weights.[9] Neither group is particularly surprising. The minimum weights, which indicate strong similarity, are usually similar or identical events, particularly those involving military conflict. The maximum weights, which indicate strong dissimilarity, primarily involve substitution of a cooperative action such as consultation or negotiation for a military action such as “Show of Strength” or “Take POWs”. The magnitude of the largest minimum weights is substantially greater—almost by a factor of ten—than the magnitude of the maximum weights. The first digit is a dyad identification code so, for example, the “Mobilization/Mobilization” pair in the maximum weights is the substitution of “mobilization by an ‘other’ actor against one of the principals” for “mobilization by one of the principals against the other principal.” Unsurprisingly, most of the large weights occur in the frequent events, and thousands of substitution weights were never changed from their initial value.[10]

[9] The weights in Tables 6.3 and 6.4 were produced with 51 iterations under the same conditions as Table 6.1. The reported weights were selected from the set of maximum and minimum substitution weights for each event rather than from the maximum and minimum events for the entire table.

[10] There were 171 distinct dyad-prefixed event codes in the sequences, so the total size of the matrix, including the insertion and deletion weights, was 172^2, or 29,584.

Table 6.3. Minimum and Maximum Weights

Minimum Weights

Code	Meaning	Code	Meaning	Weight
111633	Military victory	111633	Military victory	-687.5
111513	Clash	111633	Military victory	-579.6
121143	Change in force	121143	Change in force	-446.6
111663	Take POWs	111523	Attack	-299.1
111533	Continuous conflict	111533	Continuous conflict	-278.3
111523	Attack	111523	Attack	-237.6
211131	Military coordin.	112111	Consult	-162.7
312521	Reach Agreement	411313	Show of Strength	-128.6
111553	No code	512111	Consult	-119.8
112111	Consult	112111	Consult	-111.1
312213	Violate Territory	111313	Show of Strength	-109.9
212521	Reach Agreement	412111	Consult	-103.3
211553	No code	111353	Mobilization	-103.3
312142	Mediate	111533	Continuous conflict	-94.5
112631	Attend Internatnl Event	123151	Change Trade	-90.1

Maximum Weights

Code	Meaning	Code	Meaning	Weight
512111	Consult	111313	Show of Strength	74.9
212521	Reach Agreement	111663	Take POWs	70.5
112121	Negotiate	512111	Consult	69.4
111663	Take POWs	112121	Negotiate	69.4
114113	Subversion	111313	Show of Strength	69.4
112213	Violate Internatnl Law	111313	Show of Strength	68.3
212111	Consult	111633	Military victory	67.2
111353	Mobilization	311353	Mobilization	67.2
312173	Expel Foreign Rep	112213	Violate Internatnl Law	66.1
321133	Change Force Level	512111	Consult	63.9
311313	Show of Strength	512111	Consult	60.6
112521	Reach Agreement	514143	Assassinate	60.6
111443	Military Intrusion	111663	Take POWs	59.5
111523	Attack	112111	Consult	58.4
111513	Clash	112111	Consult	57.3

Table 6.4 reports the insertion and deletion weights for the most frequent events. These tend to be symmetric for insertion and deletion, inversely proportional to their rank orderings and substantially higher than the substitution weights. This pattern holds whether frequency-based or constant initial weights are used, and is opposite from the expectation in information theory that frequent events would have the smallest insertion and deletion weights. These high values imply that at some point in the training process, insertions and deletions were being used frequently—otherwise they would not have the high values—but their high values relative to the substitution weights would lead one to expect that eventually the sequence comparisons would be dominated by substitutions. A few codes showed negative "eights, usually on the order of 10 to 100, but almost all of the insertion and deletion weights were positive.

Table 6.4. Insertion and Deletion Weights

		Weights	
Code	Meaning	Delete	Insert
212111	Consult	2608.1	2638.9
312111	Consult	2393.6	2545.4
512111	Consult	1548.8	1951.4
111313	Show of Strength	2208.8	2077.9
111353	Mobilization	1763.3	2124.1
111633	Military victory	2083.4	1675.3
112111	Consult	862.4	652.3
111523	Attack	1180.3	1020.8
311313	Show of Strength	597.3	108.8
112121	Negotiate	731.5	735.9
112521	Reach Agreement	994.4	906.4
111653	Occupation	895.4	914.1
111513	Clash	1298.0	1285.9
212121	Negotiate	789.8	699.6
111663	Take POWs	902.0	412.5

Discrimination between multiple cases

The experiment above discriminated between only two categories, war and no war. Human decision-makers discriminate between a greater number of categories, so the obvious question is whether a single Levenshtein matrix can be used to handle multiple discrimination.

In an early article using the BCOW crisis set, Gochman and Leng (1983) classify crises into four different categories of bargaining behavior: "fight", "resistance", "standoff" and "prudence". There is sufficient overlap between the crises I am analyzing and the Gochman and Leng set that a simple test can be done of that discrimination. The following crises were used:

Fight:	*balkan* (twice)*, chaco" kash2*
Resistance:	*brtprt, spam, bosnia, italet*
Standoff:	*bangla, 1stmor, fashoda, centam*
Prudence:	*rhine, 2ndmor, schles, kash1*

This is an imperfect test of the Gochman-Leng categories in at least two respects. First, Gochman and Leng base their characterization on all of the behavior in the crisis, whereas I am looking only at the physical behavior. Second, the BCOW files do not correspond exactly to the crises discussed by Gochman and Leng: *chaco* contains both the 1928-29 dispute, which is classified as “fight” and the 1932 dispute, which is classified as "resistance"; *kash2* is a superset of the 1965 Rann of Kutch dispute which is classified as “fight”.

The algorithm used to handle multiple classifications is identical to that used for the binary classification: weights are reduced in comparisons of similar cases and increased for dissimilar cases. The number of cases in each set have to be identical since otherwise weights that are used to match sequences within a large set will be “educed simply by virtue of their being subject to a greater number of opportunities for reduction. To equalize the cases in each set, the *balkan* set was duplicated in the “fight” category; the unusually short *pastry* case was eliminated from “standoff”, and the *ansch* and *munich* cases were eliminated from the “prudence” set.

Table 6.5 shows the results of this after 31 iterations. The Gochman-Leng categories are generally differentiated, though the results are less than spectacular. The "resistance" category appears to be the unusual case: it is the category most distant from all of the other groups, and is unique among the groups in not having its within-group distance being less than the between-group distances. The “standoff” and “prudence” categories are clearly discriminated from the “fight” and “resistance” categories, which may reflect the fact that these involve acts of violence that appear as physical events. Letting the algorithm run for 81 iterations produced virtually no additional changes[11]: the distances expanded by an average of 25% but the relative distances between the pairs of groups remained the same, and the average distance from “resistance” to “standoff” and “prudence” was still less than the average distance within “resistance”. These failures are, perversely, reassuring since they indicate some falsifiability to the method: it is not capable of differentiating any grouping through pure brute force, even with a very large number of iterations.

Table 6.5. Discriminating Multiple Groupings: Average Distance between Crises by Group

	Fight	*Resistance*	*Standoff*	*Prudence*
Fight	1967.952			
Resistance	2708.714	2562.451		
Standoff	2158.901	2206.600	1461.201	
Prudence	2018.164	2067.513	1490.401	1266.701

This is merely a feasibility test but shows a potential for doing multiple discrimination of categories of sequences using a single Levenshtein matrix, much as neural networks are able to do in a single matrix of weights. The training in a multiple discrimination case takes considerably longer than that required for a single

11 The only exception was that the rank order of the distance of "fight" and "resistance" from "standoff" reversed.

discrimination case: 27 iterations are required before the final discrimination pattern stabilizes; the binary discrimination problem took only about 15 iterations to achieve a comparable level of separation.

The Levenshtein learning algorithm is clearly only a first step in the larger puzzle of learning to deal with international events as sequences. The strength of the approach lies in its inductive nature. There are clearly simpler rules for distinguishing BCOW war and nonwar crises: looking for codes involving military conflict is the most obvious. But in order to construct those simpler rules, one must first know that distinguishing characteristic; in a sense, one must already know the answer. An inductive learning algorithm does not need to know the answer; it can find the answer. The system did not know, *a priori*, the importance of the BCOW codes designating military conflict: it discovered them. If machine learning systems can discover those distinctions, they may be capable of discovering things that are not so obvious

If one had a very large set of sequences, it would be useful to find an archetypal centroid for each category. Because of the complexity of the computations involved in determining a Levenshtein distance, this cannot be done analytically, but it probably could be easily done with a genetic algorithm. The GA would start with the population of sequences in a cluster, then evaluate these by their average distance to all of the other sequences in the cluster. Recombination—for example of one crises that is typical in its early phases with another typical in its later phase—and a bit of mutation should work to produce an archetypal sequence near the center of the cluster.

Parallel Event Sequences[12]

As noted earlier, one of the problems involved in interpreting a stream of events such as those found in an event data set or newswire feed is the fact that these events are the result of multiple, parallel political initiatives. This section extends earlier work by Bennett and Schrodt (1987) that used a subset of ”000 WEIS events involving Middle East states and the North Atlantic major powers to construct common subsequences on the basis of nondirected dyad pairs (e.g. USA × USSR) using two-digit WEIS codes. These subsequences were constructed by first scanning the event sequences for the most common 2-event subsequences, then using a fraction of those to construct 3-event subsequences, then using a fraction of those to construct 4-event subsequences and so forth.

The subsequences found by this technique were very successful at covering the WEIS sequences: as a general benchmark, a set of 10 4-event subsequences could account for about 35% of the data. However, these common subsequences were very repetitive and concentrated heavily on the most common events found in this subset of WEIS: uses of force, accusations and agreements. The system discussed here modifies that earlier work by looking explicitly for event subsequences found in multiple crises coded in the BCOW data set. The BCOW data are denser and more varied than the WEIS data, and filtering is used to eliminate the common events.

The four subsets of the BCOW crises were analyzed; these are listed in Table 7A.2 in the appendix. The "Threats" set contains crises that did not result in war because one

---

12 Parts of this section appeared earlier in Schrodt 1990a.

side backed down; the two "War" sets contain crises that involved wars; and the "Mixed" set contains five nonwar crises and five wars. Common subsequences within each of these sets will be determined first, then those subsequences will be used to differentiate the different categories of crises.

Algorithm

The algorithm used to construct subsequences is a fairly simple search that focuses first on finding event codes common to as many of the target sequences as possible, then minimizing the distance between the consecutive events in a subsequence. When a subsequence has been determined, it is eliminated from all of the target sequences where it occurs, then the remaining events in the target sequences are searched for additional subsequences. The algorithm is given below in pseudo-code.

Algorithm for Finding Parallel Event Sequences

```
1. Filter and recode the BCOW sequences (see Appendix)

REPEAT
1. Set the current point in each sequence to th" beginning;
   set the subsequence to the null string
    REPEAT
    1.Evaluate each possible event and each sequence and
      select the event E' which:
        a. maximizes the number of occurrences in the target
           sequences
        b. minimizes the average distance between the current
           point and the next occurrence of the event subject
           to (a).
            [Events which have already been eliminated by
           previous subsequences are not counted in the distance.
    2.Add E' to the subsequence being constructed
    3.If E' is in a sequence, reset the current point  of that
        sequence to the location of E'.
    UNTIL  there is no event which occurs beyond the current
            point in at least a fixed number (3) of the sequences

2. Record the subsequence;

3. With each sequence, eliminate all of the events which have been
   matched by the subsequence.  A subsequence can be applied
   multiple times until less than half of its events occur in the
   sequence

UNTIL size of subsequence is less than or equal to a fixed number
(4)
```

To allow for the possibility of a non-reported (or non-occurring)'event in an otherwise complete sequence, the algorithm does not insist on the perfect matching of a subsequence. The multiple elimination of subsequences allows subsequence to be repeated, for example, when a negotiation is broken off and then reinitiated, or two short periods of hostilities occur. The core of the algorithm is the coverage-maximizing/distance-minimizing search; the remaining idiosyncratic features such as multiple subsequence elimination provide some additional coverage and change slightly the resulting subsequences but are not of critical importance.

The algorithm runs quite quickly because it is deterministically constructing subsequences rather than using a nested (i.e. exponentially expanding) search or using random experimentation. It is also quite short, about 500 lines of Pascal. The number of event codes common across the target sequences tends to be around “00 in each of the sets, so the time required to find the subsequences ”s generally a linear function of the total length of the target sequences.

Results

The subsequences found in each of the four data sets are listed in Table 6.6; Table 6.7 shows an example of how the subsequences nest within the original sequences. All subsequences that contained four or more events and were found in at least three of the target sequences are listed in Table 6.6.[13] The table presents both the 6-digit code (dyad type + 5-digit BCOW event code) and the BCOW description of the event. Note that in many cases events with the same BCOW code refer to different dyad types: for example in subsequence E in the “War2” set there are three “Reach Agreement” events prior to the “Clash” but these agreements are with “Other” parties, not between the two sides, and quite likely involve consultation with supporters prior to initiating conflict. Similarly many of the frequent “Consult” codes are not consultations between the sides of the dispute but with others or between others. A 000000 code in Table 6.7 indicates an event which occurred only once in the set and had been recoded to zero to save storage.

---

[13]The three-event subsequences War1-D and War2-G were found because the algorithm terminated when it could only find a subsequence less than or equal to 4 events in length, and the two War sets have no 4-event subsequences. These are listed in Table 6.6 because they were used when computing the coverage statistics.

Table 6.6. Parallel Subsequences

Threat Data Set

A. 111313 111313 111313 111313 312111 112111 212111
Show of Strength :: Show of Strength :: Show of Strength :: Show of Strength :: Consult ::Consult :: Consult

B. 212111 312111 111353 121133 311313 212111 312111 212111 112521 312111
Consult :: Consult :: Mobilization :: Change Force Level :: Show of Strength :: Consult :: Consult :: Consult :: Reach Agreement:: Consult

C. 112121 111333 212121 112111 212521
Negotiate :: Alert :: Negotiate :: Consult :: Reach Agreement

D. 512111 112213 114213 123151 512111
Consult :: Unknown* :: Antiforeign demonstration ::
Change in trade relations :: Consult

E. 212521 312521 111653 112111
Reach Agreement :: Reach Agreement :: Occupation :: Consult

War1 Data Set

A. 112521 212111 212521 112521 112121 112121 212111
Reach Agreement :: Consult :: Reach Agreement :: Reach Agreement :: Negotiate :: Negotiate:: Consult

B. 111353 312111 111523 311313 512111 512111
Mobilization :: Consult :: Attack :: Show of Strength :: Consult :: Consult

C. 111523 111523 111533 111533 311353
Attack :: Attack :: Continuous Military Conflict :: Continuous Military Conflict :: Mobilization

D. 212111 511313 512521
Consult :: Show of Strength :: Reach Agreement

War2 Data Set

A. 512111 212111 312111 212111 312111 212111 312111 212111 312111 212521 “212111 312111 2”2111 111523 312111
Consult :: Consult :: Consult :: Consult :: Consult :: Consult :: Consult :: Consult :: Consult :: Reach Agreement :: Consult :: Consult :: Consult :: Attack :: Consult

B. 111333 111313 111533 121133 111633 111513 112213 121143
Alert :: Show of Strength :: Continuous Military Conflict :: Change Force Level :: Military Victory (partial) :: Clash :: Unknown* :: Change in Combat Force Level

C. 114123 111523 212111 111513 512111 512111
Discrete Attack :: Attack :: Consult :: Clash :: Consult :: Consult

Table 6.6. continued. Parallel Subsequences

D. 112121 112121 112521 114213 311313 311313
Negotiate :: Negotiate :: Reach Agreement :: Antiforeign demonstration :: Show of Strength :: Show of Strength

E. 212521 312521 312521 111513 111513 111313 111523
Reach Agreement :: Reach Agreement :: Reach Agreement :: Clash :: Clash :: Show of Strength :: Attack

F. 312111 321111 112111 111633 111533
Consult :: Military Grant :: Consult :: Military Victory (partial) :: Continuous Military Conflict

G. 512521 111523 111313
Reach Agreement :: Attack :: Show of Strength

Mixed Data Set

A. 112521 311313 111523 212111 212111 312111
Reach Agreement :: Show of Strength :: Attack :: Consult :: Consult :: Consult

B. 212111 312111 111313 112121 112111 111313 512111 112121 111353
Consult :: Consult :: Show of Strength :: Negotiate :: Consult :: Show of Strength :: Consult :: Negotiate :: Mobilization

C. 111353 311313 111663 212111 111533 112521
Mobilization :: Show of Strength :: Take POWs :: Consult :: Continuous Military Conflict :: Reach Agreement

D. 512111 114213 212521 112213 111653
Consult :: Antiforeign demonstration :: Reach Agreement :: Unknown* :: Occupation

E. 112121 212121 111513 212121 212521 312521 312111
Negotiate :: Negotiate :: Clash :: Negotiate :: Reach Agreement :: Reach Agreement :: Consult

F. 111523 412111 312121 111523
Attack :: Consult :: Negotiate :: Attack

*"Unknown" corresponds to code 12213, which is in the data but not the codebook; it may be "Violate territory", which the codebook states is 12223

Table 6.7. Subsequence Positions within Sequences

	pastry	*1stmor*	*fashod*	*2ndmor*	*bosnia*	*brtprt*	*anschl*	*rhine*	*munich*
1:	AAAAAA	EEEEEE	312173	AAAAAA	000000	BBBBBB	000000	321133	41211
2:	111523	EEEEEE	312173	111443	CCCCCC	AAAAAA	BBBBBB	BBBBBB	BBBBBB
3:	BBBBBB	000000	112521	000000	112521	114223	BBBBBB	AAAAAA	AAAAAA
4:	BBBBBB	BBBBBB	000000	CCCCCC	123151	DDDDDD	DDDDDD	412111	41211
5:	BBBBBB	BBBBBB	132143	CCCCCC	311353	DDDDDD	114213	BBBBBB	121133
6:	DDDDDD	AAAAAA	111521	EEEEEE	111353	114223	DDDDDD	BBBBBB	AAAAAA
7:	DDDDDD	212161	332143	BBBBBB	BBBBBB	CCCCCC	114251	AAAAAA	121133
8:	000000	112631	132143	AAAAAA	AAAAAA	AAAAAA	214251	412521	112111
9:	111433	AAAAAA	111553	411313	000000	BBBBBB	DDDDDD	000000	111333
10:	414113	321121	EEEEEE	BBBBBB	123151	114213	212213	AAAAAA	112183
11:	112161	BBBBBB	EEEEEE	412111	DDDDDD	000000	112161	DDDDDD	AAAAAA
12:	111433	511313	132143	112121	111353	314“51	11”213	412111	412111
13:	DDDDDD	“121”1	000000	312213	321121	212213	114251	CCCCCC	121133
14:	BBBBBB	512121	000000	411313	412521	112631	CCCCCC	BBBBBB	221133
15:	BBBBBB	BBBBBB	AAAAAA	000000	111353	000000	000000	111443	121133
16:	000000	AAAAAA	112521	AAAAAA	CCCCCC	314123	112521	112183	112111
17:	CCCCCC	AAAAAA	112183	CCCCCC	CCCCCC	000000	212111	AAAAAA	114151
18:	“BBBB”B	DDDDDD	332143	312121	312121		312111	BBBBBB	412111
19:	000000	212161	EEEEEE	AAAAAA	000000		AAAAAA	000000	121133
20:	111993	512521	AAAAAA	AAAAAA	311353		000000	CCCCCC	114251
21:	112152	CCCCCC	132143	BBBBBB	EEEEEE		DDDDDD	312521	114223
22:	112363	AAAAAA	AAAAAA	CCCCCC	EEEEEE		114251	212111	321121
23:	DDDDDD	000000	BBBBBB	CCCCCC	000000		AAAAAA	312111	412111
24:	BBBBBB	312631	CCCCCC	312121	111353		221133	AAAAAA	114251
25:	111521	AAAAAA	AAAAAA	DDDDDD	311313		112121	DDDDDD	412521
26:	112152	BBBBBB	000000	112121	EEEEEE		DDDDDD	000000	414151
27:	111521	EEEEEE	BBBBBB	BBBBBB	CCCCCC		DDDDDD	000000	114251
28:	000000	AAAAAA	DDDDDD	AAAAAA	EEEEEE		DDDDDD	212111	DDDDDD
29:	AAAAAA	BBBBBB	111553	BBBBBB	312121		212111	312111	AAAAAA
30:	BBBBBB	CCCCCC	AAAAAA	311333	211313		312111	000000	BB“BBB
31:	11”173	000000	112173	000000	AAAAAA		AAAAAA	512111	412111
32:	414113	000000	111353	AAAAAA	114151		BBBBBB	212111	121133
33:	000000	112631		112121	AAAAAA		112183	312111	DDDDDD
34:	112173	BBBBBB		BBBBBB	BBBBBB		000000	212111	112111
35:	AAAAAA	BBBBBB		111993	412111		000000	312111	414151
36:	AAAAAA	512213		000000	511313		CCCCCC	212111	114251
37:	111443	BBBBBB		112213	211313		112161	312111	AAAAAA
38:	AAAAAA	BBBBBB		BBBBBB	BBBBBB		000000	212111	BBBBBB
39:	AAAAAA	321133			BBBBBB		114251	312111	BBBBBB
40:	000000	312631			312161		EEEEEE	512111	111333
41:	111523	CCCCCC			000000		311333	512111	BBBBBB
42:	EEEEEE	BBBBBB			CCCCCC			000000	111353
43:	111521	BBBBBB			112521			212111	000000
44:	000000	112631			212111			512111	000000
45:	112161	DDDDDD			312111			312111	121133
46:	112363	000000			CCCCCC			512111	112111
47:		212161			211313			BBBBBB	114251
48:		112111			321133			211353	112111
49:		BBBBBB			211313			000000	114151
50:		BBBBBB			212111			212111	412111
	pastry	*1stmor*	*fashod*	*2ndmor*	*bosnia*	*brtprt*	*anschl*	*rhine*	*munich*

The subsequences generally speak for themselves: they are plausible and they are clearly capturing more than random event frequencies. There are clear differences, for example, between the "Threat" subsequences and the two sets of "War" subsequences. Similarly, there are also clear differences between the pre-WWI and post-WWI war subsequences: the extensive communication and negotiation that accompany modern wars is evident in subsequences A and D.

What is surprising—though consistent with the underlying theory—is that many of the subsequences have a degree of internal consistency. For example, in "War1", subsequence A deals largely with consultation and reaching agreements between the sides; subsequence C is primarily military activity; in War2 subsequence A is extensive international consultation, subsequence B is the main sequence of military action, and subsequence E is international agreements followed by initial hostilities. The only feature within the algorithm that might bias the selection of the subsequences to showing this internal consistency was the sorting of event codes within days, but that process seems unlikely to fully account for the consistency because the sorted codes were the frequency-recoded dyad-prefixed integers, not the original BCOW codes, and sorting applied only to multiple events in a single day. Beyond that, the internal consistency exhibited by the subsequences is purely a product of the data and is evidence that we are actually seeing repeated patterns of events. The plausibility of the subsequences is not perfect—in particular the "Mobilization" event occurs at some rather odd places—but is still striking considering the subsequences were produced by a machine with no preconceived biases for which events should be associated together.

Degree of Fit

Table 6.8 reports the degree of fit, or coverage, of each of the sequences by the set of subsequences. The measure reported is

$$\text{Fit} = \frac{\text{number of events matched by subsequences}}{\text{L}}$$

where L = minimum(length of sequence, total length of subsequences). A fit greater than one indicates that some target sequences were matched multiple times by the subsequences.

Table 6.8. Measures of Fit by Sequence

Threat

pastry	*1stmor*	*fashod*	*2ndmor*	*bosnia*	*brtprt*	*anschl*	*rhine*	*munich*
0.6452	1.5161	0.3871	0.6774	1.4839	0.4118	0.5161	0.8710	1.3226

Total coverage = 0.428
Random coverage = 0.336

War1

schles	*rustrk*	*spam*	*centam*	*balkan*	*chaco*
0.8095	2.6190	1.3333	0.7619	1.4762	1.9048

Total coverage = 0.204
Random coverage = 0.177

War2

italet	*kash1*	*suez*	*sixday*	*bangla*	*kash2*	*palest*
1.9200	0.5800	1.1600	1.1400	1.1200	1.1000	1.7600

Total coverage = 0.368
Random coverage = 0.361

Mixed

pastry	*1stmor*	*fashod*	*2ndmor*	*bosnia*	*schles*	*spam*	*centam*	*balkan*	*chaco*
0.5676	0.7568	0.3125	0.4865	1.4324	0.5676	1.3514	0.4595	1.1892	1.2973

Total coverage = 0.351
Random coverage = 0.272

The Total Coverage measure is the total number of events matched in the data set divided by the total length of the data set. Except for the "War1" set (20%), this figure is in the 35% - 40% range. This is consistent with the WEIS results in Bennett and Schrodt (1987)—which had around 35% coverage—despite the use of a completely different data set and a somewhat different sequence construction method. As in the earlier research, the total length of the subsequences is substantially less than the total length of the target sequences, so the subsequences provide a substantial reduction in the amount of information required to describe the target sequences.

As always, it is useful to gauge the extent to which these results are accounted for by pattern rather than chance. The "null model" for event sequences is less obvious than that found in parametric statistics and one could suggest at least four different null models, of decreasing randomness. In each case, a set of null sequences of the same length as the observed sequences would be constructed; the difference is how the probability of an event occurring in a sequence would be determined:

1. Equal probability

2. Probability equal to the probability of the event occurring in BCOW

3. Probability equal to the probability of the event occurring in the set of sequences

4. Probability equal to the probability of the event occurring in each sequence within the set

These models include progressively more information about the characteristics of the sequences being studied. The strongest test is criterion [4]: it can be simulated by simply shuffling the events in a sequence and applying the subsequences to the shuffled sequences. The total coverage of the shuffled sequences is reported as "Random Coverage" in Table 6.8. The subsequences cover about 30% more of the actual sequences than the random sequences in the "Threat" and "Mixed" sets, but provide only 15% additional coverage in "War1" and almost no additional coverage in "War2". This last result was surprising and indicates that most of the regularity in "War2" is accounted for by the marginal frequencies of the events rather than the sequencing of events. The "War2" sequences tend to be longer and show a higher amount of repetition (particularly international consultations and agreements) than the other sets, which may account for the difference.

I also did some partial tests of the algorithm against random sequences created from "Threat" and "Mixed" according to criterion [3]—which is equivalent to shuffling events between sequences in a data set as well as within the sequences—and contrary to initial expectations found the random coverage to be somewhat greater than the criterion [4] random sequences—0.352 and 0.349 respectively. This may occur because the algorithm finds subsequences that are *common* to the target sequences and the random sequences produced by shuffling the entire set creates a uniform environment for detecting subsequences.

Using Subsequences to Discriminate Nonwar and War Crises

If common subsequences reflect complex but deliberately planned political activities, one would expect different types of crises to be characterized by different subsequences. If those subsequences are sufficiently distinct, they could then be used to discriminate between crisis types. Using a nearest neighbor approach, one can characterize each target sequence by a vector giving the fit of each of the subsequences to the target sequence; these fit vectors locate each target sequence in an N-dimensional space, where N is the number of subsequences. Ideally, sequences that have characteristics in common will cluster in this space.

Table 6.9 gives fit of the ten sequences in the "Mixed" data set to the six subsequences of that set. Fit in this table is measured as

$$\text{Fit} = \frac{\text{number events matched - number events not matched}}{\text{total length of the subsequence}}$$

The columns give the vector corresponding to each of the sequences in the set. The second half of Table 6.9 reports one measure of the distance between sequences: the Pearson product moment (r) of the two vectors. Sequences that have similar fits would be expected to have a high r; dissimilar sequences a low r. This expectation is borne out in general in Table 6.9, though the results are less than spectacular.

Table 6.9. Comparing Sequences by Fit to Subsequences

Subsequence fit for each sequence in Mixed

	past	*1stm*	*fash*	*2ndm*	*bosn*	*schl*	*spam*	*cent*	*balk*	*chac*
A	0.00	-0.33	-0.33	-0.66	1.00	0.00	0.00	-0.33	1.33	1.33
B	0.00	0.44	-0.11	0.00	1.22	-0.33	0.77	-0.33	0.33	1.00
C	0.00	-0.33	-0.66	-0.33	-0.33	1.00	1.33	-0.33	-0.33	0.00
D	-0.40	-0.20	-0.20	-0.60	-0.40	-0.20	0.80	-0.40	0.00	-0.20
E	-0.71	0.28	-0.71	-0.14	0.28	-0.42	-0.71	-0.71	0.00	0.00
F	-0.50	-1.00	-1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00

Correlations between sequence fits

	past	*1stm*	*fash*	*2ndm*	*bosn*	*schl*	*spam*	*cent*	*balk*
pastry	1.00								
1stmor	0.08	1.00							
fashod	0.51	0.62	1.00						
2ndmor	-0.27	0.11	-0.42	1.00					
bosnia	0.40	0.47	0.45	0.16	1.00				
schles	0.49	-0.43	-0.30	-0.19	-0.45	1.00			
spam	0.47	-0.44	-0.01	0.10	-0.38	0.62	1.00		
centam	0.28	-0.77	-0.27	0.18	-0.06	0.30	0.72	1.00	
balkan	0.39	0.04	0.41	-0.46	0.73	-0.30	-0.41	0.04	1.00
chaco	0.65	0.27	0.51	-0.16	0.92	-0.18	-0.21	0.07	0.87

Figure 6.3 uses correspondence analysis to cluster the sequences.[14] All of the five wars (filled dots) cluster in the center of the graph, with the nonwar crises on the periphery. In the attribute space (not shown), the two subsequences most strongly associated with the wars are A and F, which also are the only two subsequences containing "Attack" events. The visual examination of the subsequences in each data set also indicates that the threat subsequences and the war subsequences are quite dissimilar, and one would not expect the threat sequences to fit the war subsequences nearly as well as they fit their own subsequences.[15]

Table 6.9 and Figure 6.3 are a weak test because the subsequences were chosen on the basis of their ability to describe rather than differentiate. It would not be difficult to design a similar algorithm to explicitly search for differentiating sequences, for example modifying the selection criterion in the subsequencing algorithm to maximize the coverage in one set of case while *minimizing* coverage in the other set. To construct a war-identifying subsequence the algorithm would choose, at each stage in assembling a subsequence, the event that occurs in the greatest number of war sequences and smallest number of nonwar sequences using a weighting between such as (# war) minus (# nonwar). This could be extended to the prediction problem—that is, recognizing the

---

[14] Figure 6.3 was produced without subsequence C., whose inclusion distorted the clustering in the two-dimensional map.

[15] A quirk in the recoding of sequences precluded a direct test of this without a disproportional amount of effort...

"warning signs" that a crisis will result in a war without knowing the entire crisis—by using as training examples the initial phase of the crisis rather than the entire crisis.

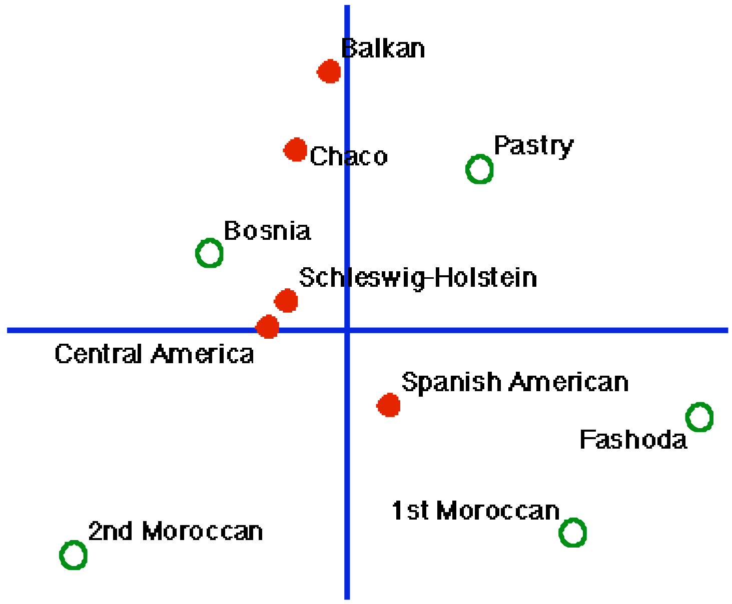


Figure 6.3. Correspondence Map of War and Nonwar Crises

Grammars and Syntactic Pattern Recognition

The two methods demonstrated here are only first attempts at dealing with the problem of sequence recognition, and they are relatively simple, relying largely on a linear structuring of the sequences. A more sophisticated approach would be to impose a grammatical structure on the sequences; this would provide a more flexible specification and there are several reasons to think it might work. The basic modeling approach would be similar to that of the syntactic pattern recognition literature, which is discussed in detail in Fu (1974, 1982)

> …[In the syntactic approach] patterns are specified as being built up out of subpatterns in various ways of composition, just as phrases and sentences are built up by concatenating words, and words are built up by concatenating characters. … The rules governing the composition of primitives into patterns are usually specified by the so-called grammar of the pattern description language. After each primitive within the pattern is identified, the recognition process is accomplished by performing a syntax analysis … to determine whether or not it is syntactically correct with respect to the specified grammar. (Fu 1974: 1)

Fu further notes "one of the most attractive aspects is the recursive mature of a grammar. A grammar rule can be applied any number of times, so it is possible to express in a very compact way some basic structural characteristics of an infinite set of sentences." In this respect, a grammar is functionally similar to a differential equation, which specifies the basic mechanisms of a process while still providing flexibility in the choice of parameters and initial values. The concept of an event grammar is by now fairly common: a useful survey of various "story grammar" concepts is found in Alker (1987).

Two recent studies have demonstrated the potential for this technique. Based on an extensive study of documents of US decision-making during the Korean War, Milliken (1994) constructs a sophisticated formal grammar of state action that can be used to characterize complex political episodes. Working in the domain of institutions rather than actions, Crawford and Ostrom (1995) develop a formal grammar of rules. Both of these projects find that a variety of political behaviors can be systematically modeled using a relatively small lexicon—in the case of Milliken, roughly the size of the BCOW coding scheme—and a set of grammatical structures.

Partially-ordered sequences are a simple form of event grammar in the sense that they can be used to generate a set of well-formed sequences. International politics involves a large number of behaviors, but those behaviors are by no means infinite. The possible set of responses to events is further constrained by resource availability, institutional operating procedures, and custom. As noted in Chapter 3, the predictability of the international environment is necessary for the system to function: if each event could evoke a full range of possible responses, from war to surrender, international affairs would grind to a halt amid chaos.

The common metaphor “the language of diplomacy” may also be accurate as description. Foreign policies can transmit meaning through actions as well as words, and those actions may assume a quasi-linguistic structure dependent upon a sequence of events rather than any single event. Because the explicit content of those events can vary—the USA/PRC ping pong match obviously involved more than young athletes furiously bouncing white balls across a table—the meaning of a sequence must lie to some extent in its structure. If this structure of action can be understood across time, cultures and policy substitutions, it probably has at least a rudimentary grammar.[16] This is not to say the grammar will be precise and unvarying, nor will it apply to all events that occur in the system: the diplomatic grammar of Kissinger differs from that of Khomeini just as the English grammar of William Safire differs from that of Langston Hughes or James Joyce. But most events most of the time can be expected to follow some sort of order. Behavior that deviates significantly from the expected order signals that one is dealing either with an unusual situation or with someone who doesn’t know (or won’t follow) the rules.

Grammars are difficult to induce using machine learning methods, but, as I will suggest in Chapter 7, very large amounts of machine-readable text describing political events are now available and it might be possible to adapt some of the newer computational linguistic methods to work on the development of political grammars. The specification of some low-level rules and the selection of regular sequences might be

---

[16] This parallel between language and action may extend further: If we assume, following Chomsky, that human linguistic abilities are at least partially genetic, then the interpretation of complex social behavior could also have a genetic component. Social interaction in most vertebrates is highly stylized and involves the communication of specific signals (i.e. "Get out of my territory", "Let's mate", "Something dangerous is coming") to invoke specific responses. These actions are frequently quite complex—particularly those involving fighting and mating—and can be at least partially described syntactically. The cognitive ability to interpret physical actions having social significance preceded the ability to process language in evolution, and linguistic abilities may have been adapted from the mental hardware used to interpret physical activity. While organizations are not under the same cognitive constraints as individuals in this respect, there may still be similarities.

sufficient to bring the problem of constructing these into the range of machine-learning—or at least machine-assisted—systems.

Appendix

The Behavioral Correlates of War data set (BCOW; Leng 1987) focuses on a limited number of historical crises using a variety of historical sources, so it has a higher events density than WEIS or COPDAB. The BCOW coding scheme is similar to that in WEIS but more detailed and arguably more precise because it is optimized to code the events that occur in international crises. While BCOW codes both physical and verbal activity, I analyzed only the physical actions on the assumption that these would be more regular than verbal actions over time and across cultures. The focus on physical actions considerably shortens the sequences: this was important due to memory constraints and because the number of operations required to compute a Levenshtein distance is proportional to the product of the length of the two sequences.

Recoding

I used the 5-digit event code in columns 25-29 of a BCOW physical action record, then added a prefix indicating which of five types of dyadic relations were involved in the event, based on BCOW's identification of the "sides" of a conflict. Denoting BCOW's "Side A" and "Side B" as the principal actors in the conflict, and all other actors as "others", the five prefixes are:

1 Interaction between principals

2 Principal as initiator, other as target

3 Other as initiator, principal as target

4 Interaction within principals (e.g. between actors on same side)

5 Interaction between others

For example:

112111 = Diplomatic consultation between Side A and Side B

312111 = Diplomatic consultation initiated by an "other" and directed to either Side A or Side B

Events that occur only once in a set of data—and as such cannot be part of a subsequence common to two or more crises—were recoded to zero to save space; these zeroed events are not used in determining subsequences. Events within a single day are sorted by code so that if identical events occur within a single day in two sequences they will be in the same order in both sequences.

In the parallel event sequence test, I experimented briefly with using nine codes that distinguished between Side A and Side B; this produced no unusual results. Side-specific codes are ambiguous since there is no firmly predetermined identity to "Side A" and "Side B"—though BCOW tends to code the victor of a dispute as Side A—and the combinatorics involved in ascertaining whether better subsequences would be found if the identities of Side A and Side B were reversed in some sequences seemed more trouble than was justified.

Filtering

A persistent problem encountered in Bennett and Schrodt (1987) was that common events—consultations, accusations and, in the case of war, acts of violence—constitute a large part of the data. Common events provide a great deal of regularity—for example much of the WEIS subset Bennett and Schrodt examined was taken up with 12-12-12-12-12 (Accuse) and 22-22-22-22-22 (Force)—but they contribute little to understanding.

From an information theory standpoint, these high frequency event are noise. They can be eliminated without loss of information about the underlying sequence because they are occurring with a much higher frequency than the frequency of the underlying signal. The ongoing shouting matches and wars in WEIS (and their BCOW equivalents) mask the slower, more significant processes of military escalation or diplomatic rapprochement; they are the event data equivalent of the static from a lightening storm intruding on a radio broadcast of the "Toccata and Fugue in D". One needs, therefore, to apply a high-frequency filter to get rid of the junk in the event stream before looking for the lower-frequency regularities.

The sequences were filtered on the basis of novelty: An event of a particular code was included in the filtered sequence only if it had not occurred in the previous N days, where N is an empirically determined parameter. Novelty filtering has some face validity: To a human analyst, the onset of hostilities is important, but after that point, the day-to-day continuation of hostility provides little new information. Since BCOW codes distinguish the cessation of action more clearly than do WEIS codes, the end of a conflict or negotiation will usually be demarcated by the occurrence of a new code, which will pass through the novelty filter.

The novelty filter also deals automatically with the "nonevent" problem—the issue of whether the *absence* of activity between two actors should be coded. The resumption of an activity after a period of inaction will cause the appearance of an event code; continual activity will not. Finally, novelty filtering insures that each event found in the original data will occur at least once in the filtered event sequence.

Experiments with three BCOW data sets—fashoda, suez and cyprus—showed a clear leveling off in the size of the sequences at a filter length of about 14 to 16 days.[17] Filtering usually resulted in a file containing 40% to 60% of the original events; this was higher in short sequences (e.g. pastry, 76%) and lower in very long sequences (e.g. suez, 29%). Parallel subsequence experiments using data processed with only a three-day filter produced no unexpected differences in the results: the shared subsequences tended to be dominated by the high frequency events and the total coverage was higher.

Levenshtein Metric Test

The four subsets of crises listed in Table 6A.1 were analyzed. The short names (e.g. "pastry") correspond to the BCOW file identifiers. "Training" sequences were used to establish weights that discriminated between the war and nonwar sequences; the system was validated with the remaining sequences. The BCOW crises not included in the study are generally those whose length in events is very long (e.g. Suez or the Cuban

---

[17] The 14-day frequency limit, determined empirically for the BCOW data, turned out to be the same "nonevent" time period as determined by guess and intuition in Bennett and Schrodt (1987).

Missile Crisis); or those I could not easily classify into war or nonwar (e.g. Trieste). No deliberate attempt was made to manipulate the results by choice of crises except that the training cases were representative of the validation cases.

Parallel Event Sequences Test

This study used four subsets listed in Table 6A.2. The "Threats" set contains crises that did not result in war because one side backed down; the two "War" sets contain crises that involved wars; and the "Mixed" set included five nonwar crises and five wars. The BCOW crises not analyzed are those directly preceding the two world wars and a set of largely post-WWII crisis that involve some military activity but do not escalate to a full-scale war (e.g. the Berlin airlift).

Table 6A.1. Data Sets Analyzed in Levenshtein Metric Test

Crises without war

Training Set

BCOW file	Crisis	Date	Length*
fashod	Fashoda Crisis	1898-1899	32
1stmor	First Moroccan Crisis	1904-1906	79
bosnia	Bosnian Crisis	1908-1909	116
2ndmor	Second Moroccan Crisis (Agadir)	1911	38
rhine	Rhineland Crisis	1936	65
	Test Set		
pastry	Pastry War Crisis	1838-1839	41
brtprt	British-Portuguese Crisis	1889-1890	15
anschl	Anschluss Crisis	1937-1938	37
munich	Munich Crisis	1938	114
berair	Berlin Blockade	1948-1949	118

Crises involving war

Training Set

BCOW file	Crisis	Date	Length*
schles	Schleswig-Holstein War	1863-1864	52
spam	Spanish-American War	1897-1898	171
centam	Second Central American War	1906-1907	71
chaco	Chaco Dispute and War	1927-1930	125
italet	Italo-Ethiopian War	1935-1936	260
	Test Set		
balkan	Balkan Wars	1912-1913	115
palest	Palestine War	1947-1948	177
kash1	First Kashmir War	1947-1949	70
kash2	Second Kashmir War	1964-1966	76
bangla	Bangladesh War	1971	108

*Length = number of events in the filtered sequence

Table 6.A2. Crises Analyzed in Parallel Event Sequence Test

THREATS: Crises without war		
pastry	Pastry War Crisis	1838-1839
brtprt	British-Portuguese Crisis	1889-1890
fashod	Fashoda Crisis	1898-1899
1stmor	First Moroccan Crisis	1904-1906
bosnia	Bosnian Crisis	1908-1909
2ndmor	Second Moroccan Crisis (Agadir)	1911
rhine	Rhineland Crisis	1936
anschl	Anschluss Crisis	1937-1938
munich	Munich Crisis	1938
WAR1: Pre-WWI conflicts		
schles	Schleswig-Holstein War	1863-1864
rustrk	Russo-Turkish War	1877-1878
spam	Spanish-American War	1897-1898
centam	Second Central American War	1906-1907
balkan	Balkan Wars	1912-1913
chaco	Chaco Dispute and War	1927-1930
WAR2: Post-WWI conflicts		
italet	Italo-Ethiopian War	1935-1936
palest	Palestine War	1947-1948
kash1	First Kashmir War	1947-1949
suez	Suez Crisis and Sinai War	1956-1957
kash2	Second Kashmir War	1964-1966
sixday	1967 Middle East War	1967
bangla	Bangladesh War	1971
MIX: Mixture of threat and conflict cases		
pastry, 1stmor, fashoda, 2ndmor, bosnia, schles, spam, centam, balkan, chaco		

Chapter 7
Conclusion

> *[I hope] at least computers might one day attain the level of the little wormlike nematode* Caenorhabditis elegans. *Blessed with only 302 nerve cells, this lowly creature is truly elegant compared to a Cray [supercomputer], able not only to adjust its squirming to suit its needs, but to learn to recognize patterns and avoid obstacles and even dangers.*
>
> *Erik Sanberg Diment*

> *But then with the throttle screwed on there is only the barest margin, and no room at all for mistakes. It has to be done right, and that's when the strange music starts, when you stretch your luck so far that fear becomes exhilaration and vibrates along your arms. The Edge… There is no honest way to explain it because the people who really know where it is are those who have gone over. The others pushed their control as far as they felt they could handle it, and then pulled back, or slowed down, or did whatever they had to do when it came time to choose between Now and Later.*
>
> *But the Edge is still out there.*
>
> *Hunter S. Thompson*

This final chapter wraps up several topics that apply generally to computational modeling. Some of these might have just as appropriately gone into Chapters 2 or 3 but were reserved to this point so that the reader would have concrete examples of computational models against which to assess the arguments. I will first consider the parallels between computational modeling and classical approaches to studying international behavior; I argue that in several respects computational models are closer to the classical methods than are other formal modeling methods. Next, I look at what computational models are *not*, and defend the inductive approach characteristic of these models. I then consider some of the implications for computational modeling of the rapid changes in the availability of computing hardware, software and data. Finally, I suggest some directions that computational modeling efforts might usefully take in the near future.

Computational Modeling and the Classical Approach

This book has made extensive reference to the "classical" or "traditional" approach to international relations.[1] Gaddis (1992/93) provides a thorough discussion of the failures of both the scientific and traditional analytical techniques to predict the end of the Cold War. While Gaddis, an historian, suggests that behavioralists need to pay more attention to history and less to the models of the physical sciences, he also notes that the

[1] I have assumed that the reader knows what this refers to: if not, see Dougherty and Pfaltzgraff 1981 or, more briefly, Bull 1966.

traditional analysts did no better on this problem than the behavioralists.[2] In another context, Khong (1992,257) notes of the advisors who produced the disastrous United States policy in Vietnam: “Critics who lament that the policy-makers knew too little history might ponder if a more distinguished and knowledgeable group of officials has since been assembled by any president.” Formal modeling has its problems, but so do historical and traditional approaches.

The failure to predict the end of the Cold War should—though probably won't—spawn a search for new methods, just as the Great Depression revolutionized macroeconomic theory and analytical techniques in the social and policy sciences generally. One interesting aspect of computational modeling are the parallels it has with the classical approach; in fact a number of the suggestions made by Gaddis for improving formal efforts are already found in computational models.

Small N, large M analysis

Statistical studies tend to look at a small number of variables (M) on a large numbers of cases (N), while traditional studies look in detail at a large number of characteristics in a small number of cases. While some computational models (e.g. ID3, neural networks) are most effective in a large-N/small-M environment, the opposite approach can also be found. Rule-based models such as JESSE or POLI follow a strategy quite similar to that of the classical case study in analyzing a single case in great detail; they differ from the classical approach in doing this formally. In addition, many computational methods are designed to be knowledge-intensive and can thus accommodate a much larger number of variables than the typical statistical analysis.[3]

Use of precedent and analogy.

One of the emphases throughout this volume has been the use of precedent and analogy. As noted in Chapter 3, this has extensive parallels in traditional international relations theory. In general, the tacit model in most traditional arguments is some variation on "If X happened in the past under conditions C_1, C_2, ..., C_n, then if conditions

[2] Gaddis may also be overestimating the influence that behavioralist methods have had on policy-making and forecasting. Laurance (1990) documents the failures of the event data approach in this regard; O'Neill (1994) does the same for game theory; Cooper (1978) for simulation; Herspring (1992) for behavioralist approaches generally. The behavioralist enterprise has paid little attention to problems relevant to foreign policy in part because no one was listening, particularly compared to the influence of the "slow journalism" approach. In this sense behavioralism has not failed as a policy tool so much as it has never been seriously attempted.

[3]The prospect of "data mining" for relationships in large data bases has attracted considerable interest in recent years. For example, a recent medical study on the blood cholesterol/heart disease linkage involved a database of about 1400 variables on over 800 patients followed for 7 to 14 years. The resulting dataset was 300-megabytes in size and, in classic scientific understatement, "the large number of variables as well as the complex relationships between them are the real challenges of the project." (Long et al 1989; also see the Piatetsky-Shapiro and Frawley 1989; Fayyad et al 1995 and *Byte* October 1995, 81-108). While this approach is anathema to the usual standards of experimental design (e.g. King, Keohane and Verba 1994) it can be justified as exploratory data analysis. If the researcher doesn't have a clue about what combinations of lifestyle factors might affect the cholesterol/heart disease relationship, systematic methods of finding these relationships in a large number of variables are more useful than trying to work through 2^{1400} subsets of those variables.

similar to $C_1,C_2,...,C_n$ occur again, X is likely to happen again". The most common evidence provided for any argument in international politics in an historical example, and most international relations theories are richly illustrated with historical instantiations of their principles.

Flexible assumptions about cognition.

Traditional theories of international behavior usually assume a cognitive process is involved.[4] In some of these theories—most notably realism—self-interest is assumed to be the dominant process, but other theories (e.g. Jervis, Allison, Lebow) invoke complex psychological and organizational decision-making mechanisms. Computational techniques provide a variety of methods for modeling information processing and learning in individuals and organizations; in particular, these are much richer than the cognitive assumptions found in rational choice modeling.

The sentient organization as primary actor.

With the exception of the "Great Man" approach, most traditional theories recognize the importance of organizations in determining behavior. While the set of organizations within a "nation-state" may be treated as a unitary actor with respect to some questions (e.g. in the simpler aspects of balance of power theory), few traditional theories go very far without delving into the decision-making of Foreign Offices, military bureaucracies, executives, business organizations, social classes, political parties and other collective entities. Rule-based models can, and have, modeled organizational decision-making in considerable detail; extensions of neural network models may be able to differentiate between situations where the organizational structure is robust and those where it is brittle.

While there remain major differences between the traditional and formal approaches—particularly with respect to the preparation required to understand them—I think these are often differences in language (natural versus formal) and technique rather than differences in objective. Clearly both approaches are empirical, seeking to study the real world rather than creating abstract mathematical or philosophical structures. Each approach contains communities who seek to be policy relevant, and this means they must be able to make predictions. Both approaches simplify historical experience into a few principles of behavior. When focused on the issue of what to study rather than how to study it, both approaches have reasonably similar definitions of what constitutes international behavior and what constitutes an interesting problem (e.g. international conflict, international economic inequality), particularly

One can imagine a situation where classical techniques coexisted with a fully developed formal theory of international relations (which we currently do not have), much as engineering coexists with physics, management coexists with formal economics, and clinical medicine coexists with statistical epidemiology. At the moment, however, many formal theories of international politics show considerably greater divergence from informal theory than physics does from engineering. Given the somewhat dubious

---

[4] A few theories of historical determinism—Toynbee, some variants of Marxism, and some theories invoking Social Darwinism—are about the only exceptions.

pedigree of many formal models of international behavior—born as they are from physics, economics and mathematical convenience—more attention to the characteristics of the classical approach might be useful.

Some problems will always exist that require a rapid solution formal methods cannot provide. Just as a physicist in a skidding car does not pause to work out the Newtonian mechanics of the situation, so a diplomat in an unforeseen crisis can hardly be expected to use regression analysis or develop a rule-based model. But formal methods retain an advantage in mapping the broader regularities of behavior. Formal techniques are superior to natural language in preserving logical consistency in a complex argument, and a computer can dispassionately process vast quantities of information that are impossible for a human brain to handle. There is a place for both approaches.

What Computational Models Can't Do

> *The world is a stupendous machine, composed of innumerable parts, each of which being a free agent has a volition and action of its own, and on this ground arises the difficulty of assuring success in any enterprise depending on the volition of numerous agents. We may set the machine in motion, and dispose every wheel to one certain end, but when it depends on the volition of any one wheel, and the corresponding action of every wheel, the result is uncertain.*
>
> *Niccolò Machiavelli*

The discussion throughout this volume has generally looked at glasses half-full and the silver linings of dark clouds, if not flying pigs. Facing the dark side for a moment, it is useful to look at some things that computational models cannot do, however sophisticated they may be, with however large an historical data base, and implemented on whatever hardware. This is my guide to where the dead ends of computational modeling are located.

A computational model is not a model of the brain!

Table 7.1 compares the characteristics of the hardware used in computational modeling and that used for human political decision-making. Not only is the basic architecture different, but also the two differ by at least five orders of magnitude in their storage capacity. While it may be possible in the future for computer hardware to get closer to that of the brain—for example with increased speed, memory and parallel architectures—at the moment we are not anywhere close.

Table 7.1. Computers versus Brains

	microprocessor (ca.1995)	**brain**
Data element	byte (0..255)	neuron (binary state) neural connection (continuous state)
number of elements	10^7	10^{11}
number of connections	10^7 - 10^8	10^{15} (?)
connections per element	1 - 10	10^4 - 10^5
mode of operation	serial	parallel
Firing time (seconds)	10^{-8} (10 nanosecs)	5* 10^{-3} (5 millisecs)
Reaction time for identifying a complex pattern using a large knowledge base	10^2 - 10^3 seconds	5* 10^{-1} secs (1/2-sec)
Mode of recall	Serial by location	Associative by content

As will be discussed below, improvements in computer hardware will have profound implications for the ease with which one can experiment with computational techniques but over the past decade these improvements have been only about two orders of magnitude. We've still got at least four to eight more orders of magnitude separating the computer from the brain.

As noted throughout this volume, these computational constraints are much less when one is dealing with organizational behavior. Judging from employment trends, microprocessors began to seriously compete as components of the middle layers of organizations around 1988, with the introduction of the Intel 80386 and Motorola 68030. But computers still can't match the individual brain, particularly in tasks involving recall. Computational models, while guided by human cognitive activity, must still be developed with the information processing limitations of computers.

Understanding political theory

A theory of political behavior that makes sense to a human being will not necessarily make sense to a machine. The term "makes sense" is here defined as a theory that can be transmitted and understood by another human and used to predict and/or guide behavior. Humans have been able to construct such theories for at least 2,500 years: We still read Thucydides and Mencius, and every year thousands of undergraduates apply the theories of Thucydides and Mencius to contemporary politics in essays and term papers, not to mention applying the theories of Cicero, Augustine, Ibn-Khaldun, Machiavelli, Hobbes, Rousseau, Marx, and Keynes. Presumably these theories have some empirical validity—we accord them a place in the curriculum distinct from that accorded to *Lord of the Rings* or *Peter Pan*—and this validity transcends time, language, culture and technology, no small feat.[5]

This does not, however, imply that those same theories will be sensible to a machine, in the sense that a machine can be expected to be able to assimilate, extend and make predictions on the basis of the theories. Any 20th century reader of Thucydides has, by virtue of that act, already experienced large scale social organization, gone through a period of social and personal discipline required to become literate, has gone through a long period of human intellectual, social and emotional development, and due to the associative capacity of the human brain, uses this information while "understanding" Thucydides. Thucydides makes sense to a modern reader because that reader, by virtue of being a reader, has much in common with the experiences and environment of Thucydides. Thucydides would probably not make sense to a 16th century Inuit with no experience with large scale political hierarchies, and demonstrably the political argumentation of Thucydides makes little sense to a 20th century seven-year-old, even if the words and sentences are understood.

Comprehension and utilization of political theory, then, is a combination of both *transmission* and *experience*. It is not sufficient simply to convey information; instead it must link with the information common between the sender and receiver. Just as a modern archaeologist must puzzle out the utility of oddly-shaped stone tools in a Neanderthal's cave, and a Neanderthal would have found equally mysterious a 20th century mechanical pencil, both objects make complete sense, and have utility, in their original environment. As 20th century humans, we find our political environment has enough in common with the environments of Thucydides' Greece and Mencius' China—to say nothing of the industrializing Western Europe of Marx and Weber—that we can understand their theories.[6] The development of aircraft could not proceed until

---

[5] There are in addition numerous works of "political theory" that we in twentieth-century North America do *not* find transcend time and culture, including virtually the entire production of the medieval Church, the pro-slavery tracts of the 18th and 19th centuries, and most of the German bellicist writers, not too mention much of non-Western political theory. [2nd edition: And in fact, due to J. R. R. Tolkein's status as on the 20th century's greatest medievalists, *Lord of the Rings* is actually a fairly good introduction to many aspects of medieval mindset. My opinion remains unchanged on *Peter Pan*...]

[6] An extreme cultural relativist would argue that true "understanding" is never achieved because there are always critical differences. In the strictest sense, this is true, but the issue is actually one of degree, not kind. Environments are never identical between individuals, much less societies: George H. W. Bush's Vice President J. Danforth Quayle and I both grew up in Indiana, in the United States, in a late 20th century Anglo-American political culture, and yet many of our fundamental assumptions about political culture are doubtlessly (hopefully...) profoundly different. In contrast, I can go to a region where there is a totally different political culture, and with time and effort still *learn* to internalize enough of that environment to be able to "understand" that culture sufficiently well to make intelligent statements about it. The validity of these statements can be ascertained either by the agreement of individuals who are native to the culture (agreement by a native being the criterion used in ascertaining whether a sentence is grammatically correct) or by making successful predictions.

I've noticed that when one becomes involved in serious first-hand cross-cultural political exchange (as opposed to debate), a tremendous amount of effort is spent exchanging information that can be used to construct an accurate model of how the other political system works ("No, no, that's not what we would do, and here's why..."). The process is eerily similar to a machine-learning protocol. Such communicated understanding is never perfect—just as an accent learned by an adult is seldom as good as that acquired in infancy by a native speaker—but much can be learned, and it is greatly facilitated by the accumulation of designers learned to pay attention to the aerodynamic shape of bird wings but not to the birds' method of propulsion; we need to make similar decisions in the design of computational models of political behavior.

In Defense of Induction

Most of the approaches I have discussed in this work are inductive: They seek to derive patterns out of a set of data rather than using data to validate general theoretical principles. This approach is at odds with much of the behavioral approach, which is deductive. The early behavioralist researchers in international relations clearly—one might suggest obsessively—compared their endeavors to those of the physical sciences. Their archetypes were the idealized models of science expounded by Popper, Hemple, Carnap and, *primus inter pares*, Kuhn.[7]

This infatuation with the supposed norms of the natural sciences neglected the not insignificant fact that very little scientific research, outside the rarefied realms of purely theoretical physics and cosmology, lives up to such standards, and the mathematical social sciences have gone well beyond the physical sciences in their standards of rigor. Reporting on the Santa Fe Institute, a collaborative effort of physicists and economists to share ideas, Pool observed:

> The physical scientists were flabbergasted to discover how mathematically rigorous theoretical economists are. Physics is generally considered to be the most mathematical of all the sciences, but modern economics has it beat. ... The flip side of the physicists' surprise at the rigor of the economists was the economists' astonishment at the physicists' lack thereof. (Pool 1989,701)

The biological sciences, ploddingly inductive, evolutionary rather than revolutionary, experienced in dealing with complex interacting systems rather than idealized points and vectors, would have been more appropriate model for the formal study of international politics, but this was not to be.

Because computer understanding is distinct from human understanding, computational models require an inductive theory of human behavior—one that can discover, rather than verify, laws or patterns. Unfortunately, induction has been given very little attention in 20th century philosophy of science:

> Bacon (1620) undertook to construct a theory of discovery that included advice and heuristics for effective discovery processes, and John Stuart Mill's (1843) famous

---

information and experience within the culture; nothing beats being there. One of the dumbest assumptions of the 1960s behavioralists was that this process of area-specific knowledge acquisition could be completely bypassed when trying to understand political behavior.

[7]The social scientific community's obsession with the Kuhnian concept of scientific progress is a mystery in light of the all but total rejection of Kuhn as a *general* model of science by historians and philosophers of science. One need go no further than the whole of the biological sciences to find a counterexample. Biology shows progressive refinement of a single model rather than the dramatic paradigm shifts predicted by Kuhn. Darwin's theory provided the theoretical underpinnings for the inductive Linnean classifications; Mendelian and population genetics provided mathematical order to the Darwinian system; the biochemical revolution beginning with the study of DNA provided a chemical explanation for Mendelian genetics. Through it all, chimpanzees were considered closer to apes than to frogs, and 17th century botanical studies can still be useful to a 20th century biological ecologist.

> rules of induction were intended to facilitate and systematize the process of generating hypotheses. However, in the late nineteenth century, philosophers of science began to express more and more doubt about the feasibility of providing "rules of discovery" and to turn their attention more and more towards the process of verification. [By the twentieth-century] it was frequently claimed that there could be no logical method or theory of scientific discovery; that discovery was a matter of "creativity"—the free play of the human creative spirit. (Langley et al 1987,37)

While the early 20th century positivists were clearly reacting in part against the admittedly excessively descriptive nature of 19th century science, one cannot help but suspect that the *image* these philosophers had of turn-of-the-century physics—pure intellect attacking problems ranging from the invisible structure of the atom to the origin of the universe—was more attractive than their image of the mid-19th century empirical inductivists, who mucked about describing the slugs they found under rocks and doing life-threatening experiments with electricity, heavy machinery and toxic chemicals.

The late 20th century renaissance of interest in induction (see for example Michalski, Carbonell and Mitchell 1983; Langley et al 1987; and Holland et al 1989), in turn, comes out of a recognition that these early philosophers of science were working with an idealized view of the enterprise. Even in physics and astronomy, the triumph of pure intellect is the exception, not the rule:[8] Maxwell's elegant unification of electromagnetic theory was possible only because it built on the ugly experiments of Faraday; the Copernican reformulation of astronomy came only after the meticulous empirical work of Brahe.

In fact, pattern recognition has often been the precursor to axiomatic systems in science. The Darwinian theory of evolution was built inductively on the Linnaean taxonomies of the 18th and 19th centuries, augmented by Darwin's own global observations during the voyage of the *Beagle*. Darwin's eventual theory, while deductive, arose from his heuristic analysis of a very large database rather than from first principles—in fact, it arose *despite* the first principles of the prevailing Biblical theories of creation.

Astronomy, physics and even mathematics also followed this progression; Babylonian and Egyptian mathematics were originally based in pattern recognition.[9] Mathematics was not put on an axiomatic base until the Greeks, and astronomy and

---

[8] Brush (1989: 1124) notes for example: "The first large-scale systematic effort to test [Popper's falsification thesis] and other methodological claims against evidence from the history of science was planned and carried out by a group of scholars at [VPI]. A striking result of their research is that predictiveness or falsifiability was not considered important by key scientists in the three cases where this factor was examined: ... Galileo's work on the Copernican system, ... Ampère's electrodynamics, and ... the reception of nuclear magnetic resonance." For other discussions of scientific creativity, see Hanson 1958, Newell and Simon 1972; Nickles 1978 and the studies of induction cited above.

[9] Eves observes

> It should be noted that in all ancient Oriental mathematics one cannot find even a single instance of what we today call a demonstration [proof]. In place of an argument there is merely a description of a process. … Moreover … these instructions are not even given in the form of general rules, but simply applied to sequences of specific cases. … It was expected that from a sufficient number of specific examples, the general process would become clear. (Eves 1983, 22-23)

physics did not gain this form until the eighteenth century. In the modern sciences, organic chemistry is still deeply involved in pattern matching, since the interactions of complex molecules are too convoluted to predict theoretically, and this is also true for much of geology, most of medicine, and most of meteorology.[10]

If one looks at what we do as opposed to what we say, most contemporary social science research is in fact quite inductive. I know of hardly any methodologists who formulate *a* hypothesis, uniquely operationalize each variable, collect the appropriate data, and then conduct *a* test.[11] More commonly, we start with a general set of hypotheses, try an assortment of alternative variable specifications, experiment with many, many different forms of the relationship, and then—recognizing that we've violated any pretense of the independent trials assumed by the significance test—hope that the model we finally publish is at least a reflection of an underlying political reality rather than an exercise in machine-assisted self-deception.

The early behavioralists studying international relations thought that they were building their empirical science upon fully-developed and logically consistent theories of international behavior: Realists and game theorists on the political right and Marxists on the political left were independently convinced their approaches provided such a foundation. But those theories survived neither logical nor empirical scrutiny. The study of international politics is not poised for takeoff into the halcyon heights of a purely deductive science; there remains instead a lot of ugly inductive grubbing to be done.

One advantage to using a computer for inductive studies is that the machine comes to the data with no preconceptions, and so can provide a "second-opinion" biased only by the choice of variables and the model presented it. For example, Bennett and Schrodt (1987) found the behaviors measured by the WEIS event data set to be exceedingly boring most of the time. We didn't go into the study thinking that the data were boring, because like most analysts, we focused on the exciting parts. But from the machine's perspective as an outside observer, the events were pretty predictable most of the time. A human in contrast comes to any data analysis with numerous preconceptions based on culture, experience, academic learning, organizational norms of interpretation and other biases. We probably come closest to pure induction when dealing with a radically different culture (e.g. Islamic fundamentalism for most North Americans), and then often as not get it all wrong.[12]

A machine is disadvantaged in having neither physical common sense nor, more importantly, a human "personal sense" that allows it to fill in missing information and infer motives. Unless programmed or taught that sort of information, a computer can

---

[10] While weather forecasting can be done deductively using numerical simulation—a technique pioneered by the Lewis Richardson of arms race fame—human weather forecasters tend to use heuristics, analogy and pattern recognition. Kerr (1989) discusses recent developments in weather forecasting using computerized pattern recognition; one such system "is as skillful as human forecasters have been during the past 25 years" in making 90-day forecasts. Computer-generated graphics have revolutionized the presentation of weather information to the point where the topic can support its own cable TV channel, with nary an equation in sight.

[11] I do know of one such case, and the test found a single rather obvious factor explained 85% of the variance in a dataset that had required a decade to prepare.

[12] [2nd edition] Yes, I wrote that in 1995…

only detect certain forms of behavioral regularities. Machines are not about to replace human political analysts, but they certainly could supplement them.

McCloskey (1985,50) notes that in the 1930s one of the consequences of the huge influx of European academics fleeing fascism was that the indigenous United States scientific modernism of for example, William James and John Dewey, was "killed off ... in favor of a harsher European type." Almond (1991) charts much the same transition in some aspects of political science. However, the biological sciences in the United States—sufficiently well established in this agricultural country to survive a European takeover—continued to thrive on their indigenous roots.

While heartily approving of certain aspects of this European influence—for example the introduction of mathematics as a logical language—I do wonder, particularly in light of my earlier remarks on the role of culture in understanding, whether the resistance shown by most beginning graduate students to internalizing methodological concepts originating in Vienna in the 1920s may stem from trying to nurture those concepts in alien ground. Given sufficient resources and protection from competition, an alien plant can survive in any environment; still, there is something to be said for the native flora.[13]

Technology

Computational modeling is dependent at a very fundamental level on the availability of computing power and large amounts of data. Designing a computational process 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.

Computers

Since the early 1960s, computing power has always been available to an elite, though 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

---

[13]Lest I be considered a cultural isolationist, what bothers me is the privileged position of *northern European* concepts in the social sciences. When French and German vocabulary are used in social science debate—preferably without translation—this is considered a sign of the *savoir faire* of one's *Gemeischaft*. Use an Islamic or Confucian concept and you're considered something of a nut. In far too many cases, what passes as "internationalism" in the social sciences in the United States is simply northern Europeanism, and bi-ethnocentrism is little improvement over ethnocentrism.

This is particularly problematic in computational modeling because northern European culture is an outlier with respect to Hall's (1976) distinction between high context and low context communications. For example, researchers from high-context cultures—Asians, Arabs, Africans and Latin Americans, as well as the high-context subcultures of many North American and European females, in short, most of the world except for the white males of varying states of motility who dominate the social science canon—would find ludicrous the premise that a valid model could be derived solely from the explicit communications within an organization.

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.

By the early 1990s, this situation had changed substantially: Table 7.2 compares the capability of an advanced personal computer that might be purchased by a average computer-oriented social science researcher in 1980 and 1995.[14] In fifteen years, speed has increased by almost two orders of magnitude, random-access memory by two and a half orders, and disk capacity by more than three orders, while the price/salary ratio has *dropped* by half an order of magnitude.

Table 7.2. Comparison of Hardware

	Apple II	**Macintosh 7100**
Year purchased	1980	1995
Processor speed	1 Mhz	80 Mhz
Data bandwidth	8-bit	32-bit
Random access memory	64 Kb	16 Mb
Disk storage	320 Kb floppy disk	720 Mb hard disk
Display	black and white	color
Time required to compile 1000-line Pascal program	2 minutes	1 second
Cost as percentage of assistant professor's salary	\$5000/\$15000 = 33%	\$2500/\$35000 = 7%

Increased speed, memory and storage will not, by themselves, solve all of the problems of computational modeling: one cannot bludgeon all problems into submission with faster hardware.[15] Given researchers have finite amounts of time, experiments that were impractical when a program required ten hours to run become practical when that

[14] 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 an ill-fated experiment in event data analysis in the 1970s (Laurance 1990).

[15] But 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. [2nd edition: As noted in Chapter 8, by 1997 that increased power overwhelmed even the best human opponents.]

same program requires two minutes. Because computational modeling methods are information intensive, the increased disk and memory capacity is also very important.

The hardware available today to the average academic researcher has about the same power as the specialized artificial intelligence workstations of the early 1980s, and substantially more power than any social science user had available in a mainframe environment in any era. This is further leveraged by interfaces and programming environments that enable an individual to manage far larger projects than would have been possible earlier. 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. Of particular interest for computational modeling is the development of inexpensive parallel computing configurations, which could be effectively applied in most of the machine learning and sequence analysis methods I've discussed.

Data Sources

The ability to code data directly from machine-readable sources has substantial implications for computational modeling. Natural language reports on political activity are readily available through commercial on-line data services such as NEXIS, and inexpensive CD-ROM and Internet sources increasingly supplement these. 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.

Coding event data from wire service reports has proven to be a straightforward task (see Gerner et al 1994; Schrodt and Gerner 1994; Schrodt, Davis and Weddle 1994), and the machine-coded data are substantially denser than those provided by human coding. For example Schrodt and Gerner's (1994) comparison of a human coded, *New York Times*-based WEIS set with a machine-coded, Reuters News Service-based set for the Middle East during the 1982-1992 period found that the Reuters set contained an average of 3 times as many events as the *Times* set. This improved density was sufficient to show patterns absent in data coded from the *Times* alone. The machine-coded data were based only on the first sentence of the Reuters reports and a system coding the entire report—or reports from multiple newswire sources—would yield an even higher density. Machine coding systems can easily be adapted to work with languages other than English and this opens the possibility of coding news sources that could provide greater detail, and a different perspective, than English-language sources currently coded in most event data sets.

While the original impetus for machine coding was increased efficiency—a machine coding system can currently code about 15 events per second, which is substantially faster than the typical undergraduate coder—two additional advantages have emerged. First, a specific set of coding rules is both transparent and fixed, so there is never a question as to why a particular sentence was assigned a code, nor a problem with the *de facto* coding rules "drifting" across generations of coders. Second, the coding vocabularies are both transferable and flexible: The vocabulary developed in one project can be used as the basis of a new project, which allows event-coding systems to be refined incrementally.

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 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.

One of the favorite parables of evangelical preachers is of a sailing ship becalmed for weeks in the Atlantic, its crew slowly dying of thirst. Sighting a passing vessel,[16] the crew appeals frantically for water. The other ship replies, "Throw down your buckets; you are surrounded by fresh water!": the beleaguered ship is resting in the outflow of the mighty Amazon River.

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

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 computational models. 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

---

[16]As with many evangelical parables, the movement of this second vessel under windless conditions is not explained...

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 areas where computational modeling may benefit from the 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.

Second, some software implementing computational methods is now becoming widely available. For example, in the month before I began the final revision of this manuscript, SPSS announced the availability of software for constructing a variety of neural networks,[17] the Sage Quantitative Applications in the Social Sciences series issued a new monograph on classification algorithms (Bailey 1995); and *Byte* (October 1995) featured a special section on inductive "data-mining" using clustering and related methods. Software designed for the construction of expert systems using production system or backwards chaining has been available for a decade, as have various "rule-generating" programs based on ID3. As this software becomes more widely available, it will be much easier to experiment with a variety of computational techniques, much as we can presently experiment with the different statistical techniques available in the standard packages.

As noted above, these increased resources will not, in the absence of appropriate choices of theories and data, lead automatically to a computational modeling utopia. But they certainly open possibilities that were unavailable even a decade ago. The fact that a computational modeling experiment has not been done in the past should not be taken as evidence it is a bad idea. Instead, the experiment may earlier have been impractical. It also still could be either impractical or a bad idea, but this judgment should be suspended until it has been assessed with current resources.

Five Challenges for Computational Modeling

The computational modeling field is still in its infancy—except for the early work of Alker, Thorson and Sylvan, and a few other researchers, serious efforts at model

[17] Commercial neural network software has been widely available since the mid-1980s; the breakthrough is seeing it offered in a statistical package such as SPSS.

development for international relations research have been underway for only about a decade. This section will outline five projects that I believe could be done with existing—or soon to exist—theory and technology and which would significantly advance the field:

Real time forecasting of events using machine learning

With the improvements in automated event data coding, real-time forecasting of events is now a possibility. A successful system—even in a limited domain such as refugee movements—would answer the continual criticisms that formal approaches to international politics have told us a lot about European alliance structures in 1912, but they can't say what will happen in the former Yugoslavia in 1997 (see for example Herspring 1992, Gaddis 1987, 1992/93).

As I've argued above, forecasting does not necessarily mean crisis forecasting, the focus of most event data research to date. There would be considerable utility in being able to forecast day-to-day events provided that actual *events* were forecast rather than some grossly aggregated numerical measure. The techniques discussed in Chapters 5 and 6 demonstrated a variety of ways that one can work with discrete events.

Understanding background patterns is probably a prerequisite to any good crisis-forecasting model. Many crisis-forecasting efforts have treated the problem as though it were similar to that of earthquake forecasting, where the geological system is usually doing nothing (or almost nothing) until it erupts into a short, distinct shock. The problem of political crisis is actually more like that of tornado forecasting, where the weather system can produce a variety of complex—and possibly quite damaging—conditions short of a tornado. The task is therefore distinguishing storms that will generate tornadoes from storms that will not, and rather than looking for faint precursors on otherwise quiet days, we must look for distinctive signals amid an already tumultuous background.[18]

Computational models, with their huge knowledge bases, are also much more complex than the models normally encountered in formal international relations research. As a consequence, the models run a greater risk of being tautological, simply incorporating the data on which they have been trained, and parroting this back when tested. Split-sample tests are one way of dealing with this, but prediction solves the problem even more credibly because neither the model nor the modeler has seen the data before.

As noted in Chapter 3, human analysts tend to predict sets of events, rather than single events, and do so in part because the actors in the international systems are themselves typically working through a number of contingent plans. One of the difficult issues in working with event-oriented predictions—as opposed to numerical predictions—is finding a metric that reflects plausibility. For example, if a model of international interactions in the Middle East had predicted war between Israel and Jordan in October of 1990, this prediction,[19] while erroneous, would be considered more

---

[18] Moving the RAND Corporation from the Los Angeles County to Kansas would perhaps increase sensitivity to this distinction...

[19] In July 1990 there was considerable talk "on the street" in Jerusalem of this possibility; the plan, if it ever existed, was preempted by Iraq's invasion of Kuwait in August 1990.

plausible than a prediction of war between Jordan and Saudi Arabia, and much more plausible than a prediction of war between Jordan and Cambodia. In prediction, as in horseshoes, closeness counts.

Evaluating the timing of forecasts presents another set of problems. As noted in Chapter 2, international behavior—in contrast to many physical and economic processes—is only weakly linked to calendar time. Analysts are usually more concerned with the sequence of events than with the precise timing of events and in circumstances where an event has a number of independent antecedent conditions—which is most circumstances—even the sequence may be partially indeterminate. But unless a model can provide some indication of whether the next step in a sequence will occur in the immediate or distant future, it has little practical utility. More work is needed on estimating and evaluating models that involve both sequential and temporal elements.

Finally, international behavior is very modal: most nations in the international system most of the time do the same things again and again. This is a general problem in political analysis, embodied in the (presumably apocryphal) story of the intelligence analyst who, on retiring in 1950, stated "Every day of my career I issued a memo predicting that Europe would remain at peace if at peace, and at war if at war. I was in error only four times." A model that predicts only modal behavior, or in a time series simply predicts autoregressively the continuation of existing behavior, will usually have a very strong empirical record. However, such predictions are of little interest to political analysts. More attention to entropy-based measures, which place greater emphasis on rare events, might be in order.

Identification of changes in policy due to adaptive behavior

Chapter 3 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 computer analysis may be more amenable to detecting these changes because the computer is unaffected by human cognitive preconceptions. Again, 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 of the crisis forecasting methods have focused only on the second part of the problem, and their general failure (aside from the sheer difficulty of the task) may stem in part from the lack of a model that shows how routine learning occurs.

Chronology Generators[20]

Political analysts spend a lot of time with textual materials simply figuring out what events happened in the course of a political episode. We now have available a great deal of machine-readable text on such events—notably newswire sources such as Reuters—but it completely mixes 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).[21] 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 AI (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.

Simulation of realistic international behavior through the self-organization of information-processing agents

Existing global models have been refined in a top-down fashion: the earliest all-computer models had very high levels of aggregation—in the widely-publicized *Limits to Growth* simulation (Meadows et al 1973), simply "The World"—but subsequent models became increasingly complex by disaggregating behavior into regions, nations, nonstate actors and so forth.

---

[20] My thanks to Ed Laurance for suggesting this one.

[21] 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.

An alternative, if ambitious, approach would be to try to build the behaviors from the bottom up using large numbers of autonomous, information-processing agents that would self-organize into recognizable patterns of political behavior. As noted in chapter 2, this is the "artificial life" approach (Goel and Thompson 1988; Langton 1989; Langton et al 1992), with the distinction that simulated human agents learn cognitively and culturally rather than through genetic evolution. This approach is currently being pursued by some formal modelers in economics (see Anderson, Arrow and Pines 1988), and considerable software has been developed at the Santa Fe Institute for such simulations.

While ideally it would be possible to generate international behavior starting from the individual, computational limitations might require a larger (but still subnational) unit to be used; Cusack and Stoll (1990) are close to this approach in their study of realism. Sophisticated "city-state" systems—which have arisen independently on a number of historical occasions including Mesopotamia, China, and the Mayan area of Central America—might be a realistic objective to generate through simulated self-organization.

By programming a sufficient number of ad hoc behaviors into a simulation, realistic behaviors can be simulated easily (see Schrodt 1988a). The problem becomes challenging if one tries to reduce these pre-programmed behaviors to a minimum—for example only preferences, constraints, memory, communication and myopic adaptation—and have complex behaviors emerging through self-organization.

This project would clearly burn a lot of machine cycles and may not yet be practical with machines at a level less than that of a supercomputer.[22] If a self-organizing simulation could be achieved, it would probably lead to substantial insights into both the origins of international systems and to the behaviors of systems that have been severely disrupted and are re-organizing (e.g. the former Yugoslavia; the former Soviet Union; post-colonial Africa). A simulation might also identify key variables—for example resource levels, military effectiveness, and communications—that are likely to change the types of organization the system exhibits.

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

This is without question the most technically challenging of the suggestions I am making, and in fact it may not be possible with current technology. As discussed extensively in Chapter 4, it is clearly possible to create realistic rule-based models of decision-making organizations that account for over 50% of observed behavior, 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—though 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

[22] The simulation itself would be practical on a smaller computer but the experiments required to develop it would not.

political economy and can answer questions about them;[23] 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) is experimenting with a variety of approaches for extracting specific pieces of information on Latin American terrorist incidents from newswire text. The performance of such systems is quite impressive, though they are 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.

For the reasons discussed above, it will still be necessary to add some ad hoc tacit knowledge to these models, though 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 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.

Another data source that has been under-exploited by rule-based modeling efforts are a variety of case-based data sets on war, crisis, and mediation. These include CASCON (Bloomfield and Moulton 1989), SHERFACS (Sherman and Neack 1993), CONFMAN (Bercovitch and Langley 1993) and KOSIMO (Pfetsch and Billing 1994). These datasets contain a substantial number of historical precedents that could be used for case-based learning or reasoning, and their complex data structures—a liability in their use in statistical models—can be easily accommodated in a rule-based model.

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.[24]

Whether the CYC project will succeed or fail is still unclear, though 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

---

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

[24] 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, though these will obviously be culturally dependent.

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.

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, I suspect one would find a significant cross-cultural divergence that could be systematically modeled: as noted in Chapter 3, the political patterns implied by the *Iliad* and the *Ramayana* are quite different.

Chapter 8
EPILOGUE

In the decade or so since the first draft of this manuscript was completed, much has changed in some areas, and little has changed in others. This epilogue is intended to serve three purposes: First, for the benefit of those outside the field of political science and newcomers (read: graduate students) unfamiliar with the current state of the discipline, it will provide an historical analysis of why the computational approach did not take off as a standard analytical method in international relations. Second, I will assess where I (and the AI/IR field more generally) got things right, and where we got them wrong. Finally, I will say a few words about the two computational methods that have entered the mainstream of the field, at least for purposes of applied analysis, since 1995: automated coding and hidden Markov models.

The Demise of the Computational Modeling Approach

The end of computational modeling effort as a collective endeavor came in the early 1990s and was due to three factors, one self-inflicted and two external. The obvious external factor was the well-documented second "AI winter"—the over-all decline in interest in AI methods that followed the excess of hype and unfulfilled promises of the early 1980s (Hiltzik 2002). Except as a field of academic research, AI as a discrete enterprise largely collapsed, particularly when off-the-shelf computers became quite capable of handling most AI applications, thus eliminating the market for specialized "AI" hardware whose premium prices had financed much of the specialized software research outside the academy. The idea that a mass-market bookstore such as Borders would not only have an entire section devoted to "artificial intelligence" but that one could track the nuances of the field by looking at the featured titles (Chapter 6, footnote 2) now seems totally anachronistic. AI specialty firms such as Symbolics and Thinking Machines went bankrupt in the early 1990s, and by the mid-1990s, attention and venture capital funding had shifted to an entirely new target (and bubble…), the Internet.

The great irony of the second AI winter is that in contrast to the first AI winter in the 1970s—the response to hype and unfilled promises in the 1960s—many of the core technologies developed in the 1980s actually worked. The apparent "collapse" was simply the transition of these techniques from the phase of esoteric research to one of applied engineering. In fact, the 1990s saw many of the promises of the first AI boom of the 1960s fulfilled through a combination of massive increased computing power and the accumulation of incremental research advances. In 1997, IBM's massively parallel computer Deep Blue defeated Gary Kasparov, the reigning human grandmaster in chess, and chess playing programs are now simply an interesting gimmick included with most computers as free software. Okay, so AI pioneer Herbert Simon had predicted in 1957 such a defeat would occur by 1967 (Hiltzik 2002: 49), but better late than never.

Automated translation of human language, while hardly flawless, is now available at the click of a mouse on many web sites; in industrialized countries one now routinely interacts with voice recognition. Rule-based systems and neural networks are now routinely used in industrial and consumer applications, and neural networks (and data-mining tools) are now part of several statistical packages, notably SPSS. Lenat's CYC project (http://www.cyc.com) for codifying common sense is still going, and has now accumulated a database of some 1.5-million assertions.

This mainstreaming of AI, however, has generally not affected quantitative analysis in international relations, with the partial exception of neural network models and automated coding. The reasons are due to a combination of the mindless self-destruction of the core AI/IR research group in the early 1990s, and the "competitive exclusion" of the AI/IR by other, quite legitimate, methodological issues and interests during the remainder of that decade.

The self-destructive component was the emergence of an intense interest by several individuals in the AI/IR group in "post-modern deconstructionism", a short-lived academic fad that tore through the humanities and social sciences in the early 1990s with the virulence and intellectual coherence of a stomach flu epidemic in a day-care center (Sokal and Bricmont 1999). Significant segments of an aging Baby Boomer generation, finding that they had nothing new to contribute, solipsistically concluded that no one else did either, and found solace in a nihilistic creed originally propounded by former Nazi collaborators following World War II.

The results within the AI/IR group were ugly: the "mean green meme" of post-modernism (Wilber 2002) did not brook criticism or alternative agendas lightly. All ideas were equally valid, except ideas that suggested that all ideas might not be equally valid. Collaborative grants for computational IR research were hijacked and re-directed to purely post-modernist agendas, and individuals with quaint notions that computational IR research might involve, say, writing code and analyzing data, were—to use the 21st century term—"voted off the island." The AI/IR panels that had once been a well-attended feature of the International Studies Association and American Political Science Association meetings came to a halt along with any further attempts at collaborative fund raising, and the collective enterprise died. Several key researchers dropped out of the academy altogether; none produced a generation of graduate students trained to carry on the methodology.

In some small way this fate was perhaps deserved: live by the fad, die by the fad. There was even a tenuous—very tenuous—link between the natural language processing focus of some of the AI/IR research and the dubious contention of post-modernists as to the infinite flexibility of humans to "construct" any world by words alone. This option would be news to anyone scrapping together a livelihood in a developing country or trapped amid a maze of military checkpoints[1], but is apparently how the world works for those comfortably ensconced with a tenured position or a fat trust fund. As the post-modernists typically spent most of their time in libraries, chatter-filled seminar rooms, and self-reinforcing conferences rather than in the field watching actual political events, their failure to note these inconvenient realities can be understood, if not excused.

While the self-inflicted wound of post-modernism was the death blow to the once-promising AI/IR effort, that agenda also failed to compete successfully in the more general marketplace of reasonably sane ideas. Two points are particularly noteworthy.

[1] Been there, done that…

First, from the perspective of international relations, statistical research during the decade of the 1990s was dominated by a single theoretical issue: establishing first the validity of the "democratic peace" hypothesis—democratic states do not fight wars against each other, although they fight wars in general with the same frequency as non-democratic states—and later trying to ascertain whether the true mechanism was the democratic peace or the "liberal peace"—wealthy states with strong trade linkages do not fight wars with each other (Schneider, Barbieri and Gleditsch 2003; Bennett and Stam 2004). These questions could be addressed, albeit imperfectly (they are difficult questions given the available data) with existing statistical methods and, with the possible exception of neural networks, the computational methods had little to contribute that would justify the necessary start-up costs.

Second, the political methodology community recognized the new worlds opened up by increased computing power, but directed this to the estimation of models that were grounded in the older statistical paradigms. In the first half of the 1990s focused on the estimation of maximum likelihood models of ever-increasing complexity (King 1989b) using numerical approximation methods; by the late 1990s and early 2000s, attention had turned to Bayesian methods (Gill 2002).

Where we are now

This section will not attempt to review all of the methods discussed in the original text—that would involve the update that I wished to avoid—but instead focus on various issues that I think a reader in the mid-2000s should be aware of before trying to proceed further with the methods

Application of basic computational methods

In terms of the basic methodologies, the core repertoire of computational methods outlined in this document—rule-based systems, ID3, genetic algorithms, neural networks, and inductive cluster analysis—remain the core. To the best of my knowledge—and since I haven't been following this area as closely as I once did, I would be delighted to hear of additional approaches—these remain the core *computational* methods with the exception of hidden Markov models, which I discuss below.

In terms of applications, the only methods that has achieved significant legitimacy is the neural network, but that has done well, both through the championing of Zeng, King and Beck, (Zeng 1999, Beck, King, and Zeng 2000) and through applications in the State Failures project (Esty et al 1995, 1998). But despite these high-visibility efforts, neural network models are still encountered only rarely in comparison to standard linear methods. If Harvard's Gary King, arguably the most influential political methodologist of his generation, can't legitimize neural networks with an article in the *American Political Science Review*, arguably the most visible journal in the political science profession, it is clear that they aren't going to catch on quickly…

Critique of rational choice

My critique of rational choice modeling, while hardly original, is certainly far more central today than it was ten years ago, and a small-scale industry—the obnoxiously (but accurately) named "post-autistic economics" movement (Fullbrook 2003; http://www.paecon.net/)/ has grown up attacking the core assumptions of rational choice, largely on experimental grounds. The intellectual investment in rational choice modeling is sufficiently great that I have no expectation that it will disappear—among other things, far too many of its adherents have tenure, and like the mean green meme of post-modernism, rational choice advocates do not take kindly to criticism—but I would gradually expect it to die away. It is one thing to adopt a theoretical approach that appears to be an over-simplification—all theories simplify—but quite another to buy into a set of assumptions that have been demonstrated to be dead wrong through a series of elegant and readily replicated experiments, and through some of the advanced technology ever used to study human thought processes, functional MRI scanning.

If current trends in academic economics continue, the next generation of political science graduate students will duly be taught rational choice methods, and dutifully evidence an interest in it through the period of qualifying exams. After that point, the weakest students will continue to adhere to the received wisdom, and bask in the unaccustomed praise they receive from certain senior and quite distinguished figures in the profession. The average young scholar will quietly abandon the approach, viewing rational choice as one of those puzzling Boomer affectations, comparable to love-beads, bell-bottoms, disco music, post-modern deconstructionism, and the Boomers' incomprehensible aversion to tattoos and body piercing. The best and the brightest intellects will, with reckless but enthusiastic glee, aggressively attack the weakened paradigm with every tool at their disposal, competing for the honor of driving the final stake through its blackened heart.

In the end, the puppy dies.[2] From the ashes is likely to emerge something that will still claim to be "rational choice" but the revised assumptions—presumably incorporating the fact that humans are motivated at least as much by social interactions entailing cooperation, altruism, and justice/revenge as by autistic self-interest—and methods (e.g. evolutionary game theory) will be so different that it is effectively a new theory.

What was missed

There are two clear misses in my treatment, though it is possible that each might be corrected in the future, albeit from two different directions.

First, the sequence analysis approach—with the exception of HMMs—has been a complete dead-end: no one has followed up any of these techniques. This is true despite the fact that sequence analysis in the cognate field of biology has exploded with the introduction of inexpensive gene sequencing methods. There are now books written about—and academic positions advertised for—"computational biology," (Gibas and Jambeck 2001, Markel and León 2003) but none of these methods have been applied in the analysis of political behavior. Nor, perhaps, should they be—this may have been too great a conceptual leap.

Second, and more surprisingly, cluster analytic methods still remain at the fringe as an analytical technique, despite the expansion of "data mining" methods—most of which are variants on cluster analysis—in business applications and the ready availability of

[2] Puppy?—more like a geriatric pit bull…

software that implements these methods in a user-friendly (perhaps too user-friendly…) fashion.

This is probably due to two factors. The first is the aforementioned problem of competitive exclusion—the increased computational capacity required for data mining has been used instead for numerical analysis in MLE and Bayesian approaches. Second—and probably even more important—is the unfortunate association of the term "data mining" with atheoretical statistical fishing expeditions, which might better characterized as computer-assisted statistical self-deception (SASSD?— sassed…one would do worse in terms of acronyms). Since failure to adequately answer the question "where's the theory?" is the kiss-of-death in job interviews and article reviews, this reluctance is understandable.

From the perspective of this presentation, that is unfortunate, since I have argued that cluster-like generalization is probably one of the most commons forms of human inductive reasoning. Furthermore, the past forty years of quantitative international relations research have many of the elements of a *collective* fishing expedition—once a data set is made available, one tends to see virtually every possible subset of the data tested for something, often with the most tenuous of theoretical justifications (and this involves only the published literature—vastly more hypotheses have been tested, failed, and left in the proverbial "file drawer"). It would, frankly, be a whole lot more efficient simply to do all of this once and be done with it, but we are still in a situation where that which is permissible for the collective is not permissible for the individual.

A resolution of this paradox may, perhaps ironically, eventually come out of the new efforts to systematize "small N" qualitative research. In that mode, high-dimensional case descriptions are used to subset cases, placing "like with like" and from those categorizations, coherent theories are produced. Human analysts have been doing this for years, and as systematic methods are developed for doing this—work on this is just beginning, in my opinion—it is likely to result in something that is similar to cluster analysis.

What is new

Two new methods have emerged since 1995 that are at least the focus of multiple research efforts, if still hardly mainstream. The most significant of these has been the emergence of machine-coded event data that attained legitimacy in the late 1990s. The KEDS project that was mentioned in Chapter 7 continued through that decade, and in the spring of 2000, produced a new automated coding system named TABARI—Textual Analysis By Augmented Replacement Instructions (http://www.ku.edu/~keds/software.dir/tabari.html)—that is based on the same sparse-parsing principles as KEDS (and hence can use dictionaries developed for KEDS) but is far faster and more flexible. KEDS was written in Pascal and worked only on the Macintosh operating system; TABARI is written as open-source code in ANSI C++ and is available on the Linux, Macintosh, and Windows operating systems. TABARI eliminates some deep-seated idiosyncrasies of KEDS and is about 70-times faster on equivalent machines, reducing the time required to recode a data set from hours to minutes or even seconds. Using this technology, we've produced about two dozen regional event data sets, primarily focusing on areas experiencing protracted conflict (http://www.ku.edu/~keds/data.dir/).

During this same period, the consulting firm Virtual Research Associates (VRA; http://vranet.com) independently developed another automated coding system, the VRA Reader (http://vranet.com/productsRead.html), which is based on full parsing, a very different natural language processing approach than that used in KEDS and TABARI. VRA has produced data for a number of government and NGO clients, and has released a global event data set available at http://gking.harvard.edu/files/bonds.zip containing about 3.7-million events based on Reuters reports from 1 January 1991 to 31 December 2000.

The KEDS project originally became involved with machine coding because, after initial start-up costs, it is dramatically faster and less expensive than human coding. Once a researcher has established vocabulary lists of actors and verb phrases, the only significant expense involved in generating event data is the acquisition of machine-readable news reports. Furthermore, a coding system developed at one institution can be used by other researchers through the sharing of vocabulary lists and coding software; this has been part of our collaboration with the PANDA project.

In working with KEDS, we discovered two additional advantages to machine coding. First, it is free of non-reproducible coding biases and is therefore both reliable and transparent. Human coding is subject to systematic biases because of unconscious assumptions made by the coders. For example, Laurance (1990) notes that even expert coders in the military tended to over-estimate the military capability of China in the 1980s because they knew China to be a large Communist country. When students do event coding part-time, coder biases are even more unpredictable and difficult to control. In contrast, with machine-coding the words describing an activity will receive the same code irrespective of the actors or time period involved. Any biases embedded in the machine coding system are preserved explicitly in its vocabulary and can be modified by the researcher; there is no such record in human coding and thus no ability to address this potential problem.

Second, machine coding allows the researcher to experiment with alternative coding rules that reflect a particular theoretical perspective or interest in a specific set of issues. Using contemporary equipment and software, our 200,000-event Arab-Israeli data set can be completely recoded in about thirty seconds. Historically, the most commonly used event data sets for international relations research have been Azar's (1982) Conflict and Peace Data Bank (COPDAB) and McClelland's (1976) World Event Interaction Survey (WEIS). These were both developed during the Cold War and assume a "Westphalian-Clausewitzian" political worldview of sovereign states reacting to each other through diplomacy and military threats; they are ill suited to dealing with ethnic conflict, low-intensity conflict, or multilateral intervention. With machine coding, alternative coding schemes can be implemented and refined with relative ease, as the PANDA project has already demonstrated.

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

With high-capacity computers, automated coding proved to be an imminently tractable problem (King and Lowe 2003), and the automated coding of event data has been accepted as both a viable—and in most cases, preferable—alternative to traditional human coding. In 2004, the last large-scale event data project using human coding—the GEDS effort at the University of Maryland—was closed down after an after its sponsor, the Political Instability Task Force, shifted to automated coding following an direct competition between GEDS and the TABARI project at Kansas. It's over—the machines have won, in this story ol'John Henry got beat by the steam engine and there's no looking back.

The second development—and really the only major methodological technique—has been the use of hidden Markov models (Bond et al 2004; Rabiner 1989; Schrodt 1999, 2000). These were another sequence analysis method originally developed in linguistics, and are widely used in speech recognition devices.

The HMM has several advantages over alternative models for sequence comparison. First, the structure of the model is relatively simple and the number of parameters is proportional to the number of Markov chain states (typically around 6) times the number of event types, whereas in the Levenshtein metric it is proportional to the *square* of the number of event types. HMMs can be estimated very quickly, in contrast to neural networks and genetic algorithms. The HMM model, being stochastic rather than deterministic, is specifically designed to deal with noisy output and with indeterminate time (see Allan 1980); both of these are present in international event sequences. An important advantage of the HMM, particularly in terms of its possible acceptability in the policy community, is that it can be *trained by example*: a model that characterizes a set of sequences can be constructed without reference to any preconceived underlying rules other than those implicit in the model itself.

Consistent with the sequence analysis methods developed in this book, but in contrast to virtually all of the statistical analysis of event data, HMMs do not require the interval-level aggregative methods using event data scales such as those proposed by Azar and Sloan (1975) or Goldstein (1992). These scales, while of considerable utility, assign weights to individual events in isolation and make no distinction, for example, between an accusation that follows a violent event and an accusation during a meeting. The HMM, in contrast, dispenses with the aggregation and scaling altogether—using only the original, disaggregated events—and models the relationship between events by using different symbol observation probabilities in different states.

The HMM also requires no temporal aggregation. This is particularly important for early warning problems, where critical periods in the development of a crisis may occur over a week or even a day. Finally, indeterminate time means that the HMM is relatively insensitive to the delineation of the start of a sequence: It is simple to prefix an HMM with a "background" state that simply gives the distribution of events generated by a particular source (e.g. Reuters/WEIS) when no crisis is occurring and this occurs in the models estimated below. A model can simply cycle in this state until something important happens and the chain moves into later states characteristic of crisis behavior.

There is, in principle at least, a clear probabilistic interpretation to each of the parameter matrices, which allows them to be interpreted substantively. More generally, there is clear probabilistic interpretation of the model that uses familiar structures and concepts such as probability vectors, maximum likelihood estimates and the like. Finally—and not insignificantly—the technique has already been developed and is an active research topic in a number of different fields. The breadth of those applications also indicates that the method is relatively robust. While there is always a danger in applying the *technique du jour* to whatever data on political behavior happen to be laying around, the HMM appears unusually well suited to the problems of generalizing and classifying international event data sequences, a task for which there are at present no particularly satisfactory solutions.

However, HMMs have proven disappointing in two regards. First, none of the methods that I am familiar with—and I have experimented with quite a few variations—produce unique parameter estimates, though it is unclear whether this is due to the equations being under-identified or to the problems of maximizing a high dimensional service with a large number of local maxima. Second, HMMs, like neural networks, employ a diffuse parameter structure that is very difficult to meaningfully interpret in practice. As a consequence of these two problems, my own assessment is that HMMs, while clearly of considerable practical interest, are still not the holy grail of sequence analysis methods that I had once thought they might be.

Where do we go from here? (reprise)

Of the five "grand challenges" I posed in 1995, only the first—real-time forecasting—has been implemented. That one has seen considerable success, although many of the applications are largely hidden behind the veil of security classification. Nonetheless, all indications are that the neural network models developed for the State Failures Project have been used for on-going monitoring; I have every reason to believe (based on a number of conversations with individuals who have no reason to want to impress or deceive me) that hidden Markov models are being used at least experimentally, and there is at least one published academic effort (Pevehouse and Goldstein 1999) to this effect, albeit using statistical methods.

The real-time forecasting efforts are likely to continue due to a combination of the potential utility of the method, and the now readily-available dense, contemporaneous event data sets produced with automated coding systems in combination with machine-readable newswire reports on services such as NEXIS, FBIS, and Factiva, all easily accessible at North American research universities. While this methodology existed in 1995, it had yet to gain full legitimacy in the academic community, a situation that has now changed. These new event data sets provide far greater detail than the older, human-coded WEIS and COPDAB sets used in the studies in this book, and also have the potential of providing easily customized coding frameworks that can be optimized for special problems. In 1995, I was impressed that these methods could code 15 events per second—quite an improvement over the human coding norms of six to ten events per hour—but coding speeds for TABARI are now on the order of 10,000 events per second. While the various coding programs are straightforward applications of natural language processing methods, the automated coding systems are, in a sense, "artificial intelligence" in that they are substituting computational methods to do a task once done by humans.

To a much more limited extent, there has been some progress on the simulation of international behavior through agent-based models. The agent-based modeling approach in general has taken off (Axelrod 1997; http://www.econ.iastate.edu/tesfatsi/ace.htm), complete with its own journal (http://jasss.soc.surrey.ac.uk/JASSS.html) and IR-oriented applications of the approach have made it into the *APSR* (Cederman 2003). I also have the impression—albeit I make this assertion with far less confidence than what I said about forecasting, being much further from the field—that serious experimentation has been done in the policy community with the method. However, it is certainly not mainstream: again, the cost of entry—computer programming—remains high despite the availability of multiple well-developed software suites for agent-based modeling.

Little or no progress has been made on the remaining three challenges, and the barrier in all three cases is largely the same: insufficient progress in semantic natural language processing. Nonetheless, as ever-greater amounts of natural language materials become available on the web, and as ever more clever methods are found for making use of that material, we may yet see progress. In particular, the U.S. government appears to be once again getting serious about funding systematic social science research related to international conflict, after a nearly two-decade absence, and this might provide the resources needed to tackle these difficult problems.

The capacity of computing power continues to climb exponentially, with limits still barely in sight. What seemed in 1995 to be a phenomenal increase in power over the Apple II of 1980 has been equally dwarfed by subsequent developments. As anticipated in Chapter 7, the latest generation of Apple computers,[3] the G5 series, was in fact advertised (with justification) as a supercomputer, and—contrary to Apple's advertising—comparable power is available in Intel systems. As I write this, the Sunday paper is advertising an Intel-based computer with a 32-bit 2.7 Ghz processor, 256 Mb RAM, 40 Gb hard drive machine for only $370, less than 1% of an assistant professor's salary.

The emergence of the open-source software movement (which both drives and is driven by the World Wide Web, an institution which barely existed when the manuscript was first completed) and the consolidation of operating environments to two systems—Microsoft's Windows and the Unix variants of everyone else—has made software far more accessible to programmers, though tools have yet to emerge (despite continual promises) that make programming any easier for computational tools.[4] The plummeting cost of high-capacity off-the-shelf hardware has opened up the possibility of massively parallel systems—usually driven by the Unix-based Beowulf system—at a fraction of the cost of the older "supercomputers".

The proposition that increased computing power alone will make a difference is still controversial. Part of this is probably a function of simple pride—we no more want our intellects to be displaced by mere machines (even machines our intellects brought into being) than old John Henry wanted to be replaced by that steam drill—but for the

---

[3] Even more remarkable is that a quarter-century after I did the first analysis on an Apple II, I'm still writing this on an Apple computer—a G5 in fact—despite Apple's organizational near-death-experience in the mid-1990s and the continuing fervent desire of the University of Kansas computer "support" that all Macintoshes would go away and we would all bow before the sacred—if virus-ridden—altar of Microsoft.

[4] This despite the fact that such tools probably *could* be created. For example, in my experience, the perl programming language reduces the size of text processing programs by at least a factor of ten compared to their C/C++ equivalents; Gauss and Mathematica perform a similar task for algorithms dependent on matrix operations.

time being, it can also be quite legitimately contested on the grounds that nothing has been demonstrated.

But the issue may still be open, and there are at least two instances where increased power has made a qualitative difference. Gary Kasparov was finally defeated in chess by fast hardware, not fancy programming, and the current Bayesian revolution in statistical methodology is primarily fueled by the fact that *finally* we can estimate the darn equations; the *theory* of Bayesian statistical estimation having been around since the 1950s. There are at least two issues in computation modeling of political behavior that clearly require massive computing power—semantic content analysis and partially-ordered event sequence recognition—and it is possible that further increases in power, combined with some supporting algorithms, will result in breakthroughs in those areas. Increased power could also simplify the application of existing computationally-intensive methods such as neural networks and cluster analysis, but these are already accessible with existing hardware and software to all but the most impatient analysts, and do not appear to be catching on.

Bibliography

Abelson, Robert P. 1973. "The Structure of Belief Systems." In *Computer Models of Thought and Language*, edited by R.C. Schank and K.M. Colby. San Francisco: Freeman.

Abrams, Robert. 1980. *Foundations of Political Analysis.* New York: Columbia University Press.

Achen, Christopher. 1995. "How Can We Tell a Unitary Rational Actor when We See One?" Paper presented at the Midwest Political Science Association, Chicago.

Aldenderfer, Mark S. and Roger K. Blashfield. 1984. *Cluster Analysis.* Beverly Hills: Sage Publications.

Aldrich, John H. and Forrest D. Nelson. 1984. *Linear Probability, Logit and Probit Models*. Beverly Hills: Sage Publications.

Alker, Hayward R. 1987. "Fairy Tales, Tragedies and World Histories: Towards Interpretive Story Grammars as Possibilist World Models." *Behaviormetrika* 21: 1-28.

Alker, Hayward R., James Bennett and Dwain Mefford. 1980. "Generalized Precedent Logics for Resolving Security Dilemmas." *International Interactions* 7:165-200.

Alker, Hayward R. and C. Christensen. 1972. "From Causal Modeling to Artificial Intelligence: The Evolving of a UN Peace-Making Simulation." In *Experimentation and Simulation in Political Science*, edited by J.A. LaPonce and Paul Smoker. Toronto: University of Toronto Press.

Alker, Hayward R. 1975. "Polimetrics: Its Descriptive Foundations," In *Handbook of Political Science*, edited by Fred Greenstein and Nelson Polsby. Reading, MA: Addison-Wesley.

Alker, Hayward R. and W. Greenberg. 1977. "On Simulating Collective Security Regime Alternatives." In *Thought and Action in Foreign Policy*, edited by G. Matthew Bonham and Michael J. Shapiro. Basel: Birkhauser Verlag.

Allais, M. 1953. "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'Ecole Américaine." *Econometrica* 21:503-46.

Allan, Pierre. 1980. "Diplomatic Time and Climate: A Formal Model." *Journal of Peace Science* 4:133-150.

Allison, Graham T. 1971. *The Essence of Decision.* Boston: Little, Brown.

Almond, Gabriel. 1988. "Separate Tables: Schools and Sects in Political Science." *PS* 21:828-824.Almond, Gabriel. 1990. *A Divided Discipline: Schools and Sects in Political Science.* Sage Publications.

Alvarado, Sergio J. 1990. *Understanding Editorial Text: A Computer Model of Argument Comprehension*. Boston: Kluwer Academic Publishers.Anderson, James A. and Edward Rosenfeld. eds. 1988. *Neurocomputing: Foundations of Research*. Cambridge: MIT Press

Anderson, John R. 1983. *The Architecture of Cognition.* Cambridge: Harvard University Press.Anderson, Paul A. 1981. "Justifications and Precedents as Constraints in Foreign Policy Decision-Making." *American Journal of Political Science* 25:738-761.

Anderson, Paul and Stuart Thorson. 1982. "Artificial Intelligence Based Simulations of Foreign Policy Decision-Making." *Behavioral Science* 27:176-193.Anderson, Philip W., Kenneth J. Arrow and David Pines, eds. 1988. *The Economy as an Evolving Complex System.* New York: Addison Wesley

Anderton, Charles H. 1985. "Arms Race Modeling: Categorization and Systematic Analysis." Paper presented at the International Studies Association, Washington.Andriole, Stephen J. and Gerald W. Hopple. 1984. "The Rise and Fall of Events Data: From Basic Research to Applied Use in the U.S. Department of Defense." *International Interactions* 11:293-309.

Andriole, Stephen J. and Gerald W. Hopple. 1988. *Defense Applications of Artificial Intelligence.* Lexington MA: Lexington.

Archibald, G.C. and Richard G. Lipsey. 1976. *An Introduction to Mathematical Economics*. New York: Harper and Row.

Ashby, Ross. 1956. *Introduction to Cybernetics.* London: Chapman and Hall.

Axelrod, Robert, ed. 1976. *Structure of Decision*. Princeton: Princeton University Press.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Axelrod, Robert. 1987. "The Evolution of Strategies in the Iterated Prisoners' Dilemma." In *Genetic Algorithms and Simulated Annealing*, edited by Lawrence Davis. Los Altos, CA: Morgan Kaufmann.

Axelrod, Robert. 1997. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton: Princeton University Press.

Azar, Edward E. 1980. "The Conflict and Peace Data Bank (COPDAB) Project." *Journal of Conflict Resolution* 24:143-152.

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

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

Azar, Edward. E., and Thomas Sloan. 1975. *Dimensions of Interaction*. Pittsburgh: University Center for International Studies, University of Pittsburgh.

Bailey, Kenneth D. 1995. *Typologies and Taxonomies: An Introduction to Classification Techniques*. Thousand Oaks, CA: Sage Publications

Banerjee, Sanjoy. 1986. "The Reproduction of Social Structures: An Artificial Intelligence Model." *Journal of Conflict Resolution* 30:221-252.

Banks, Jeffery S. 1991. *Signaling Games in Political Science*. New York: Gordon Breach.

Barr, Aaron, Paul R. Cohen and Edward A. Feigenbaum. 1982. *The Handbook of Artificial Intelligence*. Los Altos, CA: William Kaufmann.

Barrow, John D. and Frank J. Tipler. 1986. *The Anthropic Cosmological Principle.* Oxford: Oxford University Press.

Barry, Brian and Russell Hardin, eds. 1982. *Rational Man and Irrational Society*. Beverly Hills: Sage.

Bartholomew, D. J. 1971. *Stochastic Models for Social Processes.* New York: Wiley.

Beck, Neal, Gary King and Lengche Zeng. 2000. "Improving quantitative studies of international conflict: A conjecture." *American Political Science Review* 94, 1: 21-35.

Becker, Gary S. 1976. *The Economic Approach to Human Behavior.* Chicago: University of Chicago Press.

Benfer, Robert A., Edward E. Brent, Jr. and Louanna Furbee. 1991. *Expert Systems*. Newbury Park CA: Sage Publications.

Bennett, D. Scott and Philip A. Schrodt. 1987. "Linear Event Patterns in WEIS Data." Paper presented at the American Political Science Association, Chicago.

Bennett, D. Scott and Allan C. Stam. 2004. *The Behavioral Origins of War*. Ann Arbor: University of Michigan Press.

Bercovitch, Jacob and Jeffery Langley. 1993. "The Nature of Disputes and the Effectiveness of International Mediation." *Journal of Conflict Resolution* 37:670-691.

Blainey, Geoffrey. 1988. *The Causes of War*. New York: Free Press.

Bloomfield, Lincoln P. and Allen Moulton. 1989. *CASCON III: Computer-Aided System for Analysis of Local Conflicts.* Cambridge: MIT Center for International Studies.

Bobrow, Davis B., Donald A. Sylvan and Brian Ripley. 1986. "Japanese Supply Security: A Computational Model." Paper presented at the International Studies Association, Anaheim.

Bond, Doug, J. Craig Jenkins, Charles L. Taylor and Kurt Schock. 1997. "Mapping Mass Political Conflict and Civil Society: The Automated Development of Event Data" *Journal of Conflict Resolution* 41,4:553-579.

Bonham, G. Matthew and Michael J. Shapiro. 1976. Explanation of the Unexpected: the Syrian Intervention in Jordan in 1970, pp. 113-141. In *The Structure of Decision*, edited by Robert Axelrod. Princeton: Princeton University Press.

Boynton, G. R. 1991. "The Expertise of the Senate Foreign Relations Committee." In *Artificial Intelligence and International Politics*, edited by Valerie Hudson. Boulder: Westview.

Boynton, G. R. n.d. “Interpretive Triples and Reasoning About Foreign Policy.” Photocopied manuscript.

Brams, Steven J. 1985. *Superpower Games.* New Haven: Yale University Press.

Brams, Steven J. and D. Marc Kilgour. 1988. *Game Theory and National Security.* New York: Basil Blackwell.

Bremer, Stuart A. ed. 1987. *The GLOBUS Model.* Frankfurt: Campus/Westview.

Brewer, Garry D. and Martin Shubik. 1979. *The War Game: A Critique of Military Problem-Solving.* Cambridge: Harvard University Press.

Brito, D. L. and Michael D. Intriligator. 1974. “Uncertainty and the Stability of the Armaments Race.” *Annals of Economic and Social Measurement* 3:279-292.

Brown, Anthony Cave. 1975. *Bodyguard of Lies.* New York: Harper and Row.

Brunvald, Jan Harold. 1984. *The Choking Doberman and other “New” Urban Legends.* New York: Norton.

Brush, Stephen G. 1989. “Prediction and Theory Evaluation: The Case of Light Bending.” *Science* 246:1124-1129.

Buchanan, James M. and Gordon Tullock. 1962. *The Calculus of Consent.* Ann Arbor: University of Michigan Press.

Buchanan, James M. 2000. “Saving the Soul of Classical Economics.” *Wall Street Journal*, January 1st 2000:

Buchanan, Bruce G. and Edward H. Shortliffe, eds. 1984. *Rule-based Expert Systems*. Reading, MA: Addison-Wesley.

Buchanan, Bruce G. and Reid G. Smith. 1989. “Fundamentals of Expert Systems.” In *The Handbook of Artificial Intelligence*, vol 4, edited by Avron Barr, Paul R. Cohen and Edward A. Feigenbaum. Reading, MA: Addison-Wesley.

Bueno de Mesquita, Bruce. 1981. *The War Trap.* New Haven: Yale University Press.

Bueno de Mesquita, Bruce. 1984. “A Critique of ‘A Critique of *The War Trap.*’” *Journal of Conflict Resolution* 28:341-360.

Bueno de Mesquita, Bruce. 1989. “The Contribution of Expected-Utility Theory to the Study of International Conflict“ In *Handbook of War Studies*, edited by Manus Midlarksy. Boston: Unwin, Hyman.

Bueno de Mesquita, Bruce. 1985. “The War Trap Revisited: A Revised Expected Utility Model.” *American Political Science Review* 79:156-177.

Bueno de Mesquita, Bruce, David Newman and Alvin Rabushka. 1985. *Forecasting Political Events: The Future of Hong Kong*. New Haven: Yale University Press.

Bueno de Mesquita, Bruce, and David Lalman. 1992. *War and Reason.* New Haven: Yale University Press.

Bull, Hedley. 1966. “International Theory: The Case for a Classical Approach.” *World Politics* 18:361-377.

Burdick, Eugene and Harvey Wheeler. 1962. *Fail-Safe.* New York: McGraw-Hill.

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

Busch, Peter. 1970. “Appendix on mathematical models of arms races.” In *What Price Vigilance*, by Bruce M. Russett. New Haven: Yale University Press.

Butterworth, Robert Lyle. 1976. *Managing Interstate Conflict,1945-74: Data with Synopses.* Pittsburgh: University of Pittsburgh University Center for International Studies.

Carbonell, Jaime G. 1978. “POLITICS: Automated Ideological Reasoning.” *Cognitive Science* 2:27-51

Cassidy, John. 1996. “The Decline of Economics.” *New Yorker*, 2 December: 50-60.

Casti, John. 1989. *Alternate Realities: Mathematical Models of Nature and Man*. New York: Wiley.

Caudill, Maureen. 1989. *Neural Networks Primer.* San Francisco: Miller Freeman Publications.

Caudill, Maureen. 1990. “Using Neural Nets: Making an Expert Network.” *AI Expert* (July, 1990):45-49

Cederman, Lars-Erik. 2003. "Modeling the Size of Wars: From Billiard Balls to Sandpiles." *American Political Science Review* 97,1: 135-150.

Chandrasekaran, B. 1986. "Generic Tasks in Knowledge-Based Reasoning: High-Level Building Blocks for Expert System Design." *IEEE Expert* 1:23-30.

Charnick, Eugene and Drew McDermott. 1985. *Introduction to Artificial Intelligence*. Reading, MA: Addison-Wesley.

Chase, W.G. and H.A. Simon. 1973. "Perception in Chess." *Cognitive Psychology* 4:55-81.

Chase-Dunn, Christopher K. 1979. "Comparative Research on World-System Characteristics." *International Studies Quarterly* 23:601-623.

Churchill, Winston. 1948. *The World Crisis, 1911-1918*. Boston: Houghton Mifflin.

Cimbala, Stephen. 1987. *Artificial Intelligence and National Security.* Lexington, MA: Lexington Books.

Cioffi-Revilla, Claudio A. 1979. *Mathematical Models in International Relations: A Bibliography.* Institute of Research in Social Science, Technical Paper no. 4. Chapel Hill: University of North Carolina.

Cioffi-Revilla, Claudio A. 1981. "Fuzzy Sets and Models of International Relations." *American Journal of Political Science* 25:129-159.

Clancy, Tom. 1986. *Red Storm Rising*. New York: Putnam.

Clocksin, W.F. and C.S. Mellish. 1981. *Programming in Prolog*. New York: Springer-Verlag.

Coase, Ronald H. 1994. *Essays on Economics and Economists*. Chicago and London: Chicago University Press.

Cohen, Paul R. and Edward A. Feigenbaum, eds. 1982. *The Handbook of Artificial Intelligence*, Vol. 3. Los Altos, CA: William Kaufmann.

Collins, Allan and Edward E. Smith, eds. 1988. *Readings in Cognitive Science: A Perspectice from Psychology and Artificial Intelligence*. San Mateo CA: Morgan Kaufmann.

Cooper, Chester L. 1978. "Micawber versus de Jouvenel: Planning and Forecasting in the Real World of Washington." In *Forecasting in International Relations* edited by Nazli Choucri and Thomas W. Robinson. San Francisco: Freeman.

Crawford, Sue E. S. and Elinor Ostrom. 1995. "A Grammar of Institutions." *American Political Science Review* 89:582-600.

Crevier, Daniel. 1993. *AI: The Tumultuous History of the Search for Artificial Intelligence.* New York: Basic Books.

Cusack, Thomas R. and Richard J. Stoll. 1990. *Exploring Realpolitik.* Boulder: Lynne Rienner Publishing.

Cyert, Richard M. and James G. March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.

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

Davies, John, and Barbara Harff. 1998. "Dynamic Data for Conflict Early Warning." In *Risk Assessment and Crisis Early Warning Systems*, ed. John Davies and Ted R. Gurr. Pp. 79-94. New York: Rowman and Littlefield.

Davis, Lawrence, ed. 1987. *Genetic Algorithms and Simulated Annealing.* Los Altos,CA: Morgan Kaufmann.

Davis, Morton D. 1983. *Game Theory: A Nontechnical Introduction*. New York: Basic Books.

Deutsch, Karl W., Bruno Fritsch, Helio Jaguaribe and Andrei S. Markovits. 1977. *Problems of World Modeling*. Cambridge, MA: Ballinger.

Devaney, Robert L. 1986. *An Introduction to Chaotic Dynamical Systems.* Menlo Park: Cummings.

Devijver, Pierre A. and Josef Kittler. 1982. *Pattern Recognition: A Statistical Approach.* Englewood Cliffs, NJ: Prentice-Hall.

Dixon, William J. 1994."Democracy and the peaceful settlement of international conflict." *American Political Science Review* 88:1-17.

Dougherty, James E. and Robert L. Pfaltzgraff. 1981. *Contending Theories of International Relations.* New York: Harper and Row.

Downs, Anthony. 1957. *Economic Theory of Democracy.* New York: Harper and Row.

Downs, George W. and David M. Rocke. 1990. *Tacit Bargaining, Arms Races and Arms Control.* Ann Arbor: University of Michigan Press.

Doyle, Jon. 1988. "Big Problems for Artificial Intelligence." *AI Magazine* 9,2:19-22.

Duffy, Gavan. 1991. "Time-Space: Representing Historical Time for Efficient Event Retrieval." In *Artificial Intelligence and International Politics*, edited by Valerie Hudson. Boulder: Westview.

The Economist. 1992. "Minds in the Making: A Survey of Artificial Intelligence." (14 March 1992)

Ellsberg, Daniel. 1961. "Risk, ambiguity and the Savage axioms." *Quarterly Journal of Economics* 75:643-669.

Elster, Jon. 1978. *Logic and Society: Contradictions and Possible Worlds*. New York: Wiley.

Esty, Daniel C., Jack A. Goldstone, Ted R. Gurr, Pamela Surko and Alan N. Unger. 1995. *State Failure Task Force Report.* McLean, VA: Science Applications International Corporation.

Esty, Daniel C., Jack A. Goldstone, Ted R. Gurr, Barbara Harff, Marc Levy, Geoffrey D. Dabelko, Pamela Surko, and Alan N. Unger. 1998. *State Failure Task Force Report: Phase II Findings.* McLean, VA: Science Applications International Corporation.

Eves, Howard. 1983. *An Introduction to the History of Mathematics*. Philadelphia: Saunders Publishing.

Everitt, Brian. 1980. *Cluster Analysis*. New York: Wiley/Halsted.

Fain, Tyrus G., Katharine C. Plant and Ross Millroy, eds. 1977. *The Intelligence Community.* New York: R.R. Bowker.

Fayyad, Usama M., Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurusamy. 1995. *Advances in Knowledge Discovery and Data Mining.* Cambridge MA: AAAI Press/MIT Press.

Fearon, James D. 1991. "Counterfactuals and Hypothesis Testing in Political Science." *World Politics* 43,2:169-195.

Fearon, James D. 1994. "Signaling versus the Balance of Power and Interests." *Journal of Conflict Resolution* 38:236-269.

Feigenbaum, Edward A. 1983. "Knowledge Engineering: The Applied Side", pp. 37-55. In *Intelligent Systems*, edited by J.E. Hayes and D. Michie*.* Chichester: Ellis Horwood Ltd.

Feigenbaum, Edward A. and Pamela McCorduck. 1983. *The Fifth Generation.* Reading, MA: Addison-Wesley.

Fiorina, Morris. 1973. "Formal Models in Political Science." *American Journal of Political Science* 19:133-159

Forsyth, Richard and Roy Rada. 1986. *Machine Learning: Applications in Expert Systems and Information Retrieval.* New York: Wiley/Halstead.

Frank, Robert. 1987. "If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?" *American Economic Review* 77(4): 593-604.

Freeman, James A. and David M. Skapura. 1991. *Neural Networks: Algorithms, Applications and Programming Techniques.* Reading, MA: Addison-Wesley.

Frey, Bruno S. 1983. "The Economic Model of Behavior: Shortcomings and Fruitful Development." University of Zurich: Institute for Empirical Research in Economics.

Frey, Bruno S. 1992. *Economics as a Science of Human Behaviour: Towards a New Social Science Paradigm.* Boston: Kluwer Academic Publishers.

Frey, Bruno S. 2002. "Publishing as Prostitution?: Choosing Between One‘s Own Ideas and Academic Failure." Working Paper No. 117, University of Zurich: Institute for Empirical Research in Economics,

Friedman, Milton. 1953. "The Methodology of Positive Economics." In *Essays in Positive Economics*, edited by Milton Friedman. Chicago: University of Chicago Press.

Frohlich, Norman and Joe E. Oppenheimer. 1978. *Modern Political Economy.* Englewood Cliffs, NJ: Prentice-Hall.

Fu, K.S. 1974. *Syntactic Methods in Pattern Recognition.* New York: Academic Press.

Fu, K.S. 1982. *Syntactic Pattern Recognition and Applications*. New York: Prentice-Hall.

Fullbrook. Edward, ed. 2003. The Crisis in Economics. New York: Routledge.

Fürnkranz, Johannes, Johann Petrak, Robert Trappl and Jacob Bercovitch. 1994. "Machine Learning Methods for International Conflict Databases." Research Report TR-94-33, Austrian Research Institute for Artificial Intelligence, Vienna.

Gaddis, John Lewis. 1987. "Expanding the Data Base: Historians, Political Scientists and the Enrichment of Security Studies." *International Security* 12:3-21.

Gaddis, John Lewis. 1990. "Nuclear Weapons and International Systemic Stability". Occasional Paper No. 2 (January 1990), American Academy of Arts and Sciences, Cambridge, MA.

Gaddis, John Lewis. 1992/93. "International Relations Theory and the End of the Cold War." *International Security* 17:5-58.

Galbraith, John Kenneth. 1984. "Staying Awake at the Council of Foreign Relations." *Washington Monthly* 16:8.

Gans, Joshua. 2000. *Publishing Economics. Analyses of the Academic Journal Market in Economics*. Vol. Elgar. Cheltenham, UK, and Northampton, MA, USA.

Gans, Joshua S. and George B. Shepherd. 1994. "How are the Mighty Fallen: Rejected Classic Articles by Leading Economists." *Journal of Economic Perspectives* 8 (1): 165 - 180.

Garson, G. David. 1987. "The Role of Inductive Expert Systems Generators in the Social Science Research Process." *Social Science Computer Review* 5: 11-24.

Garson, G. David. 1990. "Expert Systems: An Overview for Social Scientists." *Social Science Computer Review* 8:387-410.

Garson, G. David. 1991. "A Comparison of Neural Network and Expert Systems Algorithms with Common Multivariate Procedures for Analysis of Social Science Data." *Social Science Computer Review* 9:399-434.

Gelb, Leslie H. and Richard K. Betts. 1979. *The Irony of Vietnam: The System Worked*. Washington: The Brookings Institution.

Gelman, Andrew and Gary King. 1992. "Why Do Presidential Campaign Polls Vary So Much When the Vote Is So Predictable?" Paper presented at the Midwest Political Science Association, Chicago.

George, Alexander L. 1969. "The 'Operational Code': A Neglected Approach to the Study of Political Leaders and Decision-making." *International Studies Quarterly* 13:190-222.

Gerner, Deborah J., Philip A. Schrodt, Ronald A. Francisco, and Judith L. Weddle. 1994. "The Machine Coding of Events from Regional and International Sources." *International Studies Quarterly* 38:91-119.

Giarrantano, J. and G. Riley. 1989. *Expert Systems: Principles and Programming*. Boston: PWS-KENT Publishing.

Gibas, Cynthia and Per Jambeck. 2001. *Developing Bioinformatics Computer Skills.* Sebastopol, CA: O'Reilly and Associates

Gill, Jeff. 2002. *Bayesian methods : a social and behavioral sciences approach*. Boca Raton, FL: Chapman & Hall/CRC.

Gillespie, John V., Dina A. Zinnes, G.S. Tahim, Philip A. Schrodt and R. Michael Rubison. 1976. "An Optimal Control Model of Arms Races." *American Political Science Review* 71:226-244.

Gilovich, Thomas. 1981. "Seeing the Past in the Present: The Effect of Associations to Familiar Events on Judgments and Decisions." *Journal of Personality and Social Psychology* 40:797-808.

Gleick, James. 1987. *Chaos: Making a New Science.* New York: Viking.

Gochman, Charles S. and Alan Ned Sabrosky. 1990. *Prisoners of War?* Lexington,MA: Lexington Books.

Gochman, Charles S. and Russell J. Leng. 1983. "*Realpolitik* and the Road to War: An Analysis of Attributes and Behavior." *International Studies Quarterly* 27:97-120.

Goel, Narendra S. and Richard L. Thompson. 1988. *Computer Simulations of Self-Organization in Biological Systems.* London: Croom Helm.

Goldberg, David E. 1989. *Genetic Algorithms in Search, Optimizations and Machine Learning*. Reading, MA: Addison-Wesley.

Goldhamer, Herbert and Hans Speier. 1959. "Some Observations on Political Gaming." *World Politics* 12:71-83.

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

Green, Donald P. and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory*. New Haven: Yale University Press.

Greenacre, Michael J. 1984. *Theory and Applications of Correspondence Analysis*. New York: Academic Press.

Grefenstette, John J.,ed. 1987. *Genetic Algorithms and their Applications.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Guetzkow, Harold and Joseph J. Valdez, eds. 1981. *Simulated International Processes: Theories and Research in Global Modeling*. Beverly Hills: Sage.

Hackett, John and others [sic]. 1978. *The Third World War.* London: Sidgwick and Jackson Ltd.

Hall, Edward T. 1976. *Beyond Culture*. New York: Anchor/Doubleday.

Hall, Edward T. and Mildred Reed Hall. 1987. *Hidden Differences: Doing Business with the Japanese*. New York: Anchor/Doubleday.

Hall, Stephen S. 1992. "How Technique is Changing Science." *Science* 257:344-349.

Hamburger, Henry. 1979. *Games as Models of Social Phenomena*. San Francisco: Freeman.

Handel, Michael I. 1976. *Perception, Deception and Surpise: The Case of the Yom Kippur War.* Jerusalem: Jerusalem Papers on Peace Problems.

Hanson, Norwood R. 1958. *Patterns of Discovery.* Cambridge: Cambridge University Press.

Hartley, Keith and Todd Sandler, eds. 1990. *The Economics of Defence Spending*. London: Routledge.

Hashim, Safaa and Philip Seyer. 1987. *Programming Expert Systems and Artificial Intelligence with Advanced Turbo Pascal*. Blue Ridge Summit, PA: Tab Books.

Hayes-Roth, Frederick, Donald Waterman, and Douglas Lenat. 1984. *Building Expert Systems.* Reading, MA: Addison-Wesley.

Hebb, D. O. 1949. *Organization of Behavior.* New York: Science Editions.

Hecht-Nielsen, Robert. 1990. *Neurocomputing.* Boston: Addison-Wesley.

Heise, David. 1988a. "Modeling Event Structures." *Journal of Mathematical Sociology* 13:138-168.

Heise, David. 1988b. "Computer Analysis of Cultural Structures." *Social Science Computer Review* 6:183-196.

Herspring, Dale R. 1992. "Practioners and Political Scientists." *PS* 25:554-558.

Hertz, J., A. Krough and R.G. Palmer. 1990. *Introduction to the Theory of Neural Computing.* Reading, MA: Addison-Wesley.

Hiltzik, Michael. 2002. "A.I. Reboots." *Technology Review* (March 2002). Pp. 47-55.

Hofstader, Douglas. 1985. *Metamagical Themas.* New York: Basic Books.

Hogarth, Robin M. and Melvin W. Reder, eds. 1987. *Rational Choice: The Contrast between Economics and Psychology*. Chicago: University of Chicago Press.

Holland, John H. 1975. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.

Holland, John H. 1986. "Escaping Brittleness: The Possibilities of General Purpose Algorithms Applied to Parallel Rule-Based Systems." In *Machine Learning 2*, edited by R.S. Michelski, J.G. Carbonell and T.M. Mitchell. Los Altos,CA: Morgan Kaufman.

Holland, John H., Keith J. Holyoak, Richard E. Nisbett and Paul R. Thagard. 1986. *Induction: Processes of Inference, Learning and Discovery.* Cambridge: MIT Press.

Holt, Robert T., Brian L. Job and Lawrence Markus. 1978. "Catastrophe Theory and the Study of War." *Journal of Conflict Resolution* 22:171-209

Howard, Nigel. 1971 *Paradoxes of Rationality.* Cambridge: MIT Press.

Hudson, Valerie M. 1987. "Using a Rule-Based Production System to Estimate Foreign Policy Behavior." In *Artificial Intelligence and National Security*, edited by Stephen Cimbala. Lexington, MA: Lexington Books.

Hudson, Valerie. 1991. "Scripting International Power Dramas: A Model of Situational Predisposition." In *Artificial Intelligence and International Politics*, edited by Valerie Hudson. Boulder: Westview.

Hudson, Valerie, ed. 1991. *Artificial Intelligence and International Politics*. Boulder: Westview

Hudson, Valerie, Philip A. Schrodt and Ray Witmer. 2004. "A New Kind of Social Science." Paper presented at the International Studies Association, Montréal, March 2004.

Hughes, Barry B. 1984. *World Futures: A Critical Analysis of Alternatives.* Baltimore: Johns Hopkins.

Hughes, Barry B. 1985. "World Models: The Bases of Difference." *International Studies Quarterly* 29:77-101.

Hughes, Barry B. 1993. *International Futures.* Boulder: Westview.

Hunt, Earl B., J. Marin and P.J. Stone. 1966. *Experiments in Induction.* New York: Academic Press.

Huth, Paul and Bruce Russett. 1990. "Testing Deterrence Theory: Rigor Makes a Difference." *World Politics* 42:466-501

Ingrao, Bruno and Giorgio Israel. 1990. *The Invisible Hand: Economic Equilibrium in the History of Science.* Cambridge, MA: MIT Press.

International Studies Quarterly. 1983. "Symposium: Events Data Collections." *International Studies Quarterly* 27:145-177.

Intriligator, Michael D. 1964. "Some Simple Models of Arms Races." *General Systems Yearbook* 9:143-147.

Intriligator, Michael D. 1971. *Mathematical Optimization and Economic Theory.* Englewood Cliffs: Prentice-Hall.

Intriligator, Michael D. and D.L. Brito. 1984. "Can Arms Races Lead to the Outbreak of War?" *Journal of Conflict Resolution* 28:63-84.

Iwasaki, Yumi. 1989. "Qualitative Physics" in *The Handbook of Artificial Intelligence,* vol 4, edited by Aaron Barr, Paul R. Cohen and Edward A. Feigenbaum. Reading, MA: Addison-Wesley.

Jackendoff, Ray. 1987. *Consciousness and the Computational Mind.* Cambridge: MIT Press.

Jaynes, Julian. 1976. *The Origin of Consciousness in the Breakdown of the Bicameral Mind.* Boston: Houghton Mifflin.

Jensen, Lloyd. 1972. "Predicting International Events." *Peace Research Reviews* 4:1-65.

Jervis, Robert. 1976. *Perception and Misperception in International Politics.* Princeton: Princeton University Press.

Jervis, Robert, Richard Ned Lebow and Janice Gross Stein, eds. 1985. *Psychology and Deterrence.* Baltimore: Johns Hopkins Press.

Job, Brian L, Douglas Johnson and Eric Selbin. 1987. "A Multi-Agent, Script-based Model of U.S. Foreign Policy Towards Central America." Paper presented at the American Political Science Association, Chicago.

Job, Brian L. and Douglas Johnson. 1991. "UNCLESAM: The Application of a Rule-Based Model to US Foreign Policy." In *Artificial Intelligence and International Politics*, edited by Valerie Hudson. Boulder: Westview.

Kahneman, Daniel, Paul Slovic and Amos Tversky. 1982. *Judgement Under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press.

Kandel, Abraham. 1992. *Fuzzy Expert Systems.* Boca Raton: CRC Press.

Kaplan, Morton A. 1957. *System and Process in International Politics.* New York: Wiley.

Kaplan, Morton A. 1969. "Variants on Six Models of the International System." In *International Politics and Foreign Policy*, edited by James Rosenau. New York: Free Press.

Katzenstein, Lawrence C. 1989. "PRC Relations with Hong Kong: An AI Model of Decision Making in Turbo Prolog." Paper presented at the Northeastern Political Science Association, Philadelphia.

Katzenstein, Lawrence C. 1992. "A Computational Model to Assess the Changes for Taiwan-PRC Integration." Paper presented at the International Studies Association, Atlanta.

Kauffman, Stuart A. 1988. "The Evolution of Economic Webs." In *The Economy as an Evolving Complex System*, edited by Philip W. Anderson, Kenneth J. Arrow and David Pines. New York: Addison Wesley.

Kauffman, Stuart A. 1993. *The Origins of Order.* Oxford: Oxford University Press.

Kauffman, Stuart A. and Simon Levin. 1987. "Towards a General Theory of Adaptive Walks on Rugged Landscapes." *Journal of Theoretical Biology* 128: 11-45.

Kauffman, Stuart A. and Sonke Johnsen. 1992 "Co-Evolution to the Edge of Chaos: Coupled Fitness Landscapes, Poised States and Co-Evolutionary Avalanches." In *Artificial Life II*, edited by Langton et al.. Redwood City, CA: Addison-Wesley.

Kaw, Marita. 1989. "Predicting Soviet Military Intervention." *Journal of Conflict Resolution* 33:402-429.

Kellert, Stephen H. 1993. *In the Wake of Chaos: Unpredictable Order in Dynamical Systems.* Chicago: University of Chicago Press.

Keohane, Robert O. 1988. "International Institutions: Two Approaches." *International Studies Quarterly* 32:379-396.

Kerr, Richard A. 1989. "A New Way to Forecast Next Season's Climate." *Science* 244 (7 April 1989),30-31.

Khong, Yuen Foong. 1992. *Analogies at War*. Princeton: Princeton University Press.

Kimber, Richard. 1991. "Artificial Intelligence and the Study of Democracy." *Social Science Computer Review* 9:381-399

King, Gary. 1989a. "Event Count models for International Relations: Generalizations and Applications." *International Studies Quarterly* 33:123-148.

King, Gary. 1989b. *Unifying Political Methodology.* New York: Cambridge University Press.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton: Princeton University Press.

King, Gary and Will Lowe. 2003. "An Automated Information Extraction Tool for International Conflict Data with Performance as Good as Human Coders: A Rare Events Evaluation Design." *International Organization*57,3: 617-642.

Kissinger, Henry. 1979. *The White House Years* . Boston: Little, Brown.

Klahr, Philip and Donald A. Waterman. 1986. *Expert Systems: Techniques, Tools and Applications.* Reading, MA: Addison-Wesley.

Klecka, William R. 1980. *Discriminant Analysis*. Beverly Hills: Sage.

Kohonen, T. 1984. *Self-Organization and Associative Memory.* New York: Springer-Verlag.

Kollman, Ken, John H. Miller and Scott E. Page. 1992. "Adaptive Parties in Spatial Elections." *American Political Science Review* 86:929-937.

Kolodner, Janet L. 1984. *Retrieval and Organizational Strategies in Conceptual Memory: A Computer Model.* Hillsdale, NJ: Lawrence Erlbaum.

Kolodner, Janet, ed. 1988. *Proceedings of the DARPA Workshop on Case-Based Reasoning.* Palo Alto: Morgan Kaufmann.

Kolodner, Janet. 1993. *Case-Based Reasoning.* San Mateo, CA: Morgan Kaufmann

Koza, John R. 1992. *Genetic Programming*. Cambridge: MIT Press.

Krasner, Stephen D., ed. 1983. *International Regimes*. Ithaca: Cornell University Press.

Krippendorff, Klaus. 1986. *Information Theory: Structural Models for Qualitative Data*. Beverly Hills: Sage.

Kruskal, Joseph B. 1983. "An Overview of Sequence Comparison I". In *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, edited by David Sankoff and Joseph B. Kruskal. New York: Addison-Wesley.

Kugler, Jacek, Lewis W. Snider and William Longwell. 1993. "From Desert Shield to Desert Storm: Success, Strife or Quagmire?" *Conflict Management and Peace Science* 13:113-148.

Laird, John E., Paul S. Rosenbloom and Allen Newell. 1986. *Universal Subgoaling and Chunking: The Automatic Generation and Learning of Goal Hierarchies*. Norwell, MA: Kluwer.

Langley, Pat, Herbert A. Simon, Gary L. Bradshaw, and Jan M. Zytkow. 1987. *Scientific Discovery.* Cambridge: MIT Press.

Langton, Christopher G., ed. 1989. *Artificial Life.* Reading, MA: Addison-Wesley.

Langton, Christopher G., Charles Taylor, J. Doyne Farmer and Steen Rasmussen, eds. 1992. *Artificial Life II.* Reading, MA: Addison-Wesley.

Lashley, Karl. 1956. "Cerebral organization and behavior." In *The Brain and Human Behavior*, edited by H. Solomon, S. Cobb and W. Penfield. Baltimore: Williams and Wilkins.

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

Lave, Charles A. and James G. March. 1975. *An Introduction to Models in the Social Sciences.* New York: Harper and Row.

Lebow, Richard Ned. 1981. *Between Peace and War: The Nature of International Crises.* Baltimore: Johns Hopkins.

Lebow, Richard Ned and Janice Gross Stein. 1990. "Deterrence: The Elusive Dependent Variable." *World Politics* 42:336-369

Lehnert, Wendy and Beth Sundheim. 1991. "A Performance Evaluation of Text Analysis." *AI Magazine* 12:81-94.

Leijonhufvud, Axel. 1973. "Life Among the Econ." *Western Economic Journal* 11 (3): 327 - 337.

Leontief, Wassily. 1971. "Theoretical Assumptions and Nonobserved Facts." *American Economic Journal* 61,1:1-7.

Lenat, Douglas B. and R.V. Guha. 1990. *Building Large Knowledge-Based Systems: Representation and inference in the CYC project*. Reading, MA: Addison-Wesley.

Leng, Russell J. 1987. *Behavioral Correlates of War, 1816-1975.* (ICPSR 8606). Ann Arbor: Inter-university Consortium for Political and Social Research.

Leontief, Wassily. 1982. "Academic Economics." *Science* 217:104-107.

Levine, Robert I, Diane E. Drang and Barry Edelson. 1990. *AI and Expert Systems*. New York: McGraw-Hill.

Long, John M, Erach A. Irani, James R. Slagle and the POSCH Group. 1989. "Automating the Discovery of Causal Relationships in a Medical Records Database." In *Knowledge Discovery in Databases*, edited by Gregory Piatetsky-Shapiro and William Frawley. Detroit: IJCAI-89 Workshop Proceedings.

Luce, R.D. and Howard Raiffa. 1957. *Games and Decisions*. New York: Wiley.

Luenberger, David G. 1979. *Introduction to Dynamic Systems: Theory*, *Models and Applications*. New York: Wiley.

Luterbacher, Urs and Michael Don Ward, eds.. 1985. *Dynamic Models of International Conflict.* Boulder:Lynne Rienner Publishing.

McClelland, Charles A. 1967. "Event-Interaction Analysis in the Setting of Quantitative International Relations Research." Mimeo, University of Southern California.

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

McCloskey, Donald M. 1985. *The Rhetoric of Economics*. Madison: University of Wisconsin Press.

McGowan, Patrick, Harvey Starr, Gretchen Hower, Richard L. Merrit and Dina A. Zinnes. 1988. "International Data as a National Resource." *International Interactions* 14:101-113.

McGuire, Martin. 1965. *Secrecy and the Arms Race*. Cambridge: Harvard University Press.

Mackay, Alan L. 1977. *Scientific Quotations*. New York: Crane,Russak.

Majeski, Steven J. and David L. Jones. 1981. "Arms Race Modeling." *Journal of Conflict Resolution* 25:259-288.

Majeski, Stephen J. 1987. "A Recommendation Model of War Initiation: The Plausibility and Generalizability of General Cultural Rules." In *Artificial Intelligence and National Security*, edited by Stephen Cimbala. Lexington, MA: Lexington Books.

Majeski, Stephen J. and David J. Sylvan. 1984. "Simple Choice and Complex Calculations: A Critique of *The War Trap*." *Journal of Conflict Resolution* 28:316-340.

Majeski, Stephen J. 1989. "A Rule Based Model of the United States Military Expenditures Decision-Making Process." *International Interactions* 15:129-154.

Mallery, John C. and Roger Hurwitz. 1987. "Analogy and Precedent in Strategic Decision-Making: A Computational Approach." Paper presented at the American Political Science Association, Chicago.

Mallery, John C. 1988. "Thinking About Foreign Policy: Finding an Appropriate Role for Artificially Intelligent Computers." Paper presented at the International Studies Association, St. Louis.

Mallery, John C. and Frank L. Sherman. 1993. "Learning Historical Rules of Major Power Intervention in the Post-War International System." Paper presented at the International Studies Association, Acapulco.

March, James and Johan Olsen. 1984. "The New Institutionalism: Organizational factors in political life." *American Political Science Review* 78:734-750.

Margolis, Howard. 1987. *Patterns, Thinking and Cognition: A Theory of Judgement*. Chicago: University of Chicago Press.

Markel, Scott and Darryl León. 2003. *Sequence Analysis in a Nutshell: A Guide to Tools A Guide to Common Tools and Databases*. Sebastopol, CA: O'Reilly and Associates.

Masters, Roger D. 1989. *The Nature of Politics.* New Haven: Yale University Press.

May, Ernest. 1973. *"Lessons" of the Past: The Use and Misuse of History in American Foreign Policy.* New York: Oxford University Press.

May, Robert. 1976. "Simple Mathematical Models with very complicated dynamics." *Nature* 261:459-467.

Maynard-Smith, John. 1982. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Maynard-Smith, John and G.R. Price. 1973. "The Logic of Animal Conflict", *Nature* 246:16-16

Meadows, Dennis L., William W. Berens III, Donella H. Meadows, Roger F. Naill, Jorgen Randers and Erich K.O. Zahn. 1973. *Dynamics of Growth in a Finite World.* Cambridge: Wright-Allen Press.

Mefford, Dwain. 1984. "Formulating Foreign Policy on the Basis of Historical Analogies: An Application of Developments in Artificial Intelligence." Paper presented at the International Studies Association, Atlanta.

Mefford, Dwain. 1985. "Formulating Foreign Policy on the Basis of Historical Programming." In *Dynamic Models of International Conflict*, edited by Urs Luterbacher and Michael Don Ward. Boulder: Lynne Rienner Publishing.

Mefford, Dwain. 1991. "Steps Toward Artificial Intelligence: Rule-Based, Case-Based and Explanation-Based Models of Politics." In *Artificial Intelligence and International Politics*, edited by Valerie Hudson. Boulder: Westview.

Merritt, Richard L., Robert G. Muncaster, and Dina A. Zinnes, eds. 1993. *Theory and Management of International Event Data: DDIR Phase II.* Ann Arbor: University of Michigan Press.

Michalewicz, Zbigniew. 1992. *Genetic Algorithms + Data Structures = Evolution Programs*. New York: Springer-Verlag.

Michalski, Ryszaid S., Jamie G. Carbonell and Tom M. Mitchell. 1983. *Machine Learning: An Artificial Intelligence Approach.* Palo Alto: Tioga Publshing.

Michie, D., D.J. Spiegelhalter and C.C. Taylor. 1994. *Machine Learning, Neural Networks and Statistical Classification.* London: Ellis Horwood.

Midlarksy, Manus, ed. 1989. *Handbook of War Studies*. Boston: Unwin, Hyman.

Miller, George A. 1956. "The magical number seven plus or minus two: some limits on our capacity for processing information." *Psychological Review* 63:81-97.

Milliken, Jennifer L. 1994. *A Grammar of State Action: US Policymakers' Social Construction of the Korean War.* Ph.D. dissertation, University of Minnesota.

Mills, William deB. 1990. "Rule-Based Analysis of Sino-Soviet Negotiations."*Social Science Computer Review* 8:181-195.

Minsky, Marvin and Seymour Papert. 1969. *Perceptrons*. Cambridge: MIT Press.

Maoz, Zeev and Bruce Russett. 1993. "Normative and structural causes of democratic peace, 1946-1986." *American Political Science Review* 87:624-638.

Moe, Terry M. 1979. "On the Scientific Status of Rational Models." *American Journal of Political Science* 23:215-243.

Mooney, Christopher Z. and Robert D. Duval. 1993. *Bootstrapping: A Nonparametric Approach to Statistical Inference.* Newbury Park, CA: Sage Publications.

Moray, Neville. 1984. "Humans and their Relations to Ill-Defined Systems." In *Adaptive Control of Ill-Defined Systems*, edited by Oliver Selfridge, Edwina L. Rissland and Michael A. Arbib. New York: Plenum.

Morgenthau, Hans J. 1973. *Politics Among Nations*. 5th edition. New York: Knopf.

Morrow, James. 1989. "Capability, Uncertainty and Resolve: A Limited Information Model of Crisis Bargaining" *American Journal of Political Science* 33:941-72.

Morrow, James. 1994. *Game Theory for Political Scientists*. Princeton: Princeton University Press.

Most, Benjamin A. and Harvey Starr. 1984. "International relations theory, foreign policy substitutability and 'nice' laws." *World Politics* 36:383-406

Most, Benjamin A. and Harvey Starr. 1989. *Inquiry, Logic and International Politics*. Columbia: University of South Carolina Press.

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

Myerson, Roger. 1991. *Game Theory.* Cambridge: Harvard University Press.

Negotia, Constantin Virgil. 1984. *Expert Systems and Fuzzy Systems*. Reading, MA: Addison-Wesley.

Neustadt, Richard E. and Ernest R. May. 1986. *Thinking in Time: The Uses of History for Decision Makers.* New York: Free Press.

Newell, Allan. 1969. "Heuristic programming: Ill-structured problems." In *Progress in Operations Research*, Vol. 3., edited by J.S. Aronofsky. New York: Wiley.

Newell, Allan and Herbert Simon. 1972. *Human Problem Solving.* Englewood Cliffs, NJ, Prentice-Hall.

Newman, James R. 1956. *The World of Mathematics*. New York: Simon and Schuster.

Nickles, Thomas, ed. 1978. *Scientific Discovery: Case Studies.* Dordrecht: Reidel

Nicholson, Michael. 1989. *Formal Theories in International Relations.* Cambridge: Cambridge University Press.

Nicholson, Michael. 1992. *Rationality and the Analysis of International Conflict.* Cambridge: Cambridge University Press.

Nii, H. Penny. 1989. "Blackboard Systems." In *The Handbook of Artificial Intelligence*, edited by Avron Barr, Paul R. Cohen and Edward A. Feigenbaum. Reading, Mass: Addison-Wesley.

Niou, Emerson, Peter Ordeshook and Gregory Rose. 1989. *The Balance of Power: Stability and Instability in International Systems.* Cambridge: Cambridge University Press.

Niou, Emerson M.S. and Peter Ordeshook. 1994. "'Less Filling, Tastes Great': The Realist-Neorealist Debate." *World Politics* 46:209-234.

Olinick, Michael. 1979. *An Introduction to Mathematical Models in the Social and Life Sciences.* Reading, MA: Addison-Wesley.

Olson, Mancur. 1965. *The Logic of Collective Action.* Cambridge: Harvard University Press.

O'Neill, Barry. 1985. "The Dollar Auction as a Model of International Escalation." In *Dynamic Models of International Conflict* edited by Urs Luterbacher and Michael Don Ward. Boulder: Lynne Rienner Publishing.

O'Neill, Barry. 1994. "Game Theory Models of Peace and War." In *Handbook of Game Theory*, vol. 2., edited by Robert J. Aumann and Sergiv Hart. New York: Elsevier.

Ordeshook, Peter C. 1986. *Game Theory and Political Theory.* Cambridge: Cambridge University Press.

Ordeshook, Peter C., ed. 1989. *Models of Strategic Choice.* Ann Arbor: University of Michigan Press.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge: Cambridge University Press.

Oye, Kenneth. 1987. "On Mitigating Deconstructive Destruction: A Response to Ashley and Alker." Mimeographed manuscript, Princeton University.

Parkinson, F. 1977. *The Philosophy of International Relations*. Beverly Hills: Sage.

Patrick, Edward A. and James M. Fattu. 1986. *Artificial Intelligence with Statistical Pattern Recognition.* Englewood Cliffs, NJ: Prentice-Hall.

Pazzani, Michael. 1988. *Learning Causal Relationships: An Integration of Empirical and Explanation-Based Learning Methods*. Ph.D. dissertion, University of California, Los Angeles.

Pedersen, Ken. 1989. *Expert Systems Programming*. New York: Wiley.

Pennington, Nancy and Reid Hastie. 1986. "Evidence Evaluation in Complex Decision Making." *Journal of Personality and Social Psychology*. 51:242-258.

Pevehouse, Jon C., and Joshua S. Goldstein. 1999. "Serbian Compliance or Defiance in Kosovo? Statistical Analysis and Real-Time Predictions," *Journal of Conflict Resolution* 43, 4: 538-546.

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

Petrak, Johann, Robert Trappl and Johannes Fürnkranz. 1994. "The Potential Contribution of AI to the Avoidance of Crises and Wars: Using CBR Methods with the KOSIMO Database of Conflicts. Research Report TR-94-32, Austrian Research Institute for Artificial Intelligence, Vienna.

Pfetsch, Frank R. and Peter Billing. 1994. *Datenhandbuch nationaler und internationaler Konflikte.* Baden-Baden: Nomos.

Piatetsky-Shapiro, Gregory and William Frawley, eds. 1989. *Knowledge Discovery in Databases.* Detroit: IJCAI-89 Workshop Proceedings.

Pierce, John R. 1980. *An Introduction to Information Theory*. New York: Dover.

Pool, Robert, 1989. "Strange Bedfellows." *Science* 245:700-703 (18 August 1989).

Powell, Robert. 1990. *Nuclear Deterrence Theory and the Problem of Credibility.* Cambridge: Cambridge University Press.

Prieditis, Armand. 1987. *Analogica: Proceedings of the First Workshop on Analogical Reasoning*. Los Altos, CA: Kaufmann.

Purkitt, Helen E. and James W. Dyson. 1987. "Experimentation and the Study of Political Decisionmaking." Paper presented at the International Studies Association, Washington.

Purkitt, Helen E. and James W. Dyson. 1986. "An Experimental Study of Political Decision-masking: How People Process Information about Defense and Social Welfare in Trying to Balance the Federal Budget." Paper presented at the International Studies Association, Anaheim.

Purkitt, Helen E. 1991. "Artificial Intelligence and Intuitive Foreign Policy Makers Viewed as Limited Information Processors." In *Artificial Intelligence and International Politics*, edited by Valerie Hudson. Boulder: Westview.

Pynchon, Thomas. 1973. *Gravity's Rainbow*. London: Jonathan Cape.

Quinlan, J. Ross. 1979. "Induction over Large Data Bases." Heuristic Programming Project report HPP-79-14. Computer Science Department, Stanford University.

Quinlan, J. Ross. 1983. "Learning Efficient Classification Procedures and Their Application to Chess End Games." In *Machine Learning*, edited by R.S. Michalski, J. Carbonell and T. Mitchell. Palo Alto: Tioga Press.

Rapoport, Anatol. 1974. *Fights, Games and Debates.* Ann Arbor: University of Michigan Press.

Rawlins, Gregory J.E. 1991. *Foundations of Genetic Algorithms.* San Mateo CA: Morgan Kaufmann.

Rabiner, L. R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings. of the IEEE* 77,2:257-286

Reber, Arthur S. 1993. *Implicit Learning and Tacit Knowledge.* New York: Oxford University Press.

Richards, Diana. 1993. "A Chaotic Model of Power Concentration in the International System." *International Studies Quarterly* 37:55-72.

Richardson, Lewis F. 1960. *Arms and Insecurity*. Chicago: Quadrangle Books.

Riker, William H. and Peter Ordeshook. 1973. *An Introduction to Positive Political Theory.* Englewood Cliffs, NJ: Prentice-Hall.

Rosecrance, Richard N. 1963. *Action and Reaction in World Politics.* Boston: Little Brown.

Rosenblatt, F. 1962. *Principles of Neurodynamics.* New York: Spartan Books.

Rothman, Peter. 1989. "Selecting an Uncertainty Management System." *AI Expert* (July 1989):56-62.

Rumelhart, David E., James L. McClelland and the PDP Research Group. 1986. *Parallel Distributed Processing*. Cambridge: MIT Press.

Rummel, Rudolph J. 1972. *The Dimensions of Nations.* Beverly Hills: Sage.

Russell, Stuart and Peter Norvig. 1995. *Artificial Intelligence: A Modern Approach*. New York: Prentice-Hall.

Saaty, Thomas L. 1968. *Mathematical Models of Arms Control and Disarmament.* New York: Wiley.

Salton, Gerald. 1989. *Automatic Text Processing.* Reading, MA: Addison-Wesley.

Sandler, Todd. 1992. *Collective Action: Theory and Applications.* Ann Arbor: University of Michigan Press.

Sankoff, David and Joseph B. Kruskal, eds. 1983. *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison.* New York: Addison-Wesley.

Sanjian, Gregory S. A. 1988a. "Fuzzy Sets Theory and U.S. Arms Transfers: Modeling the Decision-Making Process." *American Journal of Political Science* 32:1018-1046.

Sanjian, Gregory S. A. 1988b. "Arms Export Decision-Making: A Fuzzy Control Model." *International Interactions* 14:243-266.

Saperstein, Alvin M. and Gottfried A. Meyer-Kress. 1988. "A Nonlinear Dynamical Model of the Impact of SDI on the Arms Race." *Journal of Conflict Resolution* 32:636-670.

Sawyer, Brian and Dennis Foster. 1986. *Programming Expert Systems in Pascal.* New York: Wiley.

Schalkoff, Robert. 1992. *Pattern Recognition: Statistical, Structural and Neural Approaches.* New York: Wiley.

Schank, Roger C. and Robert Abelson. 1977. *Scripts, Plans, Goals and Understanding.* Hillsdale, NJ: Erlbaum Associates.

Schank, Roger C. 1990. *Tell Me a Story: A New Look at Real and Artificial Memory.* New York: Scribner's

Schank, Roger C. 1991. "Where's the AI?" *AI Magazine* 12:38-49.

Schelling, Thomas C. 1960. *The Strategy of Conflict.* Oxford: Oxford University Press.

Schelling, Thomas C. 1978. *Micromotives and Macrobehavior.* New York: Norton.

Schneider, Gerald, Katherine Barbieri, and Nils Petter Gleditsch. 2003. *Globalization and Armed Conflict.* Lanham, MD: Rowman and Littlefield.

Schrodt, Philip A. 1981a. "Conflict as a Determinant of Territory." *Behavioral Science* 26:37-50

Schrodt, Philip A. 1981b. "A Mathematical Model of the Persistance of International Conflict." *International Interactions* 8:335-348.

Schrodt, Philip A. 1983b. "The Lagged Logistic Equation as a Model of Sporadic Conflict." In *Conflict Processes and the Breakdown of International Systems*, edited by Dina A. Zinnes. Denver: University of Denver Monograph Series in World Affairs.

Schrodt, Philip A. 1984. "Artificial Intelligence and International Crisis: An Application of Pattern Recognition." Paper presented at the American Political Science Association, Washington, DC.

Schrodt, Philip A. 1985. "The Role of Stochastic Models in International Relations Research." In *Theories, Models and Simulation in International Relations*, edited by Michael Don Ward. Boulder:Westview.

Schrodt, Philip A. 1986a. "Set Prediction of International Events using a Holland Classifier." Paper presented at the International Studies Association, Anaheim.

Schrodt, Philip A. 1986b. "Predicting International Events." *Byte.* November.

Schrodt, Philip A. 1987a. "Pattern Matching, Set Prediction and Foreign Policy Analysis." In *Artificial Intelligence and National Security*, edited by Stephen Cimbala. Lexington, MA: Lexington Books.

Schrodt, Philip A. 1987b. "Classification of Interstate Conflict Outcomes using a Bootstrapped CLS Algorithm." Paper presented at the International Studies Association, Washington.

Schrodt, Philip A. 1988a. "PWORLD: A Precedent-Based Global Simulation." *Social Science Computer Review* 6:27-42.

Schrodt, Philip A. 1988b. "Artificial Intelligence Models of International Behavior." *The American Sociologist* 19:71-85.

Schrodt, Philip A. 1989a. "Short Term Prediction of International Events using a Holland Classifier." *Mathematical and Computer Modeling* 12:589-600.

Schrodt, Philip A. 1989b. "'If it ain't broke...': A Critical Assessment of Formal Models of International Relations." Paper presented at the American Political Science Association, Washington.

Schrodt, Philip A. 1990a. "Parallel Event Sequences in International Crises." *Political Behavior* 12:97-123.

Schrodt, Philip A. 1990b. "Prediction of Interstate Conflict Outcomes using a Neural Network." Paper presented at the American Political Science Association meetings, San Francisco.

Schrodt, Philip A. 1991a. "Pattern Recognition of International Event Sequences: A Machine Learning Approach." In *Artificial Intelligence and International Politics*, edited by Valerie Hudson. Boulder: Westview.

Schrodt, Philip A. 1991b. "Artificial Intelligence and International Relations: An Overview." In *Artificial Intelligence and International Politics*, edited by Valerie Hudson. Boulder: Westview.

Schrodt, Philip A. 1991c. "Classification of Interstate Conflict Outcomes using a Bootstrapped ID3 Algorithm." IN *Political Analysis* 2, edited by James Stimson. Ann Arbor: University of Michigan Press.

Schrodt, Philip A. 1991d. "Prediction of Interstate Conflict Outcomes Using a Neural Network." *Social Science Computer Review* 9:359-380.

Schrodt, Philip A. 1994. "Event Data in Foreign Policy Analysis." In *Foreign Policy Analysis: Continuity and Change*, edited by Laura Neack, Jeanne A.K. Hey, and Patrick J. Haney. New York: Prentice-Hall.

Schrodt, Philip A. 1999. "Early Warning of Conflict in Southern Lebanon using Hidden Markov Models." Pp. 131-162 in Harvey Starr, ed. *The Understanding and Management of Global Violence: New Approaches to Theory and Research on Protracted Conflict.* New York: St. Martin's Press.

Schrodt, Philip A. 2000. "Pattern Recognition of International Crises using Hidden Markov Models." pp. 296-328 in Diana Richards (ed.) *Political Complexity: Nonlinear Models of Politics.* Ann Arbor: University of Michigan Press.

Schrodt, Philip A. and Michael Don Ward. 1981. "Statistical Inference in Incremental and Difference Equation Formulations." *American Journal of Political Science* 25:815-832.

Schrodt, Philip A. and Deborah J. Gerner. 1994 . “Validity assessment of a machine-coded event data set for the Middle East, 1982-1992.” *American Journal of Political Science* 38:825-854.

Schrodt, Philip A., Shannon G. Davis and Judith L. Weddle. 1994. “Political Science: KEDS—A Program for the Machine Coding of Event Data.” *Social Science Computer Review* 12:561-588.

Selfridge, Oliver, Edwina L. Rissland and Michael A. Arbib,eds. 1984. *Adaptive Control of Ill-Defined Systems*. New York: Plenum.

Shafer, Glenn. 1976. *A Mathematical Theory of Evidence.* Princeton: Princeton University Press.

Sherman, Frank L. and Laura Neack. 1993. “Imagining the Possibilities: The Prospects of Isolating the Genome of International Conflict from the SHERFACS Dataset”. In *International Event-Data Developments: DDIR Phase II*, edited by Richard L. Merritt, Robert G. Muncaster, and Dina A. Zinnes. Ann Arbor: University of Michigan Press.

Shortliffe, Edward H. 1976. *MYCIN: Computer-Based Medical Consultation.* New York: Elsevier.

Shubik, Martin. 1985. *Game Theory in the Social Sciences.* Cambridge: MIT Press.

Shute, Nevil. 1957. *On the Beach.* New York: William Morrow.

Simon, Herbert A. 1979. *Models of Thought.* New Haven: Yale University Press.

Simon, Herbert A. 1982. *The Sciences of the Artificial*. Cambridge: MIT Press

Simon, Herbert A. 1985. “Human Nature in Politics: The Dialogue of Psychology with Political Science.” *American Political Science Review* 79:293-304.

Simon, Herbert A. 1990. “A Mechanism for Social Selection and Successful Altruism.” *Science* 250:1665-1668.

Smith, Peter D. 1990. *An Introduction to Text Processing*. Cambridge: MIT Press.

Sniderman, Paul M., Richard A. Brody and Philip E. Tetlock. 1991. *Reasoning and Choice.* Cambridge: Cambridge University Press.

Snyder, Glenn H. and Paul Diesing. 1977. *Conflict Among Nations.* Princeton: Princeton University Press.

Snyder, Richard C., H.W. Bruck and Burton Sapin. 1969. “The Decision-making Approach to the Study of International Politics.” In *International Politics and Foreign Policy*, edited by James Rosenau. New York: Free Press.

Sokal, Alan and Jean Bricmont. 1999. *Fashionable Nonsense: Postmodern Intellectuals' Abuse of Science.* New York : Picador USA.

Springer, Sally P. and Georg Deutsch. 1981. *Left Brain, Right Brain.* San Francisco: W.H. Freeman.

Stewart, Philip D., Margaret G. Hermann and Charles F. Hermann. 1989. “Modeling the 1973 Soviet Decision to Support Egypt.” *American Political Science Review* 83:35-60.

Stonehouse, Bernard. 1985. *Facts on File Pocket Guide to the World*. New York: Facts on File Publications.

Sylvan, David J. and Stephen J. Majeski. 1985. “Tacit Cultural Rules in Foreign Policy.” Paper presented at the Merriam Research Conference, University of Illinois, October.

Sylvan, David and Stephen Majeski. 1986. “Problems of Contextual Specificity in Computational Models of Foreign Policy.” Paper presented at the International Studies Association, Anaheim.

Sylvan, David, Stephen Majeski and Jennifer L. Milliken. 1991. “Theoretical Categories and Data Construction in Computational Models of Foreign Policy.” In *Artificial Intelligence and International Politics*, edited by Valerie Hudson. Boulder: Westview.

Sylvan, Donald A., Ashok Goel and B. Chandrasekran. 1990. “Analyzing Political Decision Making from an Information Processing Perspective: JESSE.” *American Journal of Political Science* 34:74-123.

Sylvan, Donald A., Ashok Goel and B. Chandrasekran. 1991. “Modeling Foreign Policy Decision Making as Compiled Reasoning.” In *Artificial Intelligence and International Politics*, edited by Valerie Hudson. Boulder: Westview.

Sylvan, Donald A. and Steve Chan, 1984. *Foreign Policy Decision Making: Perception, Cognition and Artificial Intelligence.* New York: Praeger.

Taber, Charles S. 1990. "POLI: An Expert System Model of U.S. Foreign Policy Belief Systems in the 1950s." Paper presented at the Midwest Political Science Association meetings, Chicago.

Taber, Charles S. 1992. "POLI: An Expert System Model of U.S. Foreign Policy Belief Systems." *American Political Science Review* 86,4:888-904.

Taber, Charles S. n.d. "The Policy Arguer (POLI) Manual." Photocopy, University of Illinois.

Taber, Charles S. and Richard Timpone. 1994. "The Policy Arguer: the Architecture of an Expert System." *Social Science Computer Review* 12:1-37.

Taber, Charles S. and Kenneth S. Rona. 1995. "EVIN: A Computational Model of Event Interpretation". Paper presented at the International Studies Association, Chicago.

Tanaka, Akihiko. 1984. "China, China Watching and CHINA-WATCHER." In *Foreign Policy Decision Making: Perception, Cognition and Artificial Intelligence*, edited by Donald A. Sylvan and Steve Chan. New York: Praeger.

Thaler, Richard H. 1991. *Quasi-Rational Economics.* New York: Russell Sage.

Thompson, Beverly and William Thompson. 1986. "Finding Rules in Data." *Byte* 11 (November 1986):149-158.

Thorson, Stuart and Donald A. Sylvan. 1982. "Counterfactuals and the Cuban Missile Crisis." *International Studies Quarterly* 26:537-71.

Tversky, Amos and Daniel Kahneman. 1987. "Rational choice and the framing of decisions and the psychology of choice." In *Rational Choice: The Contrast between Economics and Psychology*, edited by Hogarth, Robin M. and Melvin W. Reder. Chicago: University of Chicago Press.

United States Arms Control and Disarmament Agency (ACDA) (1989). *World Military Expenditures and Arms Transfers 1988*. Washington: U.S. Government Printing Office.

Unseld, Sigrid D. and John C. Mallery. 1991. "Interaction Detection in Complex Datamodels". AI Memo No. 1298, MIT Artificial Intelligence Laboratory, Cambridge, MA.

van Creveld, Martin. 1991. *Technology and War.* New York: Free Press.

Vasquez, John A. 1976. "Statistical Findings in International Politics." *International Studies Quarterly* 20:171-218.

Vasquez, John A. and Marie T. Henehan. 1992. *The Scientific Study of Peace and War*. New York: Lexington Books

Vertzberger, Yaacov Y.I. 1990. *The World in their Minds: Information Processing, Cognition and Perception in Foreign Policy Decision Making.* Stanford: Stanford University Press.

Voss, James F., Terry Greene, Timothy Post and Barbara Penner. 1983. "Problem Solving Skill in the Social Sciences." *The Psychology of Learning and Motivation* 17:194-95

Voss, James F. Jeffery Blais, Mary L. Means, Terry Greene, and Ellen Ahwesh. 1986. "Informal Reasoning and Subject Matter Knowledge in the Solving of Economics Problems by Naive and Novice Individuals." *Cognition and Instruction* 3:269-302.

Waldrop, M. Mitchell. 1988a. "Toward a Unified Theory of Cognition." *Science* 241:27-29 (1 July 1988).

Waldrop, M. Mitchell. 1988b. "SOAR: A Unified Theory of Cognition?" *Science* 241:296-298 (15 July 1988).

Ward, Michael D., ed. 1985. *Theories, Models and Simulations in International Relations*. Boulder: Westview Press.

Wasserman, Philip D. 1989. *Neural Computing: Theory and Practice.* New York: Van Nostrand Reinhold.

Wayman, Frank W. and Paul F. Diehl. 1994. *Reconstructing Realpolitik.* Ann Arbor: University of Michigan Press.

Weigley, Russell F. 1973. *The American Way of War: A History of United States Military Strategy and Policy.* Bloomington: Indiana University Press.

Weiss, Sholom M. and Casimir A. Kulikowski. 1990. *Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning and Expert Systems.* San Mateo, CA: Morgan Kaufmann.

Whaley, Barton. 1973. *Codeword BARBAROSSA*. Cambridge: MIT Press.

Widrow, G. and Hoff, M.E. 1960. "Adaptive switching circuits." *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record, Part 4.*:96-104.

Wigner, Eugene. 1960. "The Unreasonable Effectiveness of Mathematics in the Natural Sciences." *Communications on Pure and Applied Mathematics* 13:104-109.

Wilber, Ken. 2002. *Boomeritis.* Boston, MA: Shambhala Publications.

Williams, J.D. 1954. *The Compleat Strategyst*. New York: McGraw-Hill.

Winston, Patrick Henry. 1984. *Artificial Intelligence*. Reading, MA: Addison-Wesley.

Wirtz, James J. 1989. "Analogies and the Tet Offensive". Paper presented at the American Political Science Association meetings, Atlanta.

Wohlstetter, Roberta. 1962. *Pearl Harbor: Warning and Decision.* Stanford: Stanford University Press.

Wright, Quincy. 1964. *A Study of War*. Chicago: University of Chicago Press.

Yager, Ronald R. and Lotfi A. Zadeh. 1992. *An Introduction to Logic Applications in Intelligent Systems.* Hingham, MA: Kluwer Academic Publishers.

Zadeh, Lofti A. 1965. "Fuzzy Sets." *Information and Control* 8:338-353.

Zadeh, Lofti A., King-Su Fu, Kokichi Tanaka and Masamichi Shimura. 1975. *Fuzzy Sets and their Applications to Cognitive and Decision Processes*. New York: Academic Press.

Zeeman, E. C. 1976. "Catastrophe Theory." *Scientific American* 234:65-83.

Zeng Lengche 1999. "Prediction and classification with neural network models." *Sociological Methods and Research* 27, 4: 499-524.

Zimmerman, H.-J. 1991. *Fuzzy Set Logic and Its Applications.* Hingham, MA: Kluwer Academic Publishers.

Zinnes, Dina A. 1976. *Contemporary Research in International Relations.* New York: Free Press.

Zinnes, Dina A. 1983. "An Event Model of Conflict Interaction." In *Conflict Processes and the Breakdown of International Systems*, edited by Dina A. Zinnes. Denver: University of Denver Monograph Series in World Affairs.

Zinnes, Dina A. and John V. Gillespie, eds. 1976. *Mathematical Models in International Relations.* New York: Praeger.