Thinking About Foreign Policy:
Finding an Appropriate Role for Artificially Intelligent Computers

John C. Mallery

Department of Political Science and
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
545 Technology Square, NE43-797
Cambridge, MA 02139-4301
Phone: (617) 253-5966
Internet: JCMA@AI.MIT.EDU

May 19, 1994

Abstract

The growing complexity of the foreign-policy conundrum has spawned a tremendous increase in the information available to support decisions without a commensurate increase in the ability to effectively process it. Office automation offers a means of increasing the productivity and effectiveness of human decision-makers but does not help with the analysis and interpretation of information. Artificial intelligence (AI) systems have been proposed for this support role, and in some cases, such systems are already appearing in the military, intelligence, and foreign-policy sectors. This report considers some of the transformations in international systems that these systems may produce. But, the report focuses on symbolic processing techniques from AI that make new categories of political phenomena amenable to formal modeling. The new modeling opportunities require rethinking some fundamental notions of political science and call for reconceptualization of societies and organizations within an subjectively-interpreted, "conversation-processing" framework. This report argues for constitution of an academic field of computational politics to pursue these lines of inquiry. The goals of this field are to develop better, more cognitively plausible theories and computational models of political cognition, to devise criteria for evaluating these theories and simulations, and to analyze the consequences of AI systems that play roles in the international systems. A critical review of the implemented symbolic models in international relations or foreign-policy decision-making shows the existence of a literature about computational politics. The report prepares for the review by noting limitations and restrictions of the knowledge-based "expert" systems. Following the review, the report explains a series of issues related to evaluating AI models and interpreting their behavior. Following from discussion of evaluative criteria, the conclusions caution that the application of AI technologies to critical areas of national defense is ill-advised even if computational politics holds substantial promise for improving knowledge of international relations and supporting foreign-policy decision-making.

This paper was presented to the panel on "Analytical Frameworks for the Analysis of Conflict," at *The 1988 Annual Meeting of the International Studies Association,* Adam's Mark Hotel, St. Louis, Missouri, March 28 to April 3, 1988. An earlier version was submitted as a Master's Thesis to the M.I.T. Department of Political Science, February, 1988. The thesis won the M.I.T. Political Science Department's Daniel Lerner Prize for best Master's thesis, 1987-88.

Contents

1 Introduction

1.1 The Foreign Policy Dilemma

Foreign policy aims to steer a safe course through the troubled waters of international relations. The ostensible goals of American foreign policy are to promote the security and well-being of the United States, and by extension its friends and allies. An important precondition and generalization for achieving these goals is promoting a stable international environment conducive to economic progress. The domestic side of foreign policy involves mobilizing the bureaucracy to carry out specific policies, cultivating domestic constituencies to lobby for the policies, and convincing the Congress to support the policies. The substance of foreign policy ranges from the exceptional to the mundane, from nuclear crises to the price of coffee. Although the consequences of error may be immediate and severe in crisis situations, the aggregation of small errors in day-to-day policy decisions may lead ultimately, as it does in chess, to serious crises. In both cases, forseeing the distant consequences of decisions can help avoid the deleterious ones. The ability to foresee the future consequences of policy is precisely the kind of planning which many have advocated but which seems bureaucratically untenable (Bloomfield, 1982: 165-192).

The main trend in the postwar international system is proliferating complexity in all dimensions of analysis and a parallel information explosion. The major motive factors are faster communication, larger-scale economic activity, and technological progress. In economic policy, economic interdependence among non-communist countries, particularly OECD countries, has grown dramatically. The domestic economy has grown and ramified with an ever more extensive functional division of labor, yielding more numerous sectoral or geographic interests. Formulating a foreign economic policy requires an understanding of an interdependent world economy, the consequences for the American economy, the effects on America's friends, the aggregate impact on the world system, not to mention the acceptability to the domestic coalitions on whose support it depends. North-South issues have also grown in magnitude with the rising consciousness and demands in the non-OECD world. Military competition with the Soviet Union is achieving new heights of technical sophistication and destructive precision. Superpower competition also manifests itself in regional jockying for influence and in sporadic military confrontations through surrogates in the third world. The fundamental characteristic of the post-war international system is *change*; this perpetual change ensures that theories or experiential notions of the workings of the international system lag behind the phenomena. This thumbnail sketch hardly does justice to complexities, subtleties, and vagaries of the foreign policy conundrum in the modern era, but it should at least convey an inkling of the intellectual challenges facing the foreign policy practitioner.

For the foreign policy practitioner, the proliferating complexity of the external world is directly reflected in ever larger quantities of information available in the internal bureaucratic environment. Over the postwar period, the internal bureaucratic environment has grown more diverse, differentiated, and complex. Similarly, in the post-Vietnam era, the task of mobilizing public support for policies has grown more problematic with the atrophy of the earlier "foreign policy consensus." Ideally, decisions should take into account all available information. But practically, there is too much to consider. Thus the policy maker must walk a fine line between failing to consider relevant information and becoming hopelessly overloaded by too much information. The decisional time frame is typically short but sometimes extends to a week or ninety days, perhaps even a year on rare occasions. Thus, the foreign policy dilemma is that *formulation of sound policies requires consideration of more information than is possible within the time available for reaching decisions.* Moreover, the scale, inertia, and complexity of many components of the international system, with which the foreign policy practitioner must contend, are such that effective policies in certain areas may only be achieved through policies planned out in advance (Bloomfield, 1982: 184-187). Of course, this assumes that the international system is relatively well-understood – a dubious assumption at best.

Since there is little hope of simplifying the reality which foreign policy addresses, the alternative is to enhance the effectiveness of the organizations that carry out and manage foreign policy. The basic task is to improve the ability of these organizations to solve foreign policy problems, whether they appear as short-term crises or long-term consequences of their actions. There are several ways to improve organizational effectiveness:

1. Develop better theories about the phenomena through competitive explanation;
2. Collect more theory-relevant information about the phenomena;
3. Build accurate, predictive models of the phenomena based on theories and relevant information;
4. Improve the educational and bureaucratic infrastructure to support 1, 2 and 3;
5. Improve organizational information processing with office automation;
6. Improve organizational decision-making with better models of the phenomena that can help predict the consequences of policies.

The emphasis in this report is on building new types of models using the techniques of symbolic processing from artificial intelligence (AI). Although the next section sketches in the abstract how networked office workstations can help the management and timely delivery of focused information to decisionmakers, the main thrust of the report is to examine the prospects for and the bases of AI systems that could support the foreign policy decision-maker in his problem-solving.

1.2 Evolutionary Organizational Designs

The published record of AI systems in the area of international relations and foreign-policy decision-making has shown some promise, particularly in the development of research systems (see sections 4 and 5). However, there are no systems that are ready to go online in practical decision-making contexts. There are, however, alternatives that can produce immediate gains in the productivity and effectiveness of foreign-policy organizations. Collectively these alternatives fall under the rubric of office automation. The role of the computer is not to replace the foreign policy professional but to support his office work within the organization. In this combination of human and computer, each performs tasks at which they are better: The human does the thinking while the computer provides support in managing "virtual" paper, including such activities as indexing information, filing it, retrieving it, generating reports from it, communicating it to other people in the organization, and providing a personal interface to varied informational resources of the organization.

The essential idea of office automation, then, is to increase the amount of time people spend thinking by minimizing the transaction costs involves searching for information and the time lost shuffling paper to produce reports. Assuming a fixed time for executing an policy-related task, reducing the cost of accessing information and reformulating it for delivery increases the time available to contemplate the issues. The reformulation and delivery of information is straight-forward with sophisticated wordprocessors, although some designs are significantly more cumbersome than others. Thus, the main technical issue is precise, narrow indexation of information so that only relevant information is retrieved. It should further be possible to retrieve information in variable degrees of detail. The point is to deliver to the right amount of information according to the needs of the human, rather than to overload the human with more information than he had when dealing only with paper.

A modern office automation design should include the following major components:

1. Each office worker uses a personal computer workstation.

2. A local network allows communication between all of the workstations,[1] and therefore, allows each work station to provide a personal interface to any information resources,[2] whether human or machine, accessible via the network.

3. An electronic mail system allows any member of the organization to communicate virtually instantaneously with anyone else in the organization, independent of geographic location and diurnal cycle.

4. Instead of a simple word-processor each workstation supports a *non-linear text editor*,[3] roughly a special kind of text editor that provides the user with the capability to index, retrieve, annotate, and rearrange arbitrary segments of text (or images) from multiple sources.

5. Building from private indices developed by each office worker, functional divisions and different levels of aggregation in the organization construct their own higher level indices that subsume those within their jurisdiction.

6. Synonym terms allow users to use their own preferred terms but ensure correct indexing in aggregate indices.

---

[1]For many cases, network communication with other organizations is also very useful.

[2]Bringing together information from many sources is often called *information fusion*.

[3]Various non-linear text editing systems have been developed. The recent introduction of Apple's "Hypercard" (Goodman, 1987) is presently bringing non-linear text editing to the personal computer market. Xerox's "Notecards" (Marshall, 1986) is one of the better and more recent systems. Other systems include (Knuth, 1983; Handle & Whitaker, 1983; Trigg & Weiser, 1984; Pitman, 1985; Symbolics, 1986). The ideas for non-linear text editing derive from the INFO program, which runs under ITS, TOPS-20, as well as GNU-EMACS under UNIX, was in combination with EMACS (Stallman, 1981) the earliest non-linear text editing systems (circa 1975).

7. Given this office environment, additional productivity enhancing programs, AI systems, or other special-purpose computers may be made available to office workers, either directly on their personal work stations or indirectly via the (non-)local network.

Under a conversation-processing interpretation (Winograd & Flores, 1986; see section 3.4), an organization consists of a pattern of conversations that (re-)produce the organization over time, and also, lead to, or constitute, the execution of functions it performs. The pattern of information flow can be described as a graph, with each node being an office worker. The content of the information flow is usually written documents but may also be verbal communications. The nodes may be aggregated where more office workers are required to accomplish larger tasks. There are two kinds of information: control information (such as directives to take actions, requests for information, and commitments to take actions) and factual or content information.[4] The office worker can be seen as storing information for future use as well as adding new informational value by applying specialized knowledge to interpret the existing information. These interpretations may be stored or sent out to other parts of the organizations depending on the control information that applies at the node. Within this abstract characterization of the organization, the relationships between elements are conversations. One example of a conversation is a request for information and a reply fulfilling the request.

The electronic office provides each node in the organization with tools for keeping track of the information stored or interpreted at the node. Keeping track of the information involves filing, indexing, and the memory of the office worker. To the extent that the non-linear text editing system facilitates filing, indexing, retrieval and reproduction of information passing through the node, the capability of the node to respond to informational request with accuracy and speed will increase.[5] When informational requests are made of a node, the office worker need only select a series of indices from their local index to generate a rough draft containing the desired information. Some polishing and refinement of the presentation then yields a finished response to the request. Electronic mail provides a quick and error-free means of sending the response to the node originating the request.

If the organization represents a division of intellectual labor over a knowledge domain, then an important part of organizational function is to interpret, index, and record the knowledge of the aggregate domain by breaking it down into smaller chunks that subunits can process and perhaps store locally. If each subunit builds the local index for it processes and stores as preparations for responses to the requests of other units, then the local indices could be aggregated ultimately into a large index for all the knowledge present in the organization. The problem of combining the subindices is the mapping of local index terms into global synonyms.[6] The availability of the global index means that non-local knowledge can be easily retrieved without knowledge of where it resides. This counteracts the tendency toward fragmentation of knowledge according to the organizational division of labor that processes and stores it locally.[7]

[4] Control information may become factual information when quoted (*i.e.*, not intended to function as control information), as in planning documents formulating how some action is to be carried out or proposing routine activities for the organization.

[5] This means that a user friendly system will be more effective because will people use it.

[6] Assuming the absence of incomensurabilities in the interpretive frameworks responsible for the local indices, the mapping into a global index is non-problematic. Unfortunately, this is rarely the case; there will always be differences. Handling these differences is the primary research issue.

[7] For the secret organization, the global index poses of problems of valid compartmentalization. The

An important consequence of the electronic office is freeing the organizational structure from geographic location, to the extent that face-to-face communication is not necessary. Thus, the connective structure of the organization can adapt quickly to the pattern of information (conversations) processed by the organization. This means that individuals with specialized knowledge spread around the foreign policy apparatus can combine in *ad hoc* groups tailored to the specific exigencies of a foreign policy task.[8] But, for the electronic office to succeed it must be carefully designed to conform to the needs of the humans who must live with it everyday. If it is not easy to use, helpful, flexible, and friendly, people will not utilize it to make their work easier and their productivity higher.

This discussion of the electronic office is brief and abstract because others are working to deliver it. The computer technologies for realizing the electronic office either exist or are presently being refined.[9] The focus of of this report is on the artificial intelligence systems that might be connected to local networks in the foreign policy apparatus or actually run on personal workstations in support roles.[10] As the organization brings its knowledge online, AI systems can be deployed to help processing those areas of the collective knowledge which are well enough understood. Although knowledge-based systems (see section 2.1) are the main technology for AI applications, the scenario of the electronic office underscores the need for computers to understand natural language: Much information in an organization is in textual form.

In the electronic office, the human remains in the loop: This raises the possibility that the system may fail due to human error (Bloomfield, 1987). Thus, an important objective for the electronic office, and AI systems that support people, is to reduce the likihood of human failures. Although ease of information access in the electronic office can expose decision makers to more and varied information, it may do little to reduce the fixation on favored approaches that often clouds thinking during crises. Thus, a useful function for AI support systems is to help detect flaws in human analyses and to call attention to alternatives which otherwise might not receive due consideration. Good designs, then, will use the computers to check the thinking of the humans and the humans to check the inferences of the computers, and in this way, increase the effectiveness and reliability of the organization at large.

Improving effectiveness and response times in foreign-policy organizations cannot be considered in isolation. Competing foreign-policy organizations may make similar improvements in their own organizations and increases in effectiveness may lead to feedback consequences that lower stability in the international system. Like arms races, productivity races in foreign-policy may not ultimately improve the effectiveness national foreign policy and enhance national security if similar measures taken by adversaries yield a less stable and predictable international system. The next section examines one facet of this class of problems.[11]

---

research issue is to determine when non-local knowledge is indeed necessary elsewhere.

[8]Discounting textbooks theories of the foreign policy process, some observers believe that the formation of *ad hoc* decision groups is the way the process actually functions.

[9]A general maxim for computer systems is: Many try to implement systems but few do it right. The usual pitfalls arise from shortsighted designs that fail to contemplate future needs or fail to build in the extensibility necessary to meet them.

[10]An important consideration in the selection of workstations is the ability to run useful AI applications.

[11]There are many ways in which increasing effectiveness of national foreign-policy bureaucracies can reduce systemic stability and national security. Apart from being beyond the scope of this work, they are too

1.3 A Possible Strategic Future

Artificial intelligence (AI) systems are already surfacing in international relations and foreign-policy applications. For example, Dr. Paul Davis and his colleagues at the RAND Corporation recently delivered to the United States Defense Department an "expert system"[12] for studying, analyzing and simulating strategic decision-making (Bracken, 1983) as it might arise in nuclear confrontations between the United States and the Soviet Union (Davis, Bankes, Kahan, 1986).)[13] This system went from the drawing board to the final users in about four years. The rapidity of this move from the laboratory into the field adumbrates the increasing pace of technological innovation in the foreign policy sector (and perhaps technological competition between the superpowers). The developers of the RAND system maintain that it represents a qualitative improvement in the ability of strategists to study nuclear warfare scenarios. However, they warn that the RAND system should be used for *only* study and *not* in an operational role.

In the "Star Wars" era, decisionmakers have just seconds or minutes to select an action and weigh its consequences. Because of the present vulnerability of the strategic command and control infrastructure, the choice may be either to launch a full-scale retaliation or to have AI systems assess an incoming nuclear attack, select an appropriate response, and launch a carefully measured retaliation – all within seconds. According to this argument, a flexible response can only be retained though reliance on AI systems (Cimbala, 1987b). The main first-order question is: Will the AI system work as intended? This depends on the reliability of the system and the validity of the inferential model incorporated. The main second-order question is: What does the existence of these automated decisionmakers mean for the international system? As the time required to take actions and react decreases, the rate at which actions and reactions can occur increases. This increases the *gain* in the system, which in turn, increases the probability of non-linear amplifications of small initial perturbations in strategic systems. Thus, *even if the AI system works correctly, the presence of these systems can increase gain, and therefore, lower the stability of international security systems* (Ashby, 1956; Richardson, 1960; Deutsch, 1963; Pruitt & Kimmel, 1977). These instabilities can only by attenuated by introducing new control mechanisms to prevent amplification beyond homeostatic or adaptive boundaries (Deutsch, 1963).

For a glimpse into a possible strategic future, one might look at the current behavior of international foreign exchange markets and the stock markets. These markets are exhibiting unprecedented volatility due to the second order effects of new AI systems that trade according to strategies intended to optimize individual benefits without regard for collective consequences (Olson, 1965; Keohane & Nye, 1977; Keohane, 1984). The high gain in the foreign exchange markets is due to to short decision times and instantaneous actions. The efforts of governments to control these markets with monetary policy is severely constrained by the high gain in the system. If interest rate differentials between major OECD countries exceed relatively small thresholds, massive capital inflows and outflows

numerous to detail here. See Emery & Trist (1974) for discussion of a "vortical environment," an environment in which non-linear effects are so strong that planning and goal-seeking becomes impossible, for a theoretical model.

[12] "Expert systems," more technically knowledge-based systems, are discussed in section 2.1.

[13] Davis (1987) provides a short, cogent overview of the effort. The RAND system provides many interesting and useful features such as the ability to simulate decision-making under different assumption sets and to derive explanations automatically about how particular decisions were reached.

ensue. Floating foreign exchange rates serve to attenuate these flows by driving the low-interest currency down and the high-interest currency up. Thus, the price mechanism serves to attenuate non-linearities. The present volatility in the stock market apparently results from trading programs responding not just to the direction of movement in price and various fundamental indicators but also to the rate of the movement. Increasing rates are amplified as each trading program tries to take instant advantage of an accelerating trend, and thereby, amplifies the trend. Since the programs have only a formal theory of the domain, they lack the human common sense to see that trend is exaggerated and that values have been driven beyond reasonable levels (Sergeev, 1987a).

These exaggerated swings illustrate Winograd and Flores' (1986) notions of "blindness and "breakdown," in which the input to an expert system exceeds the range over which the system's knowledge applies but the system is unable to perceive this. Unlike human experts, contemporary "expert systems" have neither the ability to recognize when a task exceeds the range of their knowledge nor the capability to learn new ways of coping with unforeseen situations; they simply fail in ungraceful ways. Human custodians must determine the reasons for the failure and update or reformulate their programs' knowledge. Herein lies the hint of the next wave of stock market volatility due to computer trading. New, smarter programs will exploit second-order knowledge to throw off competing programs.[14] Here's how this can work. Knowing that first-order programs merely bet on accelerating trends, second-order programs can first bet on a non-existent trend to create a real trend as first-order programs are lured in. This simple counterplanning can lead to volatile price swings in the stock market that allow the smarter programs reap large profits. Like the foreign exchange case, the price mechanism again limits volatility, but still the range swings may be wide enough to be dysfunctional. In contrast, international security systems do not have a clear analog to price to limit volatility. Instead, high gain makes possible escalatory spirals of threat, insecurity, and tension that lead into war (Richardson, 1960; Holsti, 1970).[15]

Two assumptions can independently account for strong pressure to introduce AI systems into foreign policy decision-making.

- The competition between the superpowers manifests itself as technological competition in the politico-military sector.
- Improvements in organizational learning lead bureaucracies (with adequate budgets) to improve their effectiveness through timely adoption of modern computational technologies.[16]

[14]Emery and Trist (1974) argue that large, effective bureaucracies that counterplan against each other are the major source of instability in the world system. For them the problem is too much counterplanning and not enough cooperation on the basis of common values. In the absence of attenuating norms of socially acceptable behavior, improving organizational effectiveness (intelligence) with AI systems that help manage information, make decision faster, and act sooner, increases gain, and therefore instability, in these oligopolistic systems.

[15]Neo-classical economics assumes perfect information for many theoretical constructs. Perfect information is intended to produce perfectly rational decision-makers. However, the cybernetic perspective shows how a dimension of perfect information, namely short decision-times, destabilizes systems by increasing gain.

[16]The Central Intelligence Agency and its brethren in the intelligence community hold annually a secret conference on AI applications to intelligence. Andriole (1987) reviews some strategic intelligence applications AI and presents a sober perspective on the ability of AI systems in general and "expert systems" in specific to address intelligence problems. Although Andriole poses various unsolved problems in political and decision

These two perspectives correspond to *demand-pull* and *supply-push*) explanations. If either or both hold, AI systems will be deployed in the foreign-policy and military sectors as its possibility arises. At present, the main governmental funders and consumers of AI technology are the Defense Department and the intelligence agencies. Policy oriented agencies such as the State Department generally lag behind in these sophisticated and expensive technologies but they are moving with the "microcomputer revolution" as they install extensive word processing and document preparation facilities.

The basic facts of AI and international relations are that new technologies are being incorporated into complex bureaucratic structures as they are deemed capable of performing useful functions. The RAND expert system is an example of a simulation tool that has been deployed swiftly before scientific scrutiny determines its validity and reliability. SDI is an example of a strong demand pull for applying AI technologies in critical roles. The example of stock trading programs and foreign exchange markets illustrate the potential disruptive effects that that AI systems can introduce into social systems. As AI components expand the effectiveness of large bureaucracies, new sources of instability can be introduced into the international system.

1.4 The Need for Scientific Study

Two key questions questions need to be answered for each application of an AI system in the foreign policy sector:

- Does the system actually perform the job according to the claims of its proponents?
- What are the consequences for the organization and the international systems within which it is embedded?

Systematic scholarly study is clearly required to answer these questions. International relations scholars and computer scientists need to study the crucial consequences these systems could have for the stability of the international system and for accidental nuclear wars. Indeed organizations analogous to the Food and Drug Administration or the Environmental Protection Agency (EPA) would seem important institutional correlates of sound application of AI techniques in public policy domains. Perhaps the idea of an "ecological impact" analysis should precede the introduction of systems into critical areas. It is undoubtedly better to assess the consequences before a gross failure than afterwards.

Who will be the critics and supporters of AI systems as they are introduced into foreign policy decision-making and international simulations? Who will be the "experts" that study the merits and demerits of these systems? How will we know if these systems

theoretic analysis for AI to address, he omits applications that on the horizon. Notably absent are: 1) Satellite recognition of military installations and weapons systems based on model-based vision (Brooks, 1981); 2) Cryptographic applications of symbolic algebra or learning techniques; 3) Massively parallel computers for keyword search (Stanfill & Kahle, 1986) through many gigabyte databases of intelligence documents; 4) Non-linear text editing, *e.g.* Xerox's "Notecards" System (Marshall, 1986), for indexing documentary databases and preparing analyses drawing on them.

are performing according to acceptable standards of reliability? What second order effects will there be for relevant international systems? The task of answering these questions falls to international relations scholars and AI researchers. While the AI researchers are more concerned with developing new techniques than critiquing existing ones, those few international relations and foreign-policy scholars that employ formal models have focused on methodologies borrowed from primarily from economics, such as multivariate regression, factor analysis, and game theory. The general lack of attention to computational modeling in the political science profession severely circumscribes the literature assessing the implications of new computational technologies for foreign policy decision-making, abdicating the task mostly to other disciplines, such as computer science. But, as AI technologies are introduced into the practice of foreign-policy decision-making, their consequences will surely spill into the field of international relations as they effect the structures and processes of the international system. To meet this challenge the field of AI modeling in international relations, or more generally *computational politics* (see section 3), needs to be constituted and populated.

The first step, then, in developing a new interdisciplinary field to identify it. Thus, one task of this report is to frame the dimensions of this field. One key component is to review the published literature describing AI models of international relations and foreign policy decision-making. This literature constitutes the scholarly core of the field. The literature employs expert systems extensively but is generally insensitive to the assumptions underlying expert system technology. Thus, it is important to review these assumptions. This insensitivity and general emphasis on expert systems also suggests a need to review some of the other directions in AI. Another facet of framing the field is to link the AI modeling approach to more traditional quantitative modeling in political science, and to explain the reasons that motivate political scientists to choose to AI methods. A final aspect of organizing this field is to establish a framework for evaluating AI system in terms of not just their computational properties but also the relationship between these systems and the reality which they purport to model.

1.5 The Report Plan

The major formalism exported to political science from AI is the knowledge-based system. The literature relying on knowledge-based systems in political science seems, for the most part, unaware of the assumptions that underlie the technology. Thus, the section reviews the basic concepts of knowledge-based systems, and then, builds from the weaknesses noted in more recent directions in AI that hope to overcome these weaknesses. This understanding of AI provides a starting point for introducing computational politics as a new, emergent subfield of *polimetrics*, or formal modeling techniques for political science (Alker, 1975), . In this third section, these reasons why political science should adopt symbolic modeling techniques revolve around intentionality and cognitively plausible models. Out of this emerge the larger modeling issues of political cognition, the inadequacies of a monological AI technology, and the reconceptualization of social systems as conversation-processing entities. This preparation then provides a vantage point for the next two sections which review, and often comment, on the literature in international relations and foreign-policy decision-making that treats implemented AI models. The first of these sections looks at models using artificial formalisms while the second examines natural-language based approaches. This literature review is current through 1986. Following the review, the next section develops criteria for evaluation and interpretation of AI models. Although these criteria are somewhat premature for the level of advancement in the literature, they are timely for those who might apply AI systems in critical areas of national defense. The sections includes discussion of problems of felicitous representation of political phenomena as well as theoretical and practical considerations in model development. It culminates in suggestions for assessing the validity and reliability of computational models. The conclusion stresses four themes that run through the report: 1) computational politics is not ready for the real world; 2 & 3) AI and social science have much to offer each other; 4) Natural language modeling grounded in lexical interpretation and precedent logics are the most promising approaches for computational politics.

2 What do we mean by AI?

Artificial Intelligence[17] is an interdisciplinary branch of computer science that seeks to make computers behave in ways which would be called intelligent if attributed to a human. A more recent technical definition considers artificial intelligence to address the problem of connecting perception with action in the robot.[18] Newell (1987) considers intelligence the ability to effectively employ knowledge to solve problems that achieve goals.[19] The orienting research goal of developing intelligent robots leads the field to address issues in vision, motion planning, manipulator and locomotion design, signal processing for speech understanding, natural language understanding, knowledge representation, problem solving, control method, search, inference mechanisms, learning, cognitive modeling, and psychological studies of intelligence. It also treats areas such as special-purpose computer hardware, VLSI design, software programming, automatic programming, and philosophical foundations in so far as they support any of the central goals.[20] The complexity and difficulty of the central goals drives progress in computer hardware and software to support AI research. Two examples are The Lisp Machine and the Connection Machine (Hillis, 1986), both special-purpose computers designed with AI applications in mind. The need for AI to invent new computational and theoretical tools to meet the demands of its research goals reflects the "revolutionary science" that presently permeates the field. As yesterday's AI tools and techniques become better understood, they gradually pass into the "normal science" of computer science and receive application in other fields.

Although major advances have been made, the AI problem is *not solved*, the commercial fanfare notwithstanding. There are two general characterizations of the AI project. The *hard AI* proponents seek to create intelligent computers capable of passing the "Turing test," convincing an unwitting person that they are interacting with a human (Turing, 1950). According to the hard AI view, every component of human intelligence can be so precisely described and hardware made so effective that a computer can fully simulate human intelligence – whatever one takes that to be. On this view, AI is the simulation branch of cognitive science, a multidisciplinary field, that seeks a functional account of the human

[17]Rich (1987) provides a short overview of the field. Waltz (1985) provides a comprehensive taxonomy. Shapiro (1987) provides a comprehensive technical overview. Boden (1977) and McCorduck (1979) provide detailed histories. Gardner (1985) situates AI within the allied disciplines that make up cognitive science.

[18]Patrick Winston, personal communication, February, 1987.

[19]There is no generally accepted definition of intelligence in artificial intelligence. Definitions are controversial and many major figures refuse to venture operational definitions.

[20]The taxonomy of AI in (Waltz, 1985) clearly indicates the diversity and depth of AI.

mind.[21] While linguists, anthropologists, philosophers, psychologists, and brain scientists each study human cognition from their varying perspectives, AI researchers attempt to gain insights into natural and artificial intelligence by devising computational implementations.

Most AI researchers are more practical in their views and seek useful engineering results rather than cognitive plausibility.[22] These *soft AI* proponents simply aim to make computers more useful by adding ever more powerful capabilities, which may or may not resemble aspects of human intelligence, but whose implementation may be quite different. Soft AI is an engineering project prepared to use whatever tricks necessary to build smarter, more useful computers that solve practical problems.

Although widely accepted as classificatory criteria for the various strains of AI research, the conventional distinction between soft and hard AI is actually a false dichotomy. The primary source of weakness is the absence of an evolutionary orientation. Clearly, computers cannot be made to think like people over night; neither will they think without careful engineering. Furthermore, no account of individual psychology will be found the absence of an understanding of the relationship between the cognitive processes of the individual and their embedding in society.[23] This suggests an intermediate position for an *evolutionary* AI that seeks to simulate *some* features of psychological processing, social cognition, and social action. Evolutionary AI must develop greater insights into human psychology by attacking the problem from all angles. A major thrust must be computational implementations that connect existing technology with theories of social cognition. A cycle of theory formation, implementation, testing, and reformulation provides the basis for a scientifically cumulative research program. Proof of concept are done by theoretical argument and empirical demonstration in collective implementations. As the branches of inquiry on which it draws advance, evolutionary AI may move closer to the *hard* AI project.

Whatever the specific approach, there is one universal truth: the quest for engineered answers to deep psychological questions expands to absorb the most powerful computational tools available – and demands more. At any particular historical time, the implementationally addressable research problems are circumscribed by limitations in hardware speed and software complexity. Thus, tool building projects invariably precede waves of progress as they open up new software and hardware architectures for serious consideration. While AI pushes the frontiers of computer science, the fruits of computer science feedback in the form of better design tools, operating systems, and hardware architectures that, in turn, enable further advances. The "product life cycle" for AI techniques proceeds from a practically mystical aura when they first appear in AI laboratories to a concrete computational tool as computer science and wider use demystifies their "intelligence."

[21]See Gardner (1985) for a historical overview of cognitive science.

[22]After all, the cognitive plausibility of any program running on a computer orders of magnitude less powerful than the human brain is dubious at best, even though it may illustrate how some facets of intelligence could work.

[23]The relevance of the social embedding permeates this report. See sections 3.3 and 5.3.1 for specific discussions.

2.1 Domain-Specific Knowledge-Based Systems

While hard AI has few credits to its account, soft AI it has developed many practical technologies and disseminated them widely. The best known class of AI system is the so-called "expert" system.[24] Expert systems, or more technically *knowledge-based systems*, are computer software systems capable of representing knowledge in a form that can be used to make practical inferences in a *tightly constrained domain*.[25] The domain must be tightly constrained because the inference processes they incorporate are too simple and rigid to handle the variety and novelty of larger, less constrained domains. The knowledge possessed by these systems is generally not algorithmic.[26] It is usually a large collection of heuristics or rules of thumb that work for the domain but are not provably correct. It is important to realize that these systems mimic the *external behavior* of experts but do not capture the actual *internal* problem-solving of humans.

Depending on the formalism, knowledge is encoded as pattern-action rules for production systems, axioms for automatic theorem provers, propositions and predicates for logic programming languages (*e.g.*, PROLOG), or simply conditional branches for procedural programming languages (*e.g.*, LISP, FORTRAN). These approaches (and there are potentially more) are often formally equivalent, and in principle, can accomplish the same tasks. However, each differs in ways that make some tasks easier (more natural for the formalism) than others. For example, control knowledge is most easily expressed in procedural languages. However, the two most widespread techniques for creating knowledge-based systems are production systems and logic programming.

2.1.1 Operational Assumptions

The success of knowledge-based systems rests in their ability to exploit a large amount of information (300 to 15,000 rules) about a domain to accomplish practical tasks. They work best in domains characterized by well-defined boundaries, comprehensive domain descriptions, complete domain theories, little change, and no need for general inference methods or planning. Such domains are usually not whole fields but are instead small corners, usually termed "microworlds." In general, knowledge-based systems require certain assumptions to be satisfied:

[24] Political scientists and students of rhetoric should note the ideologically loaded nature of the term "expert system." It reinforces the cult of the expert, the naive hope for technocratic solutions, and the mystical faith in the infallibility of computer oracles. Furthermore, the main sociological result of the "expert system revolution" has been the dignification of the common-sense possessed by ordinary people. The exaltation of expertise follows as elite social groupings create barriers to entry based on specialized technical knowledge or niche-specific power. But, the ability of relatively simple programs to represent and deploy expert knowledge illustrates their true sophistication; common-sense and ordinary intelligence remain practically inscrutable. This result supports the conclusion that intellectual differences between the experts and ordinary people are far less than their respective social stati might indicate.

[25] (Brownston, *et al.*, 1985) is a recent introduction to knowledge-based systems. (Buchanan & Shortliffe, 1984) and (Hayes-Roth, Waterman, Lenat, 1983) are earlier introductions. Hayes-Roth (1987) provides a concise overview.

[26] An algorithm is an abstract characterization of a procedure that yields *correct* results for *all* cases over which it is defined. A heuristic may succeed for some cases and fail for others, even though it is defined for both. In short, an algorithm always succeed for defined cases while a heuristic may not.

- **Complete Domain Knowledge**: All relevant knowledge for the domain must be available so that system never encounters situations where insufficient information leads to spurious results.
- **Search-Control Knowledge**: In order to make inferences it is necessary to apply rules to knowledge. As the number of rules and the amount of knowledge increases the number of possible combinations of rules and knowledge can grow exponentially. Thus, a major pitfall for inference systems is a debilitating combinatorial search in a large knowledge base. Search-control knowledge tells a system the rules and facts are relevant in any situation. Without this kind of search-control knowledge complete domain knowledge is essentially useless because the system will expend all of its computational resources making useless inferences rather than the specific series of inferences required to reach the desired conclusion.
- **Comprehensive Domain Theory**: Complete domain knowledge and search control knowledge adds up to a comprehensive domain theory.
- **Narrowly Constrained Domain**: Complete characterization of a domain is typically only feasible for highly constrained domains.
- **Static Domains**: Evolving domains require human implementors to constantly introduce of new knowledge or rules. But, these additions often produce unanticipated interactions, and consequently, spurious inferences. Most contemporary systems do not have effective tools for locating these interactions in potentially complex process. Thus, debugging can be time-consuming and tedious. Rapid introduction of changes may mean the system is never successfully debugged.

Extreme domain-specificity is both the source of strength and weakness of knowledge-based systems. Whenever the task exceeds the range anticipated by its programmers, these systems fail. In the current generation technology, they have no recovery mechanisms. Even if their architectures are relatively flexible, current knowledge-based systems do not support learning or common-sense reasoning that could allow recovery. Moreover, there is little hope of adding effective learning mechanisms to these systems because they support no *informal level* of representation from which to learn.[27] In short, knowledge-based systems contain ungrounded, disembodied concepts. The kind of sudden failure behavior these system exhibit is called the "brittleness" problem. Brittleness is an important motivation for research on machine learning and related aspects of common-sense reasoning.

2.1.2 Codifying Domain Knowledge

During the 1970s, address space sizes and processor speeds limited the size of rule-based systems to between 50 and 300 rules. Recent advances have made it possible to build systems

[27] Non-deductive forms of learning, such as analogy and inductive generalization, require an explicit record of case-level or individual histories. Concepts are formed on the basis of this informal (pre-conceptual representation). Since concepts are microtheories that guide reasoning about the behavior of classes of phenomena, they can be updated from experiences to account for counter-examples. But, when a formal specification language is used to insert conceptual knowledge directly into a system, *without a mediating phenomenal grounding*, there can be no recourse to a phenomenal (or pre-theoretic) level to reformulate or revise defective theories.

with at least hundreds of thousands of rules, perhaps even a million or more. New hardware architectures also afford the possibility of matching rules in parallel, thereby gaining a substantial increase in efficiency. Thus, the hardware bottleneck of the 1970s has given way to the domain codification bottleneck of the 1980s. Codifying just a few hundred rules worth of knowledge is a significant undertaking in its own right. Culling several thousand is truly monumental. Since at least a couple of hundred rules is required to achieve adequate coverage (and interesting complexity) even for narrow domains, considerable effort must go into knowledge acquisition and codification.

Building a moderately complex knowledge-based system normally takes about two years. The usual approach is for "knowledge engineers" to "harvest" the knowledge of domain experts.[28] Their objective is to extract a comprehensive domain theory. Although experts may be able to function effectively in a domain, they are rarely, if ever, fully cognizant of the theory guiding their behavior; important components may be tacit (Polanyi, 1958; Kuhn, 1962), whether intuitive or unformalized. Since experts may not be able to articulate a comprehensive domain theory, the knowledge engineer must use clever techniques to identify and elicit the missing "informal" or tacit knowledge. One good way to extract this "informal" knowledge is by recording problem-solving protocols and systematically tracing the inference steps leading to the solution. Gaps in knowledge may also be exposed quickly by testing the incomplete expert system and analyzing its failure behavior.

2.1.3 Correctness and Validation

Theoretical computer science develops proofs of possibility (or impossibility) for computations that can be performed by computers. Typically these proofs are done for the general case and assume a universal Turing machine model (Hopcroft & Ullman, 1979; Boolos & Jeffrey, 1980; Lewis & Papadimitriou, 1981). A foundational theorem of computer science, known as the *halting problem*, proves that in the general case it is not possible to prove whether a universal Turing machine will halt (terminate). Because the universal Turing machine provides the theoretical foundation (the computational model) for work in computer science, this result effects *any* computational characterization of an AI system. If it cannot be shown that an AI system will halt (terminate), then the correctness of the system cannot be established because there is never any result whose correctness can be checked![29]

Correctness proofs for specific programs have thusfar succeeded only for simple algorithms. Correctness proofs for large programs, such as knowledge-based systems, other types of AI systems, or operating systems, are considered infeasible. This means that validation is only possible on the basis of *empirical induction*.[30] That is, given specific inputs, outputs (predictions) must be systematically compared to known results for the domain.

[28] The process of acquiring knowledge from a human expert relies on discourse. This presents the problem of accurate interpretation as the knowledge engineer translates the discourse to the target formal language. As knowledge-based system stray from the more consensual engineering domains to politics, the understanding process of the knowledge engineer becomes more problematic (Mallery, Hurwitz, Duffy, 1987). Thus, it becomes desirable to model the interpretation process of the understander in addition to the expert – a possibility relevant for natural language approaches discussed in section 5.3.1.

[29] The computational limitation of AI systems with relevance for political science modeling are discussed in greater detail in section 6.5.1.

[30] Empirical induction is discussed further in section 6.5.2.

If the system seems to be performing well as the number and variation of trials increases, we may *assume* that it will continue to perform correctly for the remaining possible inputs. But, unless all possible input combinations and associated outputs are in full agreement with domain data, *we are extrapolating (guessing) future performance from past experience.* One consequence of empirical induction as a validation strategy is that it becomes less effective as system size and domain complexity increase because the number of possible outputs and their mappings to the domain increase to unmanageable proportions. This means that knowledge-based systems for unconstrained domains will be exceedingly difficult to validate.

Medicine is often touted as a success story for knowledge-based systems, and therefore provides a good example. While medical diagnosis was among the first pioneering applications of rule-based systems and they have even shown some promise, doctors do not act without first verifying the diagnosis; medical expert systems are never left alone with the patient (Schwartz, Patil, Szolovits, 1987). Indeed, nearly all the successes of knowledge-based systems have been outside medicine, typically in commercial applications where the cost of failure is relatively low and the domain highly constrained. The accepted wisdom is that knowledge-based systems can be good descriptive prosthetics for theories and human inferencing in narrow domains but that they cannot be entrusted with any critical decisions or predictive tasks no matter how well they might seem to work.[31]

2.2 Toward Domain Independence

The major challenge in AI research today is to move beyond "microworlds" to systems with multiple domain generality, and finally, onto domain-independent[32] system capable of adapting to unconstrained everyday domains. This move demands architectures with informal levels of representation to ground multimodal learning. Systems will have to represent the phenomenal world and use common-sense reasoning strategies, like those people are believed to use, to independently learn about domains that are too complicated for knowledge engineers to program into the system. Two important research areas for domain independent learning are inductive generalization and analogical reasoning.

[31] Reasons supporting this conclusion appear through this report, and specifically, in section 6.5. They cluster into computational limitations of correctness, completeness, and decidability of inferential algorithms as well as difficulties in the codification of domain knowledge arising from sources such as vague or evolving domain boundaries, dependence on other unanalyzed domains, or the presence of unrecognized tacit knowledge.

[32] *Domain independence* refers to the independence of universal inference processes from the specific domain knowledge processed. It is intended to denote AI systems that can function in any domain precisely because their operation does not rely on domain knowledge that is wired into the design of their inference engines. It does not refer to whether the inferences performed depend on the specifics of a domain; this is always the case. The description of how or what inferences to perform is a higher order level that talks about the processing of information. The domain knowledge is object level or the content of the information to be processed. If the inference process depends (the higher-order level) *directly* on the specific content of the information (the object level), a system cannot be applied to domains without the same specifics. This means that the ability to function in multiple domains arises only when the inferential processes operate on the basis of the form of knowledge rather than the specific knowledge itself. This means that the inferential mechanism is an epistemological system designed to function according to the ontology of the representation for knowledge.

Inductive generalization, or concept acquisition, seeks to formulate characterizations of concepts by generalizing from specific examples (Dietterich & Michalski, 1983; Michalski, 1987).[33] For example, a system might learn the concept of an arch from exposure to a series of positive and negative examples (Winston, 1975). The technique can be applied, in principle, to more complex objects. If a system learns concepts, it follows that it should *classify* (Brachman & Levesque, 1984; Brachman & Schmolze, 1985; Haas, 1987) any new examples encountered as instances of known concepts.

Another more complex type of inductive generalization is *explanation-based learning* (Mitchell, *et al.*, 1985). Here, a system induces a theory from a single example of a problem-solving episode, and subsequently, refines the theory to fit other cases. The essential idea is that an AI system can acquire domain-specific inferential abilities by copying a single example. Explanation-based learning is quite similar to the precedent learning (discussed below and in section 4.4) but in the refinement process it generalizes the actions. Explanation-based learning is particularly well-suited to modeling of social processes, where demonstration effects are known to be prevalent and copying the strategy of another actor is a legitimate strategy for those with the necessary capabilities (Bennet & Alker, 1977). Considerable work has been done in this area. Since some large scale applications have been done (Dietterich & Michalski, 1983), the main research issue is developing domain independent representations and efficient induction strategies. A representation system capable of representing natural language meets the domain independence requirement. While these ideas have thus far received only limited application in natural language problems, their potential contribution is large because common-sense induction processes are ubiquitous in language (Mallery, 1987a).

Analogical reasoning is another important mode of learning for domain-independent systems that is believed to be closely related to human "common-sense" reasoning. Here, the idea is to transfer problem-solving expertise from one domain to another on the basis of analogy. In one of the most advanced systems, Winston's analogy program recursively applies analogies to solve problems and learn new concepts (Winston, 1984). Winston's program is particularly impressive because problem statements and knowledge are acquired directly from natural-language texts. More recently, Carbonell (1986) has shown how difficult problems in physics can be solved by a kind of analogy he calls "derivational problem solving." Gentner (1983; Gentner & Landers, 1985; Falkenhainer, Forbus, & Gentner, 1986) has proposed a theory of analogical reasoning based on structure mapping, in which semantic structures from one domain are mapped to another domain. Gentner advocates "systematicity," the presence of higher order unifying relations, as the criterion for good analogies.[34] Analogical reasoning and precedents logics have high relevance for social science as is developed later.

The emphasis on learning represents a quest for a genetic or evolutionary

[33]In international relations, Alker and Christensen (1972: 195-196) not only demonstrate an awareness of the problem of "concept attainment" in the psychological literature but also recognize its importance for precedent-based reasoning. In fact, their precedent logics system actually performed some feature-based inductive generalization to match different situations.

[34]Gentner's notion of analogy seems more of a cross-domain mapping from one domain theory to another individual case because participation in higher order relations is tantamount to participation in a theory. My view is that this type of mapping is actually a deductive mapping that violates the legitimate range of the theory mapped. I hold that analogy is a mapping from an individual case to another individual case, independent of a theoretical abstraction. Although, I agree with Winston (1984) that the process of analogy leads to abstraction of theories spanning the source and the target domains.

epistemology in which AI systems evolve better problem-solving strategies to cope with the world. Inductive generalization and analogical reasoning are important ways of learning. There are certainly more ways of learning, and certainly much more to AI, than presented here. Readers interested in greater detail and other topics should consult the sources noted above. The reason for emphasizing inductive generalization and analogy is that they are symbolic processes that rely on the history of an understander and they thus illustrate the hermeneutics (Mallery, Hurwitz, Duffy, 1987) of learning: Namely, that the concepts acquired or the analogies drawn depend on the specific history (*effective history*) of an understanding system. If concept acquisition and analogy (or precedents) constitute a major source of learning in human cognition, it follows that classifications, interpretations, and subsequent understandings follow from the individual histories of people and groups.

The learning processes (and also the inference processes associated with knowledge-based systems) are symbolic, not numeric. This is the major difference between modern AI and traditional uses of computers for numeric analyses. Indeed, it is the ability to devise symbolic models of politics that is of interest for international relations and foreign policy decision-making. Instead of manipulating numbers like traditional computer applications, these AI technologies manipulate symbols and symbolic representations. Because quantitative methods are already in widespread use in international relations, symbolic processing offers the greatest potential for new applications to international relations modeling problems. In the next section, we shall see how these symbolic modeling techniques may contribute to the formal modeling arsenal of political science.

3 Computational Politics

Polimetrics was used by Alker (1975) to denote formal modeling techniques for political science, whether quantitative or symbolic. According to Alker (1975: 147), "... polimetrics is the application of mathematical forms and statistical techniques [or procedures] to [the qualitative labeling and quantitative metricizing of possibly observable] political phenomena, the scientific testing of political theories, and the solution of present and future political problems." Even if this definition leans toward quantitative methods based on behaviorally observable data, it is broad enough to cover symbolic modeling techniques. In fact, Alker includes within polimetrics intentionalist methods such as game theory and cognitive models based on "conceptual dependency" (CD is discussed in section 5.2).

Three issues require attention if symbolically-based AI methods are to be smoothly integrated within polimetrics.

- The general orientation towards stochastic methods needs to be balanced with the notion of computational simulation, or the qualitative use of a universal Turing machine to simulate aspects of political phenomena.
- AI methods attempt to model cognitive processes that are not directly observable. Behaviorist correlational studies of the surface phenomena can neither identify nor confirm the "deep structure" responsible for the surface manifestations. Thus, verification must rely on indirect confirmation strategies building from functional connections between the the underlying processes and their surface manifestations.

- The strong emphasis on quantity needs to be offset by an equally strong emphasis on quality.

The term *computational symbolic polimetrics* captures the relevant distinctions. "Computational" distinguishes AI techniques from non-computational formalisms or descriptive languages such as logic, set theory, game theory, and argument analysis. "Symbolic" stresses the qualitative, non-numeric orientation of AI techniques. "Metrics" should be retained because it captures the idea of finding dimensions, in this case, qualitative dimensions. Since "computational symbolic polimetrics" is too hard to say, an abridgment to the more colloquial *computational politics* makes sense. Computational politics, thus, differs fundamentally from the stochastic and numeric approaches conventionally associated with polimetrics. A concise definition could be: computational politics refers to the technique and practice of formulating, implementing, interpreting, and evaluating computational models of political phenomena.[35]

The reason to include computational politics in the methodological arsenal of political science is to widen the range of political phenomena amenable to formal study. Social processes, including international relations and foreign-policy decision-making, can be divided into two broad classes: Material processes and ideational processes. The basic difference is that ideational processes involve the cognitive processes of individuals and groups. Although quantitative methods have produced striking successes in the study of material processes, they have not demonstrated similar successes in cognitive application, largely due to an ontological mismatch.[36] Cognitive processes are inherently symbolic (see section 3.2), and consequently, require formal methods suitable for simulating and analyzing symbolic processes. In international relations and foreign-policy decision-making, inquiry often focuses on cognitive processes of individuals, groups, organizations, and countries (George, 1959; Steinbruner, 1974: esp., 88-139; Etheredge, 1985; esp., 141-169). Specifically, political scientists are interested in why decisions (choices) are (should be) made in order to determine the actions of international actors as well as the consequences for international systems and specific actors within it. Choices are cognitively grounded in a political actor's beliefs, intentions, histories, social norms, "operational codes," and political orientations. As a problem of political cognition, the methods of computational politics provide a broad spectrum of techniques suitable to the formal study of political decisions.

[35] *Model interpretation* refers to establishing the correspondence between the elements and relationship within the model and their external referents in the world. Since the complexity of a model must, by definition, be less than that of the phenomena modeled (Ashby, 1956; Deutsch, 1963), interpretation must also identify the points at which the model no longer corresponds. Some specific issues interpretation needs to address include strategic simplification (section 6.1.7) and equifinality (section 6.1.8). *Model evaluation* refers to the more general problem of assessing the overall "goodness of fit," particularly the representational accuracy of the model (see section 6).

[36] The ontological mismatch is independent of any divergences in social ontologies. But, an ontologically adequate modeling tool will capture different social ontologies.

3.1 Reasons For Turning to Symbolic Modeling

3.1.1 Restrictions on Stochastic Methods

Stochastic methods, heralded by some as the hallmark of "a scientific social science," yield valid results only when their fundamental assumptions are satisfied. Since longitudinal studies in international relations or specific decisions types face problems of continuing changes in the international system, including economic, political, cultural, social, and technological change, stochastic methods face cases where n = 1 and basic assumptions fail. Limits on degrees of freedom prevent inclusion of large numbers of qualitative or dummy variables; attempt to capture the full range of complexity in international phenomena quickly exhaust the available degrees of freedom, forcing significant variables to be dropped. Of course, stochastic methods can effectively perform a type of extrapolation between the available data points, revealing outlines of complex systems and predicting system behavior based on inertia. But these successes also inherently lead to poor performance on the low-probability non-linearities that often account for major system transformations and policy decisions by major political units (Crecine, 1969; Alker & Christensen, 1972; Alker, 1975, 1984; Schrodt, 1984b, 1985).

3.1.2 Supplementing Differential Equation Models

Various scholars (Alker, 1976; Bennet & Alker, 1978; Alker, 1979a; Anderson & Thorson, 1982; Sylvan, 1987a, 1987b) argue that AI methods of symbolic modeling can supplement the differential equation models, such as those found in world models (Forrester 1971; Meadows, *et al.*, 1972; Mesarovic & Prestel, 1974). While dynamic equations may capture feedback relationships in macro economic processes, they do not provide an adequate account of decision-making by major political units. Decision-making by major political units is a form of goal-seeking behavior under imperfect or distorted information (Alker, 1979b). It may often be guided by historical precedents or constrained by normative considerations. Symbolic modeling techniques provide alternative mechanisms suited to modeling decision-making as a symbolic, cognitive phenomena. They are also suitable for modeling the influences of pressure groups (sectors), norms, and other qualitative influences on decision. They seek to develop better global models by incorporating new decision components that model the cognitive aspects of decision-making by the political units. More generally, this move toward *hybrid models* suggests a trend toward specialization and division of labor within complex international models. By applying techniques to those areas in which they are most effective, a hybrid model can eliminate weaknesses of each and benefit from the strengths of all.

3.1.3 Intentionality

A means for modeling intention is perhaps the major impetus underlying the turn to symbolic modeling technologies (Alker, 1970; Alker & Christensen, 1972; Alker, 1975, 1976, 1979a; Anderson & Thorson, 1982, Thorson, 1984; Ensign, 1985a; Sylvan, 1987a, 1987b). Although certain sophisticated stochastic techniques[37] may contribute some insights into intentional

[37] See (Alker & Christensen, 1972; Alker, 1975) for discussion of intentionalist statistics.

variables, they provide no adequate model of the rich representations and broad inferential strategies involved in intentional cognition. *The shortfall is due primarily to the behaviorism inherent in the method; external indicators of intention were linked with behavioral output but the actual symbolic processes were not modeled.* Thus, even if attention to some "intentional variables" could improve prediction it did not open the "black box" of cognitive processing. In contrast, symbolic modeling using precedents or rules focused squarely on the cognitive processing, including the formulation of goals and the composition of plans to realized them. An intentional symbolic model brings to bear both general and case-specific theories of information processing in the organization or cognitive processing in the individual with the external inputs and traces through the processes yielding the behavioral output. Thus, while behaviorism may be intrinsic to various quantitative methods (because they cannot account for information processing within political actors), symbolic models are post-behaviorist because they emphasize cognitive information processing. Moving beyond behaviorism is a major reason for turning to symbolic methods.

3.1.4 Economic Rationality

Perhaps the major received account of social rationality is the "rational actor model."[38] Borrowed largely from economics, the rational choice model holds that individuals and or organizations are utility-maximizing decision-makers.[39] Their decisions putatively "reveal their preferences," or underlying indifference curves. In fact, this view holds that rationality is utility maximization. This approach models rationality with probabilistic expected utility models. But reduction of knowledge and cognitive processes to numeric indices lacks cognitive plausibility. It lacks cognitive plausibility because individuals and organizations make decisions based on plans, inferences, and knowledge.[40]

Understanding rational action has been a longstanding problem for international relations as it attempts to explain decision-making by nation-states and other international actors. While some IR theorists have imported the "rational actor" model from economics, AI researchers (*e.g.,* Cohen & Levesque, 1985, 1987a, 1987b) have endeavored to develop structural models of purposeful behavior. These models attempt to explain how an intelligent agent formulates and executes intentions, goals, plans, and actions. In the structural model of rationality, the cognitive processes such as learning, common-sense reasoning and concept formation are presumed to operate on a large semantic representations accruing from the agent's history. The two approaches are very different: While economic rationality is numerical and information poor, the structural rationality is symbolic and information rich.

The community of researchers developing symbolic models generally does not discuss the rational actor paradigm. They vote it down by pursuing other approaches. Those writers who address the question consider economic rationality inappropriate for international relations applications (Schrodt 1984b, 1985; Sylvan, 1987a, 1987b). Sylvan (1987a,

[38]Allison (1971: 10-38) and Steinbruner (1974: 25-47) overview of the rational actor model with specific reference to foreign-policy decision-making.

[39]Beuno de Mesquita (1981) is a recent application of "rational choice" to international relations.

[40]Of course to the extent that organizations use expected-utility models and other stochastic approaches to model their environments, and then, feed the results into their decision-making, their behavior approaches the predictions of "rational choice."

1987b) shows an example of how intentional inferencing performs better than economic rationality because it captures certain non-rational features of decision. An illustrative selection of objections to rational choice theory which are taken as reasons for pursuing symbolic modeling approaches are:

- The international system is not an equilibrial system. System transformations occur before an equilibrium is ever reached. Thus, methods suited for describing equilibrial systems are inappropriate (Schrodt, 1984a, 1985).
- Choices are made only once. The absence of comparability makes the application of expected value calculations dubious (Schrodt, 1984a, 1985).
- The psychological evidence indicates people make decisions in ways substantially more complicated than expected utility calculations. Decisional processes are information rich and socially influenced (Bennett, 1981; Schrodt, 1984a, 1985).
- The rational choice theory is not falsifiable; it is offered as a normative model of how people *should act* rather than how they *actually act*. Thus, while there is negative evidence, there is no empirical support for rational choice models (Schrodt, 1985).
- The logics expressible within the representational ontologies of rational choice do not fit the phenomena (Diesing, 1962; Habermas, 1979, 1981; Bennett, 1981; Alker, 1984, 1986, 1987). Human rationality is communicative (Habermas, 1981) and symbolic (Newell, 1980; Simon, 1985). It involves formulating goals on the basis of discursively expressed and validated premises and organization of these premises into arguments considered coherent by other members of the social group.[41]

3.1.5 Rigor, Counterfactual Simulation, Inspectibility

Like conventional mathematical modeling techniques, intentional symbolic models are formally and precisely specified. This yields rigor and replicability. The combination of a representation for knowledge and an inference engine makes it possible to modify aspects of the model and run counterfactual simulations. Sensitivity analysis of counterfactual simulations can disclose key components of the model responsible for specific aspects of an actor's behavior. More generally, symbolic models involve a series of functional transformations on a representation. Since they are explicitly represented, these transformations are available for detailed inspection, cumulative criticism, and subsequent refinement.

An important inspection technique involves verifying the logical consistency of inferential chains. A deductive model, such as one based on theorem proving, guarantees the veracity of its inferences. Non-deductive models, such as systems relying on heuristics or analogy, may perform erroneous inferences. In these cases, a *truth maintenance system* (TMS), also sometimes referred to as a *reason maintenance system*, can be used to detect the introduction of contradictions into the representation.[42] *Contradiction detection* is an

[41]Rationality or the process of formulating plans and making choices falls under cognitive processing. The case of political cognition is discussed in section 3.2.

[42]Truth maintenance systems create a dependency graph to track the justifications inferences. Sophisticated TMSs can do certain types of theorem proving. A recent special issue of *Artificial Intelligence*, 1986, volume 28, focuses on technology for truth maintenance systems.

important analysis to perform on large theories or data sets. Similarly, *gap detection* locates gaps in theory or data when inferences cannot connect premises with conclusions (Mefford, 1986d). In general, symbolic methods afford the possibility of detailed inspection of models, including the theories and data they incorporate, in ways a human might employ to evaluate theories and data, but with greater systematicity and methodological variety.

3.1.6 A Mental Prosthestic

Formulating symbolic models provides a means to clarify theories in the course of the model building exercise. In AI research, good theories are generally wrong but they are good enough to guide an implementation. Once in place, the implementation helps identify the weaknesses and underspecifications of the theory. Thus, cycles of theory extension or reformulation and implementation lead to convergence on correct, computational theories. A good theory, then, is a beginning for an implementation, whereas a bad theory cannot be implemented in any useful way. The reason that theories are not right from the start is that the phenomena to be explained are too complicated for *a priori* theorizing to succeed without the power of a strong deductive tool. While mathematics provides such a tool for physics, symbolic modeling provides a similar tool for sciences concerned with non-numeric "symbolic equations." The importance of a deductive formalism[43] is that it enforces consistency or checks inconsistency between distant parts of the same theory. While numerically grounded equations perform this function for physics and other natural sciences, computational models may perform an equivalent function for symbolically structured domains. Thus, computational models enforce systematic relationships between distal components of a symbolically-based theory. Computational models also allow examination of the logical structure of theories. In general, the rigor of formalizations exposes weaknesses that may lead to theory revisions that improve accuracy and breadth of coverage. Thus, symbolic models of foreign policy decision-making and international relations provide a strong deductive formalism that can help scholars cope with the inherent complexity of their subject and hone their theories to better approximate the phenomena. In this use, computational models constitute a mental prosthestic.[44]

3.1.7 Evolutionary Epistemology

Learning and non-learning is an important part of what political actors do in response to (or occasionally in anticipation of) changes in the international system. Modeling the learning and the ideational restructuration of individuals, organizations, and other social collectivities seems possible, in principle, only by means of symbolic modeling. In the literatures on psychology (Campbell, 1960, 1974) and philosophy of science (Popper, 1974, 1977), "evolutionary epistemology," or the processes by which biological systems and cog-

[43]Note that the deductive formalism need not model deduction. It may mode other inferential processes, including learning. It is a deductive formalism because it builds on a base (the computer) whose operation is necessarily deductive and syntactic – even if it can simulate at higher levels non-deductive and non-syntactic processes.

[44]Mefford's (1986d) report on his computer implementation of balance of power is an illustration of using a computational model as a mental prosthestic.

nitive systems accumulate knowledge, provides an account of the growth of knowledge.[45] This formulation of evolutionary epistemology is, however, precomputational. In the AI literature, evolutionary epistemology is partly treated by learning and models of the scientific discovery process.[46] The machine learning literature emphasizes inductive generalization, concept acquisition, theory formation, discovery, analogy, metaphor among other computationally-grounded learning strategies. A synthesis between the computational and the pre-computational approaches promises to supply better cognitive (and perhaps genetic) learning strategies than "blind variation and selective retention."[47] The prospect of simulating the learning of political actors, and even, the learning of the social-scientific investigator, will lead to a major transformation in the formal modeling of international relations and foreign-policy decision-making.[48]

3.2 Modeling Political Cognition

Modeling tools are rarely theoretically neutral. When the tool is multivariate regression, the data is time series and cross-sectional variables. International events data clearly reflects this orientation. However, when the tool is a symbolic processing technique, such as a rule-based system, the data becomes symbolic characterizations of decision rules for the principle international actors. This suggests a change in the meaning of international events data. The remainder of this section explores the reconceptualization of the modeling task that necessarily accompanies the advent of symbolic modeling. It is intended to sensitize the reader to this crucial issue before the following section reviews the implemented models.

Political cognition divides into cognitive processes at the level of the individual and social collectivities. Both levels are interpenetrating. Political cognition in the individual is constrained by his embedding in society (Bennett, 1981; Habermas, 1981) while political cognition in collectivities builds from the cognitive processes of individuals. The two key features of political cognition are:

[45] For a recent edited volume containing articles by the main proponents of evolutionary epistemology in psychology and philosophy, see Radnitzky and Bartley (1987). Bartley (1987) provides an overview of evolutionary epistemology.

[46] See Michalski, Carbonell, & Mitchell (1983, 1986) for recent overviews of the machine learning literature.

[47] The computational difficulty with blind or random variation is that it implements an undirected and unconstrained search. Without massive parallelism, such a brute force search is unlikely to converge at reasonable rates. Greater constraint in the sources of variation, such as Waddington's (1976) epigenetic landscapes (topologies of possible evolutionary trajectories), are necessary to reformulate evolutionary epistemology for greater computational plausibility.

[48] Although precomputational, Etheredge (1985: 66) provides a definition of organizational learning and sets out to understand foreign-policy decision-making from the perspective of organizational learning. Etheredge assesses intelligence in terms of:

- *realism*, recognition of the different elements and processes at work in the world;
- *integration* of the different elements and processes in coherent thought;
- *reflection* upon the first two processes (sometimes called *deutero learning*, or learning about learning).

- Individual cognition is grounded in "procedural rationality" (Simon, 1985), a non-stochastic, symbolically-mediated, functional account of human reasoning processes, where these reasoning processes are conceived as "physical symbol systems" (Newell, 1980).

- Social cognition is fundamentally a linguistic phenomenon, centered on interpretations of the meanings and intentions of social actors (Deutsch, 1953; George, 1959; Gadamer, 1960; Alker, 1975, 1979a, 1986; Habermas, 1979, 1981; Bennett, 1981; Dallmayr, 1984; Shapiro, 1984).

Modeling political cognition recognizes that Man is not merely *homo faber*, but is, perhaps more fundamentally, *homo loquens* (Dallmayr, 1984). People live in linguistic communities and exchange information through language. Significant action can be treated much like linguistic statements found in texts, because both convey meanings and are subject to interpretation that reveals motivating intentional structures (George, 1959). By building representations of intentionality in addition to belief, it becomes possible to model social action (Alker, 1971; Ricoeur, 1971; Searle, 1983; Gilbert & Heath, 1985). The beliefs and intentions underlying planning and counterplanning finds their expression in strategic language (Mallery, 1987b). Strategic language expresses not just the beliefs and intentions of the speaker but also those attributed to competitors, and reflexively, those the competitor attributes to the speaker. For international relations and foreign-policy decision-making, the modeling task builds from this kind of language – even if most standard logics cannot express statements of embedded belief (Mallery, 1987b).[49]

[49]Strategic language is not explicitly included within Habermas' (1979: 41, 117-119; 1981: 94-101, 305-319, 333)) formulation of strategic action or communicative action. For Habermas, *strategic action* refers to the case of competing opponents who are each attempting to influence each other's decisions in a purposive-rational way. *Communicative action* is a kind of discursive practice for reaching understandings, or agreements, about legitimate norms governing social action. Habermas restricts the linguistic grounding for his discussion to speech acts, ignoring the more general issue of belief or modal contexts and their role in the empirical knowledge mustered to justify choices. The notion of strategic language could be incorporated into Habermas' notion of communicative action but it might have significant consequences for the theory, such as eliminating the possibility of an objective truth independent of the subject (favoring Gadamer in the Habermas-Gadamer debate (Mallery, Hurwitz, Duffy, 1987: 366-367)), changing the focus of validity redemption to discourse about the opacity of belief contexts, improving the analysis of validity claims by grounding it in belief contexts rather than speech acts – which do not cover all the linguistic cases where belief contexts effect validity. More importantly, revising his theory to account for belief contexts and the associated opacities (see section 6.1.3) would poses difficulties for Habermas' existence claim for an objective truth, and consequently, would compel Habermas to accept some version of Gadamer's position in the Habermas-Gadamer debate! Grounding his work in a linguistic analysis limited to speech acts presents significant difficulties for his implicit theory of internal structure of individual action and the edifice he builds on top of it. Recent work by Cohen & Levesque (1985, 1987a, 1987b) suggest that speech acts are not primitive and that an underlying theory of individual action can be developed from more basic constructs. The position I maintain in my conception of strategic language is presently agnostic concerning the specific theory of individual action but holds that strategic language reflects significant elements of such a theory, including the patterns of inference from belief contexts derivable from it.

3.3 Limitations of the Monological Model

Traditionally, most AI researchers have assumed an asocial or monological model of the cognitive agent (Chomsky, 1965, 1981; Minsky, 1987; Batali, 1987) in which cognitive processes are reformulated and studied independent of their social embedding. Apel (1980) argues that the monological model, specifically Chomsky's linguistic research programme, is inadequate for exploration of cognitive competences and for use in social science because it ignores the role in cognitive processes, whether linguistic or ideational, of culture, conventions, social role-playing and norms. In short, it ignores the social embedding of the cognitive agent. Apel's critique centers around *a priori* grounds for communication between social actors. The argument easily generalizes from an attack on transformational syntax to an attack on single-agent cognitive science or standard AI research. By ignoring intersubjectivity, typically linguistic interaction with other cognitive agents in society, the monological model fails to problematize differential understanding of the cognitive agents as a function of their relative embeddings in society. While a monological model might be adequate for studying aspects of the *a priori* structure of cognition, it fails as a scientific research design whenever social interaction comes into play, such as in the acquisition of linguistic capabilities, stages of cognitive or moral development, or more generally ideational development.[50]

In all fairness, it should be noted that the monological approach is not monolithic within AI. Some prominent figures in AI, *e.g.*, (Winograd, 1980; Winograd & Flores, 1986), are explicitly sensitive to social constraints on cognition (Bennett, 1981; Klockner & Bennett, 1987). Others find intersubjectivity intimately related to their research, particularly in natural language utterance planning (Appelt, 1985) and theories of rational action (Cohen & Levesque, 1985, 1987a, 1987b). In the information processing context of the office, Hewitt (1986) addresses the problem of distributed computers systems (computer systems with different informational embedding) attempting to talk to each other but meaning different things with the same terms. But these efforts typically do not go as far as social scientists might desire. Thus social scientists, particularly but by no means exclusively, ethnomethodologists and social psychologists, can contribute desiderata for symbolic modeling on basis of the social constraints on individual psychology. More generally, the recognition that intelligent agents have acquired a history in a specific social milieu is crucial for modeling the concrete ways in which they understand the world.[51]

There are two ways in which the monological model needs to be transcended. First, models of cognitive agents must incorporate histories for the agents. In contrast to typical engineering domains where a single objective interpretation often seems self-evident, social systems, and in particular international relations, are riddled with examples of divergent and incommensurate interpretations of the same phenomena. This divergence of interpretations is explained by a fundamental insight of the German philosopher Martin Heidegger (1927), which was elaborated and extended to language by his student Hans-

[50] Importing tools from AI into social science methodology will always carry the danger of distorting the phenomena to be modeled because of the biases engineered into the tool; the social scientist must select tools suitable for the modeling problem, or if the do not exist, create his own, *e.g.,* Relatus (Duffy & Mallery, 1986).

[51] A dialectical alternative would emphasize the collectivities over the individuals and understand intelligent agents as instantiations of constraints specified by their cultural and social embedding. This view, however, encounters difficulties because it easily loses touch with the grounding of society in individuals and the exigencies of everyday life that compel them to act.

Georg Gadamer (1960). The insight is simply that understanding is inherently subjective because it must always be based on the information available to the understander from his experiential history in the world. In international relations, this problem has surfaced as "differential perceptions" (Jervis, 1976). It is axiomatic that ideological, cultural, social, sectoral, and even organizational, differences can produce differential understandings. Modeling these differences demands natural language representations of the beliefs and interpretations of understanders based on their individual histories. These representations must allow for biases and distortions in understanding due to differential histories.

Second, once history-sensitive symbolic models become the goal, the data collection task becomes acquiring documentary traces for the histories of political actors. Here, detailed case studies provide the means. In sum, symbolically modeling social phenomena demands an alternative to the monological model, in terms of both the characterization of cognition and the requirements of detailed case histories. This alternative has been dubbed *computational hermeneutics* (Alker, 1975; Alker, Lehnert, Schneider, 1985; Duffy & Mallery, 1986; Mallery, Hurwitz, Duffy, 1987).[52]

3.4 Social Systems as Patterns of Conversations

The external corollary of *homo loquens* is a "conversation processing" view of social functions and interrelationships. The emergence of this view is reflected by a growing momentum of linguistically-based studies of organizations, societies, politics, and international relations (Lasswell, Leites, *et al.* 1949; Leites, 1951; Lasswell, Lerner, & Pool, 1952; Deutsch, 1953; George, 1959; Habermas, 1979, 1981; Dallmayr, 1984; Onuf, 1985, 1987; Duffy & Mallery, 1986; Mallery, Hurwitz, Alker, Duffy, 1986; Alker, 1986; Winograd & Flores, 1986). If social systems are "conversation processing" entities, their informational currency is primarily natural language utterances and actions – to which interpretations can be assigned. Through complex conversational patterns, or "social talk" (Alker, 1986), societies organize (re-)production of not just their material bases but also the practices and structures of rights and responsibilities (Harre, 1985) in the socio-political superstructure.[53] Socially significant actions are typically embedded in communicative situations in which the action can be interpreted as a speech act whose modality has been shifted. Thus, the received cybernetic dichotomy between action and control information loops (Ashby, 1956; Deutsch, 1963; Steinbruner, 1974) is superseded by a unifying framework in which action and control streams are understood within a conversational model of linguistically-mediated communication. In short, language provides the tool that intersubjective communication uses to organize social action and reify conventional practices in culture. For international relations,

[52]Some readers might believe that functional simulations of AI phenomena might be indifferent to the monological model because no interpretation seems involved. This argument will stand as a crude approximation. However, modern social methodology (Alker, 1981) stresses the need to situate the subject (the observer) within the analysis so that the biases inherent in the subjects position in society can been considered by readers (receivers of the wisdom). Thus, the monological model fails to withstand the test of methodological self-reflectivity.

[53]Society as a pattern of self-producing conversations is an idea that traces from the notion of autopoeisis found in (Maturana and Varela, 1980), its application to cognition (Maturana, 1970, 1977), and its extension and elaboration within social contexts by Winograd and Flores (Winograd, 1980; Flores, 1981; Winograd & Flores, 1986).

conversations may range from diplomatic exchanges, negotiations, and deterrent statements (Bennett, 1987) through war – violent communicative action or a diplomacy by other means (Clausewitz, 1832). Within a theoretical framework that considers the fundamental informational unit of social systems to be conversations and beliefs derived therefrom, ontological adequacy demands a modeling technique suited to processing natural language utterances and felicitously representing beliefs.

Granting the view that social systems are (re-)constituted by their conversations (including internalized conversations, *i.e.*, thinking), what are the content of the conversations? One approach to this question is to consider the conversational ecology of the system and the informational niches it contains. Productive, sectoral, class, geographic, ethnic, and affilliational divisions provide a communicational matrix within which individuals and groups interact with each other and the rest of society. Paralleling the division of labor, the pattern of social communication calls forth and defines rules for appropriate (inter-)action within each niche.[54] Different conceptual structures may arise as categories are induced or abduced from the differing historical experiences of individuals and collectivities (Luria, 1974; Sergeev, 1987b). In short, participation in particular conversations reflects, reproduces, and transforms conceptual differentiations in the social and organizations systems. This conceptualization demands a methodology that is sensitive to the informational differentiation reflected in the division of labor as well as other ideational differentiations in social systems, specifically:

- The ability to model the linguistic histories of political actors;
- The ability of learn (induce) the categories structure that political actors would induce from their histories;
- The ability to locate and apply precedents, analogies, and metaphors whose applicability (and existence) depends critically on the specific history of the political actor;
- The ability to simulate not just the problem-solving behavior of a political actor on the basis of their intentions and goals as they arise within the actor's specific history.
- The ability to simulate the political actor's interpretation of other purposive agents in the environment.

Although this manifesto for a hermeneutically-informed computational politics may seem farfetched, advanced work in this field is not just targeted in this direction but actually developing these kinds of capabilities.

[54]Some approaches have proposed "schemas," (Lau & Sears, 1986) and "scripts" (Schank & Abelson, 1977) to capture stereotypical mental models for niches but these typically suffer from assumptions of categorial universality or the absence of dynamic reformulation as learning responds to changes in the communicative matrix. Habermas (1981) refers to stereotypical roles associated with informational niches as "dramaturgical self-presentation." Harre (1985) discusses the "ruling following" (Wittgenstein, 1953) associated with these niches, considered as sociological categories.

4 Models Using Artificial Languages

This sections reviews the research using artificial languages to model political phenomena. It covers information processing models, cognitive mapping, knowledge-based systems, and precedent logics.

4.1 Information Processing Models

The information processing models of Simon and Newell and their students (Simon, 1969; Newell and Simon, 1972; Crecine, 1969; Simon, 1985) are an early tradition of symbolically-based work in artificial intelligence with political relevance. These models attempt to model social phenomena through the lens of problem-solving and planning behavior as formulated in their "general problem solving" (GPS) (Newell, Shaw, & Simon, 1957, 1959). GPS involves applying *means-ends analysis* to find a path from an initial state to a desired final state. Information processing models aim at elucidating goal-seeking behavior using notions of "bounded rationality" and "satisficing." Goal-seeking behavior is formulated as a search through a *problem space*, the set of all possible paths from an initial state to a final state.[55] Since computers, individuals, and organizations can deploy only finite computational resources, all possible aspects of the problem space cannot be examined. Hence, "bounded rationality" expresses the inability of real decision-makers, in contrast to idealized "rational" decision-makers, to consider all possibilities. If a decision-maker does not consider all possible problem solutions, he must select an adequate one as a function of the order in which solutions are enumerated. "Satisficing" captures the notion of accepting an adequate problem solution as it is found even though it may not be optimal. GPS, then, provides both a computational technique for problem-solving, and perhaps more importantly, a richer vocabulary for describing computational processes found in organizations and other social processes. Indeed, the work of Newell and particularly Simon constitutes an early transfer of computational vocabulary to the social sciences with far-reaching effects.

Despite its conceptual successes, GPS faces series computational difficulties. In practice, GPS style problem-solving is intractable because it involves search in an exponential problem spaces.[56] Thus, it is only practically computable for the simplest problems. Even if it evinces poor computational properties, GPS provides a highly general conceptual framework for understanding both human and computer problem-solving. More recent work aims to develop problem-solvers that are both computationally tractable and cognitively plausible. A modern descendant of GPS is SOAR (Laird, Newell, & Rosenbloom, 1987). Building on top of OPS5[57] (a production rule kernel), SOAR remembers and re-uses successful problem solving sequences in new situations. This remembering process is termed "chunking." In cases when these "learned" chunks do not work for a novel problem situation, the system introduces variations into the "chunked" problem-solving sequence. Once

[55] GPS and problem spaces are discussed in any introductory textbook on AI, *e.g.* (Winston, 1984).

[56] Although other ways could be used, one way to formally demonstrate the intractability of GPS is to show that it can implement non-linear conjunctive planning, which is known to be computationally intractable and also undecidable (see section 6.1.4).

[57] Students of rhetoric will appreciate that OPS5 stands for "Official Production System," version five. Brownston *et al.* (1985) provide an overview of OPS5.

a successful variation is found, the old "chunked" problem-solving sequence can be refined to identify the novel situation and apply the new variation to it. In this architecture, computational resources are conserved by applying methods known to work all at once without any search. Just as the earlier GPS offered conceptual insights for macro-social phenomena so too does the novel SOAR architecture. Specifically, it argues for convergence with other traditions that have focused on precedent-based models of organizational and individual decision-making (see section 4.4). The reason is simple. Once an organization has solved a problem by whatever means, it is far simpler to apply the stock solution rather than to rederive the solution from first principles. Thus suggests a corollary to "satisficing:" In seeking a solution to a problem, organizations will accept any known precedent from the organization's history which applies and solves the problem adequately. As a further note, Allen Newell has recently argued to philosophers and psychologists that SOAR represents a plausible architecture for a unified theory of cognition.[58]

4.2 Cognitive Mapping

In contrast to problem-solving's focus on plausible search, cognitive mapping approaches (Bonham and Shapiro, 1975; Axelrod, 1976) attempt to capture some aspects of affect in causal algebras based on "cognitive consistency." In the earlier formulations, "cognitive consistency" refers to the propagation of positive and negative affect according to the conceptual associations. Cognitive mapping represents one branch in a family tree descending from "cognitive balance" theory (Heider, 1946) and the "cognitive consistency" models of belief developed by Abelson and his associates (Abelson & Rosenberg, 1958; Abelson, 1959; Abelson & Carroll, 1965; Abelson, 1968; Abelson & Reich, 1969). Cognitive mapping attempts to construct essentially symbolic model of the affective associations of a decisionmaker. After mapping the relevant affective structure, studies in this genre attempt to postdict (or predict) decisions for specific decision makers according to affect propagation and associations.

The need for interviews with the decisionmakers to construct affect maps circumscribes the applicability of the approach to accessible leaders. One drawback is that affect maps were static. Mefford (1979a, 1979b) proposes introducing a transformational component into cognitive mapping in order to cope with structural change over time. Another major shortcoming is that the affective relations were essentially undifferentiated. Working within the "conceptual dependency" tradition[59] (Schank, 1972; Schank & Abelson, 1977), Lehnert (1981, 1982; Lehnert & Loiselle, 1985) develops a richer theory based on "plot units,"

[58]Newell makes this argument in his 1987 William James Lectures at Harvard University. For Newell, a unified theory of cognition must propose a comprehensive mechanism to account for human cognitive abilities. It should have a computational implementation that conforms to and explains the large quantities of experimental constraints currently known to psychology. Although there are indeed important elements of SOAR's architecture that a unified theory of cognition must incorporate, there are also important facets which are untenable. Some implausibilities include:

- The theory of meaning is reductionist rather than constructivist and the representation impoverished;
- If-then rules have low cognitive plausibility because they are ungrounded in a historical representation;
- The learning modalities do not include metaphor or even analogy.

[59]Discussed in section 5.2.

a collection of "primitive" units of affect that can be combined in larger "affect molecules." These larger affect chains are considered to represent the emotional skeleton underlying narrative stories. The approach has been used to study the affective core of politically significant texts, such as Christ's "Sermon on the Mount" (Alker, Lehnert & Schneider, 1985). This political application clearly illustrated problems of intercoder reliability that demand attention. While the "plot units" enrich the approach by differentiating affective relations and allowing them to combine into interesting composites, no (cross-)cultural validity has been established for the primitives or the patterns in which they can combine. As the research on affect becomes more sophisticated, it gradually shades from the simple representations of cognitive-mapping into the richer memory structures and inferential processes of cognitive modeling (Pfeifer, 1982; Dyer, 1983; Mueller, 1987)[60]

4.3 Knowledge-Based Systems

4.3.1 Foreign Policy Decision-Making

Thorson and Sylvan (1982; Anderson & Thorson, 1982) describe an interactive cognitive model[61] that supports counterfactual simulations of President Kennedy's decision process during the Cuban Missile Crisis. Their aim is to model the role of incremental advice and new information in the selection of options and assessment of probable Soviet responses. A CD-like representation (see section 5.2) is used to represent Kennedy's knowledge at any point during the crisis and a set of 63 production rules captures Kennedy's beliefs about relevant causal relationships in international affairs. Sentence-like inputs are interpreted according to the current context and the active rules. As new information is received by the system, plausibilities are assigned to beliefs and options. Responding to inputs consisting of either Soviet actions or some new advice, the system searches for an acceptable option that will move the current context toward an acceptable state of the world. Option selection is sensitive to the order in which alternative are considered in order to simulate "satisficing," but this is considered a weakness (Anderson & Thorson, 1982). After initially reproducing early stages of the historical crisis, several counterfactual runs tested the outcome sensitivity to various parameters, including the risk adversity of the Soviets and changes in Soviet responses. Thorson (1984) emphasizes intentional inferencing in a expressly CD representation used by a successor model of decision-making in Cuban Missile Crisis.

Anderson and Thorson (1982) report on an interactive hybrid simulation of Saudi Arabian decision-making, treating the government as a unitary actor. Difference equations simulate Saudi oil production, crop production and population dynamics while the government is modeled using rule-based programming. The government decision module implements policy by manipulating equation parameters in order to achieve production goals. The cycle periodicity is one month for the dynamic equations. Sentence-like inputs provide information which the system interprets and acts on to achieve goals of the simulated

[60]Of course, most cognitive modeling ignores emotions even though figures such as Simon (1985) and Abelson have repeatedly stressed its importance, including its role in structuring memory. To date, Mueller's (1987) Ph.D. thesis is the most sophisticated cognitive model of emotions within a conceptual dependency framework. For those interested in pursuing emotion modeling, Mueller (1987: 57-69) provides a useful review of previous work.

[61]The program consists of 800 statements written in SPITBOL.

government. Syntactic analysis reduces the input to a "basic meaning." This is then interpreted context-sensitively, according to rules about how the environment works, for use by a decision module. The vocabulary consists of 23 nouns, 19 verbs, and 17 actors.[62] Based on user input describing the international system, the system controls production parameters and provides responses to queries. Thus, by providing different input or varying its order,[63] different scenarios are played out.

Lenat and his coworkers (Lenat, Clarkson, & Kiremidjian, 1983) discuss an expert system designed to aid a human intelligence analyst in performing the "indications and warnings" task, essentially tracking reports about military movements and preparations in order to predict military attacks or operations.[64] The task involves maintaining a domain model representing the current situation and potential scenarios. This system maintains two parallel situation models. Rules with strong matching criteria are used to infer the high confidence model from the report stream. Rules with weak matching criteria are used to maintain a low confidence model. The true situation is believed to rest between the two models. Inverse rule pairs provide forward-chaining rules to infer consequences from events and backward-chaining rules to infer antecedents from events, and thereby, help fill in information gaps.[65] The representation uses a frame system that represents objects and events in terms of attribute-value pairs. Temporal duration is an important parameter associated with each event frame. The systems uses knowledge of characteristic durations to determine the military processes active at any time. Using characteristic temporal durations and knowledge of active processes, the system can estimate lower and upper bounds for times required to achieve specific military states, which accomplishes the "indications and warnings" task. The system provides an explanation facility that allows an analyst to inspect the inferential processes leading to specific conclusions. An analyst can use this facility to identify the key break points and critical pieces of information. A window-oriented user interface facilitates system use. In 1983, the authors were expanding the number of reports types handled and incorporating a scenario generation capability based on open-ended exploration methods or "learning by discovery" (Lenat, 1983).

In more recent published work, Lenat and Clarkson (1986) provide a high-level overview of the application of AI to C^3I, which they understand as:

- Managing SDI;
- Enriching and making readily available the intelligence information to policy makers;
- Refining command and control of strategic forces C^2;
- Improving tactical C^2.

[62] There are a total of 215 knowledge elements in the simulation; only a small fraction are active at any one time. The decision making component contains 109 rules and is implemented by 2619 lines of SPITBOL.

[63] Like the Cuban Missile simulation, the system is sensitive to input order.

[64] The system is written in Interlisp-D and runs on Xerox D (1100 series) Lisp Machines. The prototype system contains 60 rules and 170 frames representing objects and processes. Rules are presented to a user in a stylized English form. On a Dolphin (1100 series), a lower end machine, the system can process reports at a rate of 30 per minute.

[65] The authors note that an improved design would incorporate bidirectional causal relations instead of inverse rule pairs.

Their view of the directions for AI in C^3I is grounded in specific technologies with which they are familiar, particularly machine learning (Lenat, 1983) and encyclopedic knowledge bases (Lenat, Prakash, & Shepherd, 1986). They anticipate introduction of major AI components into C^3I, including numerous expert subsystems and a machine learning component, operating with 10^4 to 10^5 rules. The note a major difference between strategic and tactical C^2: Tactical strategy is more computationally tractable for AI system to explore. While a learning system may attempt to discover or learn new tactics, flaws in tactics, and synergistic relationships between tactics, it cannot expect anywhere near as much success at the strategic level because of the higher complexity found there. In the intelligence field, they see AI systems:

- fusing data from multiple sources,
- highlighting indicators for various important events from a morass of information,
- verifying hypotheses based on confirmation or disconfirmation by collected information,
- generating hypotheses on the basis of stored information.

Introducing AI systems into the analysis apparatus is considered to help reduce "wishful thinking," where only information supporting preconceived hypotheses is highlighted, improving institutional memory by carrying over expertise across cycles in human analysts, and freeing analysts to cover more areas in greater depth. Lenat and Clarkson contend that the most difficult aspects of C^3I will require *machine art*, AI systems which reason about mythical archetypal and historic knowledge by drawing on not just expert-specialist knowledge but also common-sense knowledge. The premise is that mythical knowledge together with heuristics for creating dramas provide scripts of strategic eventualities suited for specific times, places, and conditions. Assuming that these mythical scripts are "powerfully imagined" – that is, unquestioned cultural predisposition guiding individual behavior – they constitute a constraint on strategic thinking that an AI system can exploit to guide its contemplation of the domain.

Sylvan and Majeski (Sylvan & Majeski, 1983; Majeski & Sylvan, 1985; Majeski, 1985) discuss their ROSTOW model.[66] Theoretically informed by "bureaucratic politics" (Allison, 1971; Steinbruner, 1974), this case-specific simulation attempts to reproduce the decision-making behavior of a key presidential advisor within an intersubjective organizational context. The specific historical period treated is the Kennedy administration's decision in late 1961 to increase the number of U.S. advisors in Vietnam. The advisor is Walter Rostow who was then a deputy to McGeorge Bundy, the special assistant for National Security affairs. The objective of the simulation is to reproduce Rostow's recommendations during about a dozen stages in the decision process. A set of rules simulating the decision process act on a propositional representation of the decision task and the bureaucratic environment. Organizational constraints, essentially reflecting Kennedy's views, are an important set of propositions that constrain the direction and framing of recommendations. One category of rules represents general norms for bureaucratic behavior and another group expresses heuristics followed by the individual decisionmaker for interpreting situations. Giers (1986) concludes that although ROSTOW aims at generality by clear separation of theory and data, the effort slides into case specificity as theoretically motivated bureaucratic norms

[66]The system is implemented in MICRO-PROLOG, a subset of PROLOG.

become intertwined with particular scenarios. Important weaknesses include the inability to represent the beliefs of different actors and the implementation in one of the least powerful computational environments. But, this data intensive exercise is an important research direction because it represents an initial attempt to build a model that follows the paper trails which condition bureaucratic cognition and constitute the matrix of ideational niches within bureaucracy. It therefore, recognizes bureaucratic decision-making as a process of organizational communication with a documentary trace. These researchers are focusing now on identifying and formalizing role of "bureaucratic culture" in framing situational interpretations in ways that making war recommendations conceivable or not (Majeski, 1987). This move is an attempt to get beyond case-dependent models to a generalizable yet formal theory of decision-making explaining war recommendations.

The RAND model is one of the major implemented AI systems for foreign policy decision-making. It is intended to provide a computational laboratory for testing different theories about American and Soviet decision-making in strategic crises. This is an improvement over earlier methods. As long as the military strategists heed the implementor's warning that the system is only a tool for studying the domain and not a predictive tool on which real-world decisions should hinge, the system makes a positive contribution to strategic thinking. While the RAND expert system for simulating strategic decision-making uses some symbolic decision criteria expressed in the form of heuristic rules, it mostly relies on models of rationality grounded in a utilitarian outlook — an ontologically anomalous juxtaposition with symbolic modeling. The decision-making assumes an ideal strategic decisionmaker. The model of decision[67] incorporates elements of:

- **Decision-analytic models** which seek an optimal solution by maximizing a preference function (Raiffa, 1970; Keeny and Raiffa, 1976);
- **Game-theoretic approaches** which combine preference functions with game-theoretic constructs (Brams, 1985);
- **Bureaucratic politics approaches** which stress the influence on decision of competing factions, interests, and divisions within the bureaucracy, as well as their external supporters (Allison, 1971; Steinbruner, 1974);
- **Organizational cybernetics** which emphasizes feedback, control, and local adaptation in organization decision-making (Steinbruner, 1974; Janis & Mann, 1977; Simon, 1969, 1982);
- **Rule-based heuristic models** which deploy heuristic rules based on situational context (Carbonell, 1978; Schwabe & Jamison, 1982);
- **Information processing models** which focus on receiving and processing information within the constraints of "bounded rationality" (Janis & Mann, 1977; Simon, 1969, 1982; Kahneman, Slovic, & Tversky, 1982).

This decision model proceeds from a utilitarian principles that presuppose a preference function to select actions to maximize some preferred values. Much like the scientific style of

[67]For details of their model, see (Davis, 1987 10-17; Davis *et al.*, 1986: 10-30).

neo-classical economics, the unrealistic assumptions are gradually relaxed to achieve successively closer approximations to the real world. Notions such as "bounded rationality" (Simon, 1969) relax assumptions about decision-making based on perfect information and unlimited computational resources. Incorporation of additional situational constraints on the decision process further relax utilitarian assumptions. These include introduction of competing interests with differential capabilities as well as the structure of command, control, and information flow, bureaucratic politics and organizational cybernetics.

Although the system incorporates a full complement of standard decision theories, there remains no adequate account for the role history in a decisionmaker's interpretation of a situation and selection of appropriate actions. Naturally, learning modalities such as concept induction, analogy, and metaphor cannot be considered without first accepting history as an important component in decision. The RAND researchers note the importance of learning (Davis, *et al.*, 1986: 92-97) but model it as changing assumptions — an unrealistic construal for learning by metaphor which gives new significances to old terms or learning by analogy which employs a familiar functional relationships in new contexts or concept induction which creates classificational categories. Aligning themselves with Schank,[68] they claim that no general theories of learning exist. This is almost true in regard to Yale school except for some work on analogy based on causal relations (Schank, 1982; Burstein, 1986) and some work on concept induction (Lebowitz, 1986) but certainly not true in regard to other approaches that do not rely on semantic universalism, *e.g.*, Winston (1984) or Minsky (1987). The additional learning strategies possible in lexicalist representations arise because of the presence of an informal level of representation (Winograd & Flores, 1986; Mallery, Hurwitz, Duffy, 1987).

Job and Johnson (1986) present a preliminary report on the development of UNCLESAM, a rule-based simulation of U.S. decision-making regarding the Dominican Republic between 1961 and 1965. They aim to develop a general decision-making model that reproduces the chain of assessments and decisions observed in U.S. policy on the basis of incremental information about events in the Carribbean. But they report on their first case-specific effort. Data for the simulation were acquired by having a specialist on Central America extract 250-300 reports between 1959 and 1965 from the *New York Times.* These reports were then translated into event statements according to a "political action language." Their system converts event statements into "quasi-English" statements which are processed by the simulated decisionmaker. Once events are input to the system, they are assessed according to their context-specific policy relevance. Assessment may change values for attributes associated with actors and relationships. If policy goals are affected, a "report" is issued that states the significant changes. Based on the policy posture toward the country, modeled as a set of rules, an internal or external action may be selected and executed. While the system can record fixed attributes for a situation, its representation system is not yet sophisticated enough to recognize analogies or explore counterfactual consequences of alternate actions. The results they report are based on 38 event inputs that were processed by 10 "algorithms" for assessment of political stability, and 3 policy postures, apparently comprised by 10 rules. The authors note that their model fails to incorporate delays for acting on knowledge and that such delays played an important role in the historical case.

Donald Sylvan (1987a, 1987b) presents a model of congressional decision-

[68]This is correct; they are grounded in a semantic universalism because token significance, *ergo* interpretation, is invariant across belief systems. See section 6.1.2.

making regarding expenditures for research and development in the energy sector.[69] The model represents at an individual cognitive level one "modal" and two other representative members of the U.S. Congress as each address a hypothetical energy research and development funding bill. A central assumption of the theory expressed in the model is that problem solving by congressmen in the energy arena is directed by short-term considerations such as reelection, not alienating constituents, and their general public image rather than long-term issues such as the eventual cost of energy to consumers and national security. These motivations are reflected in the simulated congressmen. A binary decision logic, reminiscent of cognitive consistency, is to determine how the simulated congressmen vote. By simulating the intentional inferencing of congressmen for a specific issue area (different logics may apply for other issue areas), Sylvan's model predicts lower expenditures than would be predicted by optimization models based on economic rationality. From his empirical research, Sylvan concludes that his model provides a superior account of congressional decision-making than alternatives based on economic rationality precisely because he is able to capture some of the intentional aspects of decision. Sylvan and his colleagues are now turning to modeling Japanese foreign policy decision-making in the energy sector (Bobrow, Sylvan, Ripley, 1986; Sylvan, Ripely, Bobrow, 1986).

4.3.2 General Foreign Policy Decision-Making

In contrast to the RAND system's grounding in conventional decision theory, Mefford's (1986b) system approaches the foreign-policy decision process from the perspective precedent-based reasoning.[70] Mefford's idea is to provide decisionmakers, or foreign policy researchers, with a database of precedents and allow them to build hypergames for particular decision situations. The hypergame approach yields, in effect, a decision tree of actions available for each actor as perceived from each actor's perspective. A key assumption underlying the approach is that contingency plans can only be generated by mechanisms based in analogy, or precedent logics (see section 4.4). Since this class of reasoning can produce spurious results, an evidential reasoning module weeds out spurious precedents. The system's evidential module reasoning (Mefford, 1986c) uses "models of evidence,"[71] hierarchical Bayesian analysis, and hierarchical Dempster-Schafer to justify the plausibility of specific scenarios. Games of partial information can be generated by varying the partitioning of the database.

4.3.3 Legal Decision-Making

Despite some early work in international law and AI (Alker, 1971), no subsequent work has developed knowledge-based systems for international law and international legal decision-making. However, there is considerable interest in developing legal reasoning systems for constrained areas of domestic law.[72] Most AI applications law have treated well-established

[69] The model is simulated by a procedural formulation in LISP. Sylvan (1987b) includes a copy of the LISP code implementing the model.

[70] This expert system shell runs in PROLOG in an Interlisp environment on Xerox LISP machines.

[71] Mefford attributes this to Martin, 1986 but provides no citation.

[72] Conferences on AI and Law began to appear in the mid-1980s. The First International Conference on Artificial Intelligence and Law took place at Northeastern University in May, 1987. It was sponsored by Northeastern's Center for Law and Computer Science.

area such as assault and battery (Meldman, 1975, 1977), corporate taxation (McCarthy, 1977; McCarthy, Sridharan & Sangster, 1979), product liability (Waterman & Peterson, 1981), contract law (Gardner, 1984). In contrast, Grunbaum (1986) reports on a system that predicts Supreme Court decision-making for discrimination cases by modeling steps of the legal argument.[73] Grunbaum argues that legal reasoning is suited for expert system simulation because decisions are essentially binary in nature and because most, but not all, reasoning is explicated for each individual case. The system uses a decision tree based on 37 rules and 41 conditions to determine the constitutionality of a case. A user provides a characterization of the case by providing answers to a series of yes-no English questions, and the system reports a decision or its inability to do so for reasons of overbreadth or vagueness. The knowledge base is divided into categories and rules. Careful attention is paid to rule precedence, the order in which rules are applied. The system can identify critical rules that swing a decision different ways. It also weighs evidence according to a certainty system based on fuzzy logic and an ability to detect contradictions. Based on theoretical argument and simulation runs, Grunbaum concludes that legal reasoning, specifically supreme court decision-making, can be modeled by predicate calculus. He feels that the computational approach helps identify the core logical steps leading to decisions.

4.3.4 Decision-Making in Political-Economic Development

Phillips and Ensign (1982) model governmental decision-making in the context of political-economic development in less developed countries. They develop sets of decision rules to simulate several "ideal-type" decision-makers. They characterize the decision-makers with several combinations of three economic strategies (externally reliant growth, autarkic or "self-reliant" growth, and neo-classical policies) and four political development strategies (political development based on a modernizing elite, social transformation from rural agriculture to urbanized industry, social integration, and reassertion of traditional religious or cultural values). Their method employs a rule-based system originally designed for symbolic simulations of causality in physical mechanisms (Reiger, 1975, 1976). In addition to relations of instantaneous cause, continuous cause, and causal enablement, Reiger's simulator distinguishes structural relationships between states of equivalence, inhibition, and enablement. Inhibition and enablement relations between states are used when the intervening causal relations are unknown but some empirical relationships are observed. After expressing their decision-makers and specifying initial conditions for the hypothetical countries according to Reiger's descriptive system, Phillips and Ensign tested the responses of their idealized decision-makers to stimuli such as poor agriculture performance as well as a combination of growth and political instability. They found that modernizing strategies faced the greatest political instability while traditionalist strategies faced little instability when confronted by economic or political disruptions.

Ensign (1985a) extends earlier research with Phillips on decision-making for development. She revises the decision rules in terms of four economic strategies (primary specialization, balanced growth, industrial specialization, and balanced autarky) and three political strategies (political development, socio-economic transformation, and social integration). After building twelve theoretical strategies, she constructs country specific models for Sri Lanka, Brazil, and Korea. In the Sri Lankan case, decision-models for the party in power and the opposition were tested for reactions to increased foreign direct investment

[73] The system is implemented in a commercially available version of PROLOG.

and domestic political instability. She found that foreign direct investment produced more political strife than other sources of political instability.

Ensign (1985b), again employing the Reiger causal modeling techniques, modeled[74] the decision-making of international bankers to determine credit worthiness of underdeveloped countries. The simulation was based on data from the early 1970s for six major U.S. banks. Based on interviews with top executives, Ensign identified a "conservative" lending strategy followed by Chase Manhattan, Morgan Guaranty, and Manufacturers Hanover and an "aggressive" lending strategy followed by Citibank, Bankers Trust, and Bank of America. The "conservative" strategy was more concerned with political risk than the "aggressive" strategy which was primarily concerned with economic risk. But, the two strategies converged in 1977 to form a merged strategy concerned with both political and economic risk. After constructing decision models for assessing credit worthiness during the 1972 through 1974 period, Ensign ran simulations replicating the domestic and international economic environment facing the bankers. She found that, for the entire period, the merged strategy predicted lending behavior better than either the conservative or aggressive strategies.

More recently, Ensign (1986) proposes developing educational rule-based systems that allow experimentation with different decision criteria within a simulated political economic environment. The substantive domains to be modeled are international trade and finance in their relation to economic development and the international political economy of poverty and hunger. By capturing domain expertise about the causal structure of problem areas in a simulation, students may examine the structure and trace justifications back to the literature or they may experiment by varying the problem-solving strategies of actors as a form of counter-factual analysis. The inspectibility of the simulation allows students to fill specific deficits in their knowledge. The ability to modify strategies of different actors are deemed an important pedagogical tool because it encourages counter-factual thinking.

4.3.5 Functional and Teleological Simulations

Mefford (1985) uses PROLOG to model the universe of debate and decision about the resort to force of arms. This work uses accounts from Thucydides' history of the Peloponesian wars. The paper argues that logic programming provides a means for systematic and reproducible models of argument and debate. Mefford (1986d) recasts Morton Kaplan's "balance of power" as a rule-based system, and shows the logical gaps in Kaplan's verbal theory that are left for human readers to fill. The rigor of his PROLOG implementation allowed Mefford to identify and fill the gaps in Kaplan's theory. Mefford thus illustrates how computational simulations expose the incomplete coverage found even in verbal theories that employ systems-theoretic methodologies to gain completeness.

Focusing on the internal dynamics of societies within Wallerstein's (1974, 1980) world system perspective, Banerjee (1986a) models the reproduction of social structures using Piaget as theoretical orientation and PROLOG as a computational tool. Building from Piagetian schemata (1977), Banerjee devises social action schemata. Social action schemata represent the interpretations and preferred actions of social actors in situations. In a simulation, each actor uses its schemata to select actions on the basis of prediction of their consequences and the likely responses by other actors. Banerjee applies the method in

[74] This research was originally reported in Ensign (1982).

simulations of Skocpol's (1979) analysis of China's sociopolitical structure in the 1930s, and subsequently, O'Donnell's (1973) analysis of bureaucratic-authoritarianism in Latin America during the 1960s. The conclusions of this work are that Piaget's action schemata provide a building block for Geertz's (1973) notion of culture and that AI, PROLOG in this case, provides tools for proof by simulation. Substantively, Banerjee finds that the simulation supports Skocpol's and O'Donnell's analysis of social reproduction. Even if these claims are overly strong, they suggest an interesting line of research.

In a subsequent model intended to provide a structural-reproductive alternative to rational choice approaches from economics, Banerjee (1986b) simulates the selection, retention, and abandonment of techniques by firms seeking a competitive advantage. More recently, Banerjee (1986c) applied his method to simulating US and Soviet foreign policy decision-making in the early Cold War period. Approximating a hermeneutic approach, Banerjee models the USSR and the US separately. This allows each country to have different beliefs and interpretations of "objective" reality, represented by a separate model of exogenous events and actions by the actors. A means-ends type of reasoning is performed by each unitary actor in order to select actions achieving to actor goals. Differential perceptions (Jervis, 1976) arise as different beliefs about other actors and about the causal structure of the world are applied to interpret the significance of actions by others and to foresee the consequences of actions. Although simple, the simulation demonstrates, according to Banerjee, that the pattern of failures and mistakes in the history of the early Cold War is (re-)produced by the beliefs of the superpowers.

4.4 Precedent Logics for Decision-Making

International relations scholars (Alker, 1970, 1971; Alker & Greenberg, 1971; Alker & Christensen, 1972) discovered precedent logics in the late 1960s. Alker and his students at M.I.T. invented precedent logics as ways to model decision-making, learning, and adaptation by international organizations. Their idea was that organizations and individuals solve problems by selecting past situations most closely resembling present situations, and applying the prescriptions of the past to the present. They also noted that precedent logics provide an account of learning in so far as they explain the salience of precedents over time and their recombination into composites. A related information retrieval system, CASCON, was independently developed at M.I.T. by Bloomfield and Beattie (1971). CASCON retrieved retrospectively coded crisis descriptions that were "similar" to a present crisis. Although CASCON's search algorithms resembled Alker's, the systems had fundamentally different orientations. CASCON arose from crisis gaming (Bloomfield & Gearin, 1969) and was conceived as a decision-support tool to provide policy-makers or analysts with relevant historical precedents. In contrast, Alker's precedential reasoning models were intended to actually simulate problem-solving.

An important side effect of this "precedent logic" approach is an implementation of Ludwig Wittgenstein's (1953) concept of "seeing-as." The seeing, or perception, of the situation is prestructured as it is found similar to previous situations and as past precedents are called forth to interpret it. This prestructuring is a propensity to understand in a particular way rather than some other plausible way. It sets the stage, almost preconsciously, for interpretation. In the international relations literature, May (1973), Jervis, (1976), George (1979), Gilovich (1981), and Mefford (1986a) discuss the prestructuring of the present by focal elements of the past. An important property of precedent logics for the individual, the organization, and the modeler is that they provide criteria for ignoring irrelevant information and focusing on the core structure of decisions, the precedentially grounded abstractions spanning multiple situations. While they provide foci for individuals and organizations, precedents suggest the inferential core that strategic simplification must retain in plausible models. Thus, precedent logics promise a felicitous account of the central aspects of social decision-making in both theory and simulation.

Under the influence of Newell and Simon's ideas of generalized problem solvers, Alker and his students originally brought forth the idea of precedent logics to model organizational problem-solving. Interestingly, SOAR, the GPS of the 1980s, incorporate an important precedential component in its "chunking" mechanism (see section 4.1). Some AI researchers take as an article of faith the premise that human reasoning is grounded in precedents or analogies (Winston, 1984; Minsky, 1987). The underlying reason for this convergence is that most AI researchers have recognized the computational difficulty of solving any non-trivial problem, and especially, solving it repeatedly. The intractability of GPS (see section 4.1) and non-linear conjunctive planning (see section 6.1.4) serves as an illustration. Recall, however, that organization theory draws significantly from GPS, as can be seen in the arguments of Simon and others about "bounded rationality," "satisficing," and "muddling through" (Simon, 1957, 1969, 1982; Lindbloom, 1959, 1965). Thus, Bloomfield and Alker with their colleagues would seem to be rewriting organization theory according to a computationally more effective theory based on precedents.

In summary, precedent-based explanations of decision-making are an important theoretical development because:

- They afford a history to role in decision;
- They provide an account of learning;
- They explain decision prestructuring;
- They reduce the information required for a model;
- They provide a computationally tractable account of problem solving.

Although important cumulative progress has been made, no definitive theory of the role of precedents in decision-making or individual cognition exists as yet. Further advances depend on better AI models of natural-language processing, semantic representation, and precedent-based reasoning as well as a thorough theoretical formulation in light of operational AI systems.

The CASCON project (Bloomfield & Beattie, 1971; Bloomfield, 1986) is a two decade effort to develop a computer system for retrieval of international conflict or crisis situations on the basis of identifying features derived from crisis histories. CASCON aims to provide an "institutional" memory and to aid imagination of a conflict analyst. Since foreign-policy decision-makers could retrieve cases similar to a current crisis, this "institutional memory" helps the analyst see the crisis in the light of historical precedents which are similar according to some analytical dimensions although perhaps not all. The underlying assumption is that lessons of the past provide a guide for contemporary decisions. By comparing historical crises with a current crisis, a conflict analyst can suggest possible futures, identify information gaps, and propose collecting missing facts to fill in an emerging picture. The CASCON III database contains over 50 cases coded for more than 500 variables provided by experts with different perspectives. The retrieval mechanism uses statistical matching on the variables to retrieve narrative preces of the essential facts for cases. The system has found application in government, crisis gaming simulations (Bloomfield & Leiss, 1969), and international relations research (Choucri, 1974). It represents one of the earlier computational applications of case-based or precedential reasoning in international relations. In this work, the "reasoning" is done primarily by a human with computer support for retrieving cases. Recent work on CASCON aims to deliver a personal computer version of the system for analytic uses in government and academia.

Alker and Christensen (1972) describe research with William Greenberg between 1969 and 1971 that led to the development of the PRECEDENT program.[75] Their aim is to formulate theories about UN peacekeeping efforts in the postwar period. The study recounts the authors' transition from an initial methodological orientation based in "causal modeling" grounded in multivariate statistics to a new artificial intelligence approach. Their new artificial intelligence approach is theoretically informed by Herbert Simon's symbolic problem-solving and practically grounded in a precedent-based understanding of UN decision-making. After reviewing the limitations of advanced statistical methods, they explain their notion of precedent and detail the operation of their program. By precedent, they mean "rules for accumulating experience in an action predisposing way" (Alker & Christensen, 1972: 194).[76] The role of precedents in the model gradually evolves from a partnership with multivariate equations incorporating precedent variables to an essentially symbolic model using feature matching for precedent retrieval. The key research questions about UN decision-making become how precedents are determined and how precedents are changed by unsuccessful applications. They develop the dynamic notion of precedent evolution in which an initial stock of precedents (the 1945 UN Charter) is modified incrementally by the addition of new precedents from successful peacekeeping cases and by the extinction of existing precedents as they fail in new situations. This "one-shot" learning and forgetting, or selection and retention, is an *evolutionary decision model*. Alker and Christensen find that the norms contained in the UN Charter and expectations engendered by historical case similarity are important for explaining UN decision-making in peacekeeping efforts during the postwar period. They also suggest that, to the extent that precedent-based models capture evolution in individual precedents and the underlying norms, they provide a means to model structural transformations in decision-making regimes.

Alker and Greenberg (1971, 1976) extend the Alker-Christensen UN peacekeeping model. Their aim is to ascertain the effect of the Cold War on successful conflict resolution by simulating learning in UN peacekeeping efforts between 1945 and 1965. An evolutionary precedent logic uses an "operational UN Charter," that is updated by learning, to select both conflict cases amenable to UN peacekeeping efforts and courses of action to promote peace. The operational charter begins as the initial 1945 charter but is modified according to the success or failure of conflict-resolution precedents as they are applied to cases.

[75] The program is written in PL-1.

[76] Their notion of precedent emerged primarily from interactions with Lincoln Bloomfield and his CASCON co-workers at M.I.T. as well as Philip Stone other scholars working on "concept attainment" at Harvard. The use of precedents by Alker and Christensen, however, drops the human analyst out of the picture and frames the problem as modeling organizational decision-making on the basis of precedents.

The success of conflict-resolution actions in cases is determined by probabilistic equations derived from historical data. Unsuccessful precedents are "forgotten" from the operational charter according to an exponential decay rate of 1.33, which means they are dropped from the charter typically within several failures. "New" precedents are added to the operational charter by cross-classificational mappings, mapping a precedent from one conflict category to a case in another conflict category. The accumulation of new precedents (the rate of learning) in the operational charter, therefore, depends on the amount of generalization performed by the matching process that retrieves precedents.[77]

The matcher matches feature-vectors describing conflict cases. It selects candidate precedents on the basis of a minimum number of importance-ordered, mandatory features (either 2, 3, or 4) shared between the case and the precedent. Candidates are then ordered according to their past success. Only when ties remain does the number of additional features effect closeness of match.

The pattern of learning and forgetting in this evolutionary precedent logic causes the operational charter (the current stock of precedents) to generalize the application of successful precedents to conflict cases in which they continue to succeed and to drop precedents from application to cases in which they fail. Thus, the operational charter evolves toward success cases and away from failure cases. The "1965 operational charter" was derived in an initial simulation of UN conflict resolution efforts between 1945 and 1965. Observing that conflicts involving Cold War issues accounted for a large number of UN peacekeeping failures because the superpowers vetoed UN efforts, Alker and Greenberg ran another simulation of the 1945 to 1965 period to learn a "maximalist" operational charter by assuming security council agreement on ceasefire plans. In subsequent runs, Alker and Greenberg reran the past, performed counterfactual simulations for the postwar period using the 1965 operational charter and the "maximalist" operational charter. They also constructed a future conflict scenario on the basis of the conflict history for the postwar period. For the counterfactual simulation of the past, they found that the 1965 operational charter performed better than the "minimalist" charter, the record, and that the "maximalist" performed better than the 1965 operational charter. The same relationship held for the 1965 and the "maximalist" charters in their future projection. Thus, a conclusion is that Cold War disagreements hampered the effectiveness of the UN's conflict resolution but that the UN became better at conflict resolution on the basis of adaptive learning.

Bennett and Alker (1977) report on their extensive multi-level regional simulation of the War of the Pacific, a 19th century regional conflict in South America.[78] The simulation cumulatively incorporates many methodological ideas from earlier precedential and dynamic models. Although they recognize and stress recombination and evolution of precedents as a model of learning, the major aim is to model goal-seeking behavior of countries with a regional subsystem. Dynamic equations represent the international economic transactions at regional and global levels of aggregation, the international stratification of actors, and relationships of dependence and interdependence. Thus, the dynamic equations provide both the environment and the stimuli that actors respond to by formulating goals ac-

[77]Interestingly, Alker and Greenberg formulate precedent learning as a combination of inductive generalization to make cases match and deduction to project the peacekeeping action from the historical case to the present situation or some future or counterfactual past.

[78]Implemented in PL-1, the simulation contains over 100 procedures, over half of which implement the actor decision components. Circular lists provide a simple but CD-like representation. Bennett (1978) provide details of the implementation.

cording to gaps between their aspirations and their achievements. Actors may interact along multiple dimensions with each other and pursue rich multi-game strategies. Precedent-based decision-making represents problem-solving of actors as they attempt to attain their goals within this regional environment. The sources for precedents are not just the actors' own historical experiences but also the examples of others (imitation). Drawing on organization theory (Cyert & March, 1963), Bennet and Alker reject the unitary actor assumption; they model actors' internal structure with a central decision-maker supervising lower-level subordinates. The subordinates carry out specific decision tasks under the constraints of their assigned resources. Each decision unit context-sensitively applies an evolutionary and recombinant precedent logic to strategies drawn from international relations theory and the historical case. To investigate different precedent retrieval strategies (different generalization heuristics), Bennett and Alker varied the precedent search strategies to identify the best strategies for each actor, assuming a specific international system. They find that outcomes were influenced significantly more by the set of precedents from which actors seeks to realize self-proposed goals than by whether the actor seeks to equalize stratificational ranks, avoid security vulnerability, or realize successive increases in status. Although precedent coding strongly influences final outcomes – and domain expertise is essential for accurate coding – the self-structuring of the simulation (actors proposing goals and solving them) produces system-level structural transformations as a function of slight differences in initial memory.

As part of a consolidation and proposal for further research, Alker, Bennett, and Mefford (1980) review and assess key ideas from efforts to apply precedent logics developed since Bloomfield's early CASCON work. They characterize the core of international relations as collective insecurity dilemmas, in which multiple countries or actors seek security through conflicting strategies. They further assume that these dilemmas involve contractions between short-term individual goals and long-term collective goals. Since the actors are interdependent in the longer term, they must therefore employ "reflective logics" to reason about the beliefs and intentions of other actors in order to empathize, negotiate, and achieve the collective long-term goals despite the obstacles of the short-term term. Transcending short-term goals is conflict resolution. According to their experience, sequential prisoner's dilemma (SPD) games largely parallel both the experimental literature on interpersonal conflict dilemmas and real-world international conflicts. In order to develop precedential theory of breakpoints leading to thematic shifts in SPD, such as cooperation and defection lock-ins, they wish to turn more systematically to artificial intelligence techniques. They consider models of history-based learning in collective insecurity dilemmas necessary for cumulative scientific study of conflict behavior. The task they set for AI is to apply precedent logics developed earlier to representing narrative, textual accounts of this kind of reflective understanding. By explicit representation of the cognitive processing, with particular attention to moral (Kohlberg, 1969), ego (Loevinger, 1976), and cognitive (Piaget, 1972, 1977) development, they hope to generate realistic psychological models for experiment data from SPD game play. They also discuss ideas for improving their earlier organization-level models and explain the need for similar AI methods there. The general conclusion is that better models of memory, retrieval and problem-solving are required for further progress in precedent-logics models of individual and collective action.

Tanaka (1981, 1984) reports on CHINA_WATCHER[79], a computer model of Chinese foreign-policy decision-making in the post World War II period that brings together elements from cognitive consistency, cognitive mapping, and precedent logics. Tanaka models

[79]CHINA_WATCHER was implemented procedurally in FORTRAN. The data set of conflict cases was compiled by Consolidated Analysis Centers, Inc. (CACI, 1979) under the sponsorship of DARPA.

decision using an Alker-Christensen-Greenberg precedent logic (Alker & Christensen, 1972; Alker & Greenberg, 1971, 1976). When presented with a new event, CHINA_WATCHER interpretes it based on "cognitive consistency" (Abelson & Rosenberg, 1958; Alker, 1979a), updates a world map accordingly, and selects an appropriate response based on a precedent logic. Arguing that Chinese leaders are inclined to divide the world into friends and enemies, Tanaka considers the four valued Abelson-Rosenberg symbolic psychologic (friend, enemy, ambivalent, indifferent) a good approximation. Current actions of countries are interpreted as friendly, hostile, ambivalent or indifferent. Six transformation functions update the world map, mapping classifications for countries from old ones to new ones on the basis of their current actions. These transformation procedures apply only to countries whose actions directly effect China. CHINA_WATCHER implements "cognitive balance" (Heider, 1946) or equilibrating propagation of affect because other countries are classified indirectly according to their relationship to these "front line" countries. By comparing current conflict cases to unfinished cases, context routines attempt to situate present actions as phases of ongoing episodes. But, the simulation's major component is a featured-based similarity matcher that retrieves historical precedents used to determine Chinese responses to current events. After reviewing earlier matching schemes, Tanaka proposes matching criteria that require essential important features of the current conflict to match the precedent and utilizes preferences on other features to order successful precedents.[80] Learning in the implementation follows Alker and Christensen (1972) closely; the historical success or non-success (failure) of a precedent effects its possibility of future selection. Specifically, unsuccessful precedents are "forgotten" using a yearly exponential decay while successful precedents are reinforced. For cases in which no precedents are retrieved, two fall-back strategies exist. First, if the case has been identified as a phase in a ongoing episode and the episode's conflict type is known, then CHINA_WATCHER employs an inertia-escalation logic, based on the type of episode, to determine Chinese actions. Otherwise, CHINA_WATCHER uses an "operational code" system (George, 1969), analogous to the Alker-Christensen-Greenberg "operational charter" for their UN model.[81] Tanaka reports that for 383 conflict cases CHINA_WATCHER predicted outcomes with a 60 percent success rate for physical involvement in conflicts and 82 percent for verbal statements. An important aspect of CHINA_WATCHER is derivation of precedent chains leading up to specific actions. He concludes that the representation of knowledge requires improvement, perhaps along the lines of conceptual dependency. Further, Tanaka believes that precedent-logics require further conceptual clarification. He suggests that similarity measures could be improved along lines found in Winston's (1977, 1980) computational models of analogy.

Mefford (1984, 1986a) reports on a program that draws historical analogies to determine responses of the Soviet Union to crises in Eastern Europe, specifically the Czechoslovakian crisis of 1968. The program matches histories against cases to assemble composite precedents representing courses of action leading from the present into the future. After evaluating the cost and benefits of each path, the system prints an analysis and justification for the best response. To escape case specificity, Mefford constructs a coding language, containing a vocabulary for describing events and a classification hierarchy for linking terms in the language to other terms, events, and actors.[82] Since cases are repre-

[80]Each of the 383 cases are coded for features derived from 32 descriptor categories.

[81]This decision strategy assumes that precedent histories are accumulated and reduced to practice in standard operating procedures for the organization.

[82]In in the appendices of an earlier version of the paper, Mefford (1984) provides a taxonomy of actions by actor as well as the narratives and codings for the 1956 Hungarian Crisis and the 1968 Czechoslovakian Crisis.

sented as feature vectors, string matching provides the technology for retrieving precedents. To address partial matches, Mefford incorporates Levenstein metrics (Fu, 1982: 246-275), a general measure of distance between feature vectors based on the number of insertions, deletions, and substitutions required to transform one to the other. To address the problem of misaligned feature vectors, Mefford incorporates techniques based on syntactic pattern matching, specifically error-correcting parsers (Fu, 1982). Casting the modeling task as a form of problem solving, Mefford interprets the program's operation as generating the problem space on the basis of composite analogies. This problem space is then searched for the best solution. His program, like some critics of Soviet policy, finds a strategy which is better than the historical Soviet invasion of Czechoslovakia in 1968 because it achieves the same political objectives with half as many Soviet troops.[83]

Schrodt (1984a, 1985, 1986, 1987) reports on two programs that predict short-term foreign policy outcome by applying a precedent logic approach to international events data, including *World Events Interaction Survey* (WEIS) and Azar's (1982) *Conflict and Peace Data Bank* (COPDAB).[84] Schrodt defines *set prediction* as the ability to predict the immediate consequences of a political event and argues that humans experts perform it using analogy or precedents. Like his predecessors, Schrodt maintains that the success of a policy precedent reinforces its selectability. Schrodt discusses four general methods for feature matching:

- **Levenstein metrics** (Fu, 1982: 246-275) are a general measure of distance between feature vectors based on the number of insertions, deletions, and substitutions required to transform one to another other (Schrodt, 1984, 1985, 1987);

- **Syntactic pattern recognition techniques** (Fu, 1974) for recognizing combinations of features according to feature grammars (Schrodt, 1984, 1987);

- **Holland classifiers**[85] (Holland, 1975, 1986), a so-called "genetic algorithm" that learns through simulated "evolution" to classify objects by selecting "successful" feature detectors and combinations of feature detectors for recognizing grammars (Schrodt, 1986, 1987).[86] The goal of the model is to predict discrete set of events for 20 days, following a randomly chosen date, based on the previous 40 days. Working with COPDAB data with 15 feature codings, Schrodt uses the US/UK, US/France, US/West Germany cases because a large number of interactions are required to train the classifier. Schrodt (1986) finds that the Holland classifier achieves a performance level of about 95 percent of a mathematically optimized estimator.
  87

[83]Carbonell noted a similar problem with his analogy-based problem solver: It solved problems in a cognitively implausible but effective way. He was unsure if this was good or bad. J. Carbonell, talk on "Derivational Analogy," *Analogica-85*, Rutgers University, November, 1985.

[84]Schrodt admirably advertises his willingness to supply interested scholars with copies of his programs.

[85]The Holland classifier was implemented in UCSD Pascal on an Apple II microcomputer. It involves about 700 lines of code, two thirds of which deals with the classifier itself. Practically all runs used 32 rules in order to speed execution time – which was slow.

[86]As Holland notes, (1975: 49-53), genetic algorithms are similar to Markov chains with transformation probabilities that change as a function of history.

[87]Axelrod (1984, 1986) has also used "genetic learning" techniques to model conflict and cooperation. Although the methodology "genetic algorithms" fall within an earlier branch of artificial intelligence research,

- **"Connectionist learning" approaches** learn, after exposure to a sufficiently long training set, to recognize patterns by minimizing entropy given a prespecified energy function (Schrodt, 1987).[88]

Like earlier work in precedent logics, Schrodt focuses on feature vector matching. But unlike earlier work, Schrodt sacrifices detailed models of decision processes for general methods suitable for shallower applications to large datasets.

Following up on the 1980 discussion of SPD by Alker, Bennett, and Mefford, Alker (1985; Alker, Duffy, & Mallery, 1985) sketches a research programme for SPD stressing qualitative models of narrative with specific attention to modeling argument, debate and dramaturgical self-presentation. Precedent logics, efficient precedent retrieval, and felicitous semantic representation figure prominently. This paradigmatic statement presages the computational modeling of SPD narratives (Hurwitz, Mallery, Alker, & Duffy, 1986; Mallery, Hurwitz, *et al.*, 1986) using the RELATUS Natural Language System[89] (Duffy & Mallery, 1986). Thus far, approximately 100 hundred pages of narrative protocols for SPD game play have been syntactically parsed, semantically represented, and analyzed. The resulting representation is referentially integrated (co-referring terms are merged). It captures the relation structure of subject and object that is expressed in natural language sentences. The question of cognitively plausible representations of belief and intention has received attention (Mallery, 1987b) because of its striking salience in the strategic thinking found in SPD protocols. The kind of precedential reasoning found in previous precedent-based models is only now becoming operational in RELATUS. Mallery and Hurwitz (1987) discuss the application of precedent-based reasoning in RELATUS to foreign-policy decision-making.

The RELATUS system represents an important departure from previous prece-

their application in this work do not fall within computational politics. The reason is that the learning that these system incorporate is inherently non-symbolic. It depends on random mutation or "blind variation" (Axelrod, 1986: 1097, 1099) rather than cognitive processes). The non-symbolic nature of the enterprise follows from the formulation of the problem within a utilitarian framework. This non-symbolic orientation is reinforced by a methodology with low relevance for symbolic problem-solving. Indeed, the great leap from the simple account of genetics to conclusions about normative processes in political cognition is weak from the standpoints of a social science and computation. Contemporary biology tells us that "genetic learning alogrithms" are far simpler than the phenomena the purport to model. Modern cognitive science tells us that once a computational device as power as the human brain contains many emergent properties such as normative or deontic reasoning and that they are not directly coded in genetics. To claim a genetic basis for normative behavior is a line of social biological reasoning that has more relevance for ideology than political or cognitive science. The various emergent levels required to move from genetics to deontic reasoning mean that any mapping from genetics to deontic reasoning in the phenotype must be underdetermined were it to exist at all. Not the least of the problems is the mapping from cultural terms for relevant norms to brain or genetic locations that can influence reasoning. Thus, common sense suggests the burden of neurophysiological proof should fall with those arguing practicality. In more recent work, Axelrod (1987) has introduced a 3 move memory to his system. This new system, which still retains the Markov-like features of the "genetic algorithm" found a superior strategy to "tit for tat" (namely, initial defection followed by either continuing defection against cooperators or "tit for tat" against defectors). But, Axelrod (1987) had not yet recognized the crucial role of memory in these models but should report on this soon.

[88]Because these approaches are so computationally difficult, special-purpose hardware, for example, the Boltzmann Machine (Hinton, Sejnowski, & Ackley, 1984; Ackley, Hinton, & Sejnowski, 1985), is being developed to run connectionist programs.

[89]Section 5.3.1 further discusses RELATUS.

dent models in political science in terms of the approach to matching. All the earlier systems used some form of feature vectors and string matching techniques. Following along the path broken by Winston's analogy work (1980; 1984), RELATUS uses a new graph matching technique based on declarative constraints and graph traversals (Mallery, 1987a). Matching in both RELATUS and Winston's system is based on *relational graphs*. This differ from string matching of feature vectors because the terms matched can be linked by specific relations, or specific graph traversals. In other words, string matchers match fixed sequences of terms but not relations among the terms. Moreover, the significance of terms in a RELATUS or in a Winston binary-relation graph[90] is a function of their relational embedding rather than just the type of the term (a symbol). Thus, the significant representational difference is that fixed-length feature vectors cannot effectively represent relational structure while binary-relation graphs can. While binary relation graphs are arbitrarily expressive, fixed-length feature vectors are fundamentally limited in expressive power. Matchers for each formalism are similarly expressive or inexpressive. One major consequence following from relational graphs is that RELATUS style matchers can be used to construct referentially integrated semantic representations from syntactically-analyzed natural language input[91] whereas feature vector matchers have no hope of accomplishing this task.[92] In terms of representation and matching, the RELATUS approach to precedent logics marks a fundamental departure from earlier techniques because it has the expressive power to represent and match semantic structures derived from natural language texts.[93]

[90]These graphs are a special kind semantic network comprised of labeled binary relations, also known as ternary relations. The significant difference between the Winston graphs and the RELATUS graphs is that in RELATUS graphs the bidirectionality of relations is *more fully* exploited. The Winston system used relations, during matching of two situations, to generate both the relational subject and the relational object. RELATUS uses the bidirectionally of relations to get from the object of a relation to its subject. In fact, it is this bidirectional interpretation that allows the RELATUS reference system to achieve efficiency while retaining completeness (Mallery, 1987a).

[91]This is not true for contemporary Winston-style matchers; they are not intended to solve this problem. Those matchers achieve efficiency at the expense of completeness and they do not process declarative constraints — which could describe a syntactic parse. Completeness is lost when matches are filtered for importance criteria, such as causal links. Thus, the RELATUS matching system represents an improvement over the earlier Winston-style matchers, primarily, because it achieves efficiency without sacrificing completeness and it processes an extensible set of declarative constraints describing graph structures.

[92]Stanfill and Waltz (Stanfill & Waltz, 1986) have developed some experimental algorithms based on massive parallelism for performing precedential reasoning building from an underlying keyword search mechanism. This approach, however, is essentially statistical in nature and involves parallel feature vector matching. It therefore inherits the standard expressive limitation of the feature vector approach. Following from this limitation, it does not construct a felicitous semantic representation from narrative cases by performing a structural analysis of the text, involving syntactic analysis and referential integration.

[93]The graph structure matching used in the Winston and RELATUS matchers is also more powerful than Rete pattern matching used in OPS5 (Brown, *et al.*, 1985) and unification (Stickel, 1985) precisely because they are full graph matchers rather than term matchers. While Rete pattern matching can use tokens denoting frames in a frame representation, it cannot represent the connections between frames other than according to slot or relation types. This is, in effect, the same class of matching performed by unification which can match literals and variables as they appear in predicate formulae. Neither can capture the instances of predicates or slot relations. To do so would make them graph matchers. Thus, while Rete matching and unification are more powerful than feature vector matching because they can handle variable length feature vectors, they are not as powerful as full graph matching.

5 Models Using Natural Language

5.1 Content Analysis

5.1.1 Standard Computerized Content Analysis

Content-analytic techniques[94] (Lasswell, Leites, *et al.*, 1949; Lasswell, Lerner, & Pool, 1952; Pool, 1959; Stone *et al.*, 1966; Hays, 1969; Krippendorff, 1980; Weber, 1985; Namenwirth & Weber, 1986) have succeeded in certain macrosociological applications where general sociopolitical value orientations are the research focus (*e.g.*, Namenwirth, 1973; Weber, 1978), and in practical political applications, where the research orientation is discovery of fifth columnists or an enemy's intentions (Lasswell, Leites, *et al.*, 1949). Content analysis deploys a sophisticated form of keyword search to find the frequencies and correlations of interesting words. Word frequency is considered to reflect salience, and in turn, importance for a speaker. Word senses are disambiguated according to word co-occurrences. Standard content analysis programs neither analyze syntactic structures nor construct referentially-integrated semantic representations. But this weakness is also a strength, because it avoids the great technical difficulties required for computer understanding of unrestricted texts. For this reason, content analytical techniques could benefit from recent advances in parallel computers. Massive parallelism makes possible the application of standard computerized content analysis to very large unrestricted corpora and the performance of even more complex analyses with far greater rapidity than possible on serial computers.[95]

5.1.2 Semantic Content Analysis

While standard content-analytic techniques can provide information regarding the kinds of value orientations expressed in particular textual corpora, it cannot provide information about the *direction* of value orientations (Duffy, 1987: 21-40). The source of this shortcoming is the absence of a referentially-integrated semantic representation – something that massively parallel hardware will not fix. And, for political texts, this is of critical importance. Since understanding of natural language text by computers is AI-complete,[96] its full solution may be quite distant. Nevertheless, the technology is advancing and content analysis based on referentially integrated representations is now becoming possible. An early demonstration

[94]Holsti (1968), Krippendorff (1980), and Weber (1985) provide overviews of content analysis.

[95]Massively parallel computers, such as the Connection Machine (Hillis, 1986), have been applied to simple keyword search tasks. The results are startling. 200 megabytes of text can be searched in 200 milliseconds! However, it takes 2 minutes to load the machine readable text from serial storage devices into the connection machine's 64,000 independent processors. This bottleneck will be removed from the next version. This work is described by Stanfill and Kahle (1986). An obvious extension of this work is to replicate standard computerized content analysis using this fast new parallel architecture. Of course, parallelism will not in itself eliminate the problems for content analysis that arise from the absence of a structure analysis of sentences and a referentially integrated semantic representation.

[96]By analogy to NP-completeness in complexity theory, "AI-complete" is a term, first coined by Fanya S. Montalvo, to indicate that the difficulty of a computational problem is equivalent to solving the central AI problem, *i.e.* making computers as intelligent as people.

of semantic content analysis has been shown, beginning with natural-language protocols of sequential prisoner's dilemma games (Mallery, Hurwitz, Alker, Duffy, 1986; Mallery, 1987b). Although the English text must conform, at present, to a highly restricted "literal and explicit" form, the RELATUS system has been used to analyze over one hundred pages of protocols. The natural-language processing model embodied in RELATUS is presently moving from single-sense to multi-sense processing (Mallery, 1987b). A lexical classification system (Mallery, 1987b) identifies semantic representations of lexical items (words) corresponding to conceptual categories developed in advance by an analyst. The analyst can extract (or label) all occurrences of a conceptual category in a referentially integrated representation derived natural-language texts. The analyst can also construct taxonomies of categories and associate each category with a set of lexical-semantic constraints that pick it out. The semantic constraints include the various lexical items that may express the concept and constraints on relationships to other elements in the semantic representation. A lexicon of lexical categories contains the taxonomic links between a category and its specializations and generalizations. This allows categories to be differentiated as finely as required for lexical recognition while maintaining the taxonomic relationships between categories, and thus, enabling an analyst to identify categories with large varieties of lexical-semantic realizations. The existing lexical classification system, and an earlier version, have been used to locate game-relevant statements concerning belief, intention, and causal attributions in a study of the patterns leading to transitions between phases of defection and cooperation (Mallery, Hurwitz, Alker, Duffy, 1986; Hurwitz, Mallery, Alker, Duffy, 1987).[97]

5.1.3 Non-Computational Content Analysis

As natural-language processing technology advances, ideas from manual and computer-assisted content analyses can be incorporated into computerized semantic (pragmatic) content analysis systems. These "softer" kinds of content analysis range from approaches based on literary criticism through argument analysis to cognitive mapping and cognitive consistency affect graphs. In fact, growing numbers of political scientists are publishing work with either theoretical or practical relevance for semantic content analysis or cognitive models building from natural language.

The relationship of language to politics is becoming ever clearer to the political science community (Dallmayr, 1984; Shapiro, 1984). While hermeneutic studies (White, 1986) and deconstructivist literary analysis (Alker & Sylvan, 1986) have become quite fashionable at professional meetings, they are undoubtedly a response to the important role of language in societies, and the central problem of interpretation.[98] Indeed, language is the fundamental communicative tools that makes social organization possible. The ability to organize social action in turn gives rise to political power, which figures so centrally in the self-description of political science (Lasswell & Leites, 1949; Lasswell, Lerner, & Pool, 1952).

Some computational models of argument are beginning to emerge (Cohen, 1983, 1987a, 1987b; Alvorado, Dyer, Flowers, 1985; Alvorado, Dyer, Flowers, 1986; Ashley & Rissland, 1986, 1987; Rissland & Ashley, 1987). Except for Cohen (1983, 1987a, 1987b), this work is grounded in semantic universalism, and although it often illustrates interesting

[97]Section 5.3.1 further discusses RELATUS.

[98]Shapiro (1988) applies deconstructionist techniques to lay bare the political content found in visual representations such as photographs or advertising.

ideas, even the strongest efforts seem to lack the generality required for broad application. It is therefore not surprising that argument analyses relevant to international relations and foreign-policy decision-making remain largely precomputational. A few examples will illustrate this genre. Alker (1980) proposed partial formalization of Thucydides account of the Melian Dialogue using a dialectical logic of argumentation (Rescher, 1977). Devereux (1985a) shows that the formalism for argument analysis of Toulmin (1964), Birnbaum,[99] and Rescher (1977) are equivalent, develops a convergent procedure for graphing debates which he applies to the Lincoln-Douglas Debates. Devereux (1985b) manually applies his argument graphing technique to a narrative protocol game play in a sequential prisoner's dilemma.[100] Karapin & Dixon (1985) provide an rigorous characterization of arguments concerning deterrence based on argument graphs. The interpretation of the initial natural language and coding it according to the formalism is a potential source of bias introduced by the analyst. Bennett (1987) investigates empirical debates concerning SDI, using both computationally-informed and pre-computational approaches to argument, *e.g.*, (Eemeren & Grootendorst, 1984). Bennett concludes that analysis by computers is unlikely due to the high degree of vagueness exhibited in actual SDI arguments. Although the interpretive neutrality of this mapping may be doubtful, advances in natural language processing may provide a means for manipulating representations of arguments and debates in substantially more complex and rigorous ways. That argument and debate figure so prominently in international politics (*e.g.,* international law, negotiation, ideological competition, and policy formulation) suggests a fertile area for AI applications.

Onuf (1985) deploys concepts from ordinary language philosophy to attack the positivist view of international law and to advance an explanation of how international legal norms emerge from practice. Onuf understands legal norms as a form of "rule governed behavior." Drawing from Searle (1969) and Habermas (1981), Onuf erects an analytical framework that classifies rules in terms of three categories of speech acts: assertives, directives, and commissives. Onuf notes that certain tropes, specifically metaphor (assertives), metonymy (directives), and synecdoches (commissives) correspond to the possible fits of a speech act to the world. He argues that historical practices (customs) yield international laws as legal discourse about the speech acts associated with the practices expedites their acceptance as precedents to be cited in matters of principle, claims to rights, and justification of conduct. In more recent work, Onuf (1987) incorporates Piagetian developmentalism and Kohlberg's position on moral development to explain how culturally-mediated normative systems are transmitted, control social interactions, and justify distributions of rewards.

5.2 Semantic Universalist Approaches

Cognitive modeling approaches reflect the realization that symbolic-structural models of natural language processing and interpretation must move beyond superficial correlational belief attributions (Alker, 1975, 1979a). Thus, the other descendant of "cognitive consistency" pursued a natural language processing approach (Abelson & Carroll, 1965; Abelson & Reich,

[99] Devereux notes that his presentation of Birnbaum's theory is based on several chapters from an uncompleted Ph.D. thesis. However, some indication of McGuire's approach may be found in (Flowers, McGuire, & Birnbaum, 1982).

[100] These protocols are the same ones developed by the M.I.T. group during the 1980s and computationally analyzed using the relatus system. See the discussion in section 5.3.1.

1969; Abelson, 1973; Alker, 1975; Schank & Abelson, 1977; Carbonell, 1978; Abelson, 1979; Carbonell, 1981b; Kolodner, 1980, 1983a, 1983b; Schank, 1982; Dyer, 1983) grounded in a specific version of semantic universalism which they called "conceptual dependency" (Schank, 1972). Semantic universalism refers to those semantic theories that seek to capture common meanings through reduction to a set of putatively universal semantic primitives.[101] While Katz and Fodor (1963) proposed an early and widely-read theory of semantic universalism, the computer scientist Schank and the political psychologist Abelson (Schank & Abelson, 1977) reduced this general approach to a specific computational practice.[102] They proposed a specific set of 16 primitives (which was later expanded to 64 and more) for which they claimed cognitive plausibility. Composite representations are constructed by combining semantic primitives. Stereotypical characterizations of situations, or *scripts*, provide "slots" for the various events, actions, objects, and relationships for constrained domains. The restaurant script, for example, describes the process of eating at restaurant, including sitting down, reviewing the menu, selecting items, ordering, and so forth. Understanding a text about a restaurant, then, involves reducing surface word forms to CD primitives and fitting them into slots of an active script. This kind of "expectation-driven" or top-down understanding when conjoined with semantic primitives finds its natural complement in discrimination network models of perception (Kolodner, 1980, 1983a, 1983b; Simon & Feigenbaum, 1984). Like the primitives, Schank and Abelson have made universalist claims for scripts. Indeed, the CD approach has been tentatively applied to semantic translation between languages (Lytinen, 1984). Because of the universalist claims for specific primitives, scripts, and other knowledge-specific constructs, the CD approach is neo-platonist. Work within the CD paradigm or the "Yale school" of natural language processing has spread quite widely and today constitutes the major semantic universalist research tradition in computational linguistics.

Carbonell's "Politics" program (1978, 1981b), which represents an update of Abelson's "Goldwater machine," simulated ideological debate between a conservative, modeled on Barry Goldwater, and a liberal. Carbonell's thesis is that subjective understanding, or differential interpretation, arises from the presence of different motivational priorities, modeled as a goal tree, operating on a common set of factual knowledge structures using universal inference mechanisms. Since ideological thinking exhibits great individual variation, Carbonell selected it as a testbed for this thesis. Goals channel understanding by focusing attention on aspects of a situation effecting the understander most directly, *i.e.*, personally. In this way, goals call forth inferences to analyze consequences of events with the greatest personal affect. Then, these interpretations are remembered. Subjective understanding arises as understanders draw different interpretations because they focus on how events effect them personally and defocus the effects on others. A corollary of this thesis is that understanders know this, and therefore, understand other agents as understanding as a function of their own goals. But, different actors may attribute different goals to other agents because they each focus on the elements that effect them most. According to this view, intersubjectivity can be modeled by a mechanism that simulates the planning and counterplanning (Carbonell, 1981a, 1981b).[103] This view of understanding individuals is also applied to aggregations such as institutions and countries. Carbonell's thesis (1981b) contains a discussion of the TRIAD module which models social power relations and conflict situations. By holding the semantic coding invariant (inherent in CD), "politics" illustrates

[101] See Wilks (1987) for an overview of primitives in AI.

[102] Scripts are a version of Minsky's (1975) notion of stereotypical "frames" for understanding but differ in the choice of primitives.

[103] In his strong emphasis on planning, Carbonell combines a conceptual dependency approach to natural language processing with a the information processing emphasis on planning and problem solving.

the centrality of teleology in understanding.

Other politically relevant research has been done in the Yale tradition. The "Cyrus" program answered questions on the whereabouts of Secretary of State Cyrus Vance based on newspaper articles which it read into a CD representation. Another line of CD research investigates argument and debate. Flowers, Ringle, and Birnbaum (1982) report on a theory and a partial implementation that treats the role of personal attacks in adversary arguments with special reference to the Arab-Israeli case. More recently, the UCLA group reports efforts to use CD to model the argument structure of politico-economic newspaper editorials (Alvorado, Dyer, Flowers, 1985; Alvorado, Dyer, Flowers, 1986). Even if these texts are domain specific, they raise many of the problems of understanding unrestricted texts, including deliberative references (Mallery, 1987a), deictic representations, and difficult tropes such as metaphor.

For various technical reasons, these programs from the "Yale school" have not been scaled up beyond hand-crafted microworlds nor produced cognitively felicitous representations. One major reason is that since top-down processing requires extensive background knowledge already coded in pre-existing datastructures, the range of application is limited by the amount of background knowledge available to a system. Systems constructed within the CD approach have relied on text-specific deep semantic understanding provided by the implementors, *e.g.* (Dyer, 1983). Reliance on domain-specific knowledge and processing strategies have limited the volume of text handled by these systems. In the absence of adequate cognitive theories with sufficiently broad scope, the desire for high levels of performance leads into the trap of *ad hoc* implementation and limited generality or complete case-specificity.[104] In cases where free text from newspapers is analyzed, a preprocessor filters out articles outside the domain, passing articles in the domain through to the semantic parser. The system then does the best it can; memory organization strategies descended from scripts pick out whatever features they recognize. Further processing proceeds based the various slots in the scripts. Since the rate and cumulativity of knowledge codings by programmers control the amount of background knowledge available, CD is a natural language counterpart of knowledge intensive and domain-specific expert systems.

The cognitive plausibility of the CD approach is also suspect. Reduction to semantic primitives is generally considered cognitively implausible (Fodor *et al.*, 1980). The discrimination network model of perception also faces significant anomalies (Barsalou & Bower, 1984). Propositional attitudes, and related opaque constructs, also pose very serious problems for natural language approaches that reduce surface semantic structures to semantic primitives, such as CD (Mallery, 1987b).[105] Ontologically speaking, CD depends on meanings being a function of external reference (Perry & Barwise, 1983), yielding a strong dependence on the correspondence theory of truth, wherein the truth of a proposition depends

[104]However, intensive study of narrow domains and attempts to model cognitive processes considerably beyond the available theoretical analysis can be a useful means of identifying and cumulating some knowledge about the problems involved insofar as the actual cognitive processes are independent of earlier undiscovered levels. If they depend on earlier undiscovered cognitive processes, the exercise becomes essentially descriptive.

[105]Specifically, reduction to primitives is a form of substitution of equals during the perceptual process, a discrimination-net based mapping into the semantic representation, which means that referential opacity must be detected prior to representation in *ad hoc* structures. This prevents converting referential opacity to inferential opacity and using standard inferential techniques for the approach (within the standard representational format) to determine the effects of opacity on inferences. Section 6.1.3 discusses referential opacity.

on its correspondence to reality. This is, however, quite problematic for the social scientist interested precisely in modeling the differential perceptions of actors. A further difficulty is the synchronic nature of the theory. Because it is clear that languages (and interpretations using languages) evolve historically, a viable theory of language must explain linguistic innovation and evolution. Metaphor provides the main vehicle for semantic innovation identified thus far (Ricoeur, 1973). But, the assumption of literalism in language, a necessary consequence of a theory of meaning that finds equivalent meanings through reduction to primitives, precludes the non-literalism inherent in metaphor (Mallery & Duffy, 1987). The inability to coherently account for metaphor and referential opacity deals a serious blow to CD's claims for cognitive plausibility.

5.3 Lexicalist Approaches

There are various lexicalist approaches to syntactic analysis. Although most remain theoretical, some computational implementations have been done. However, the concern of these computational linguistics is typically with syntax and they often ignore semantic representation. The only existing lexicalist approach known to this author that builds domain-independent referentially-integrated semantic models of sufficient scale for credible social scientific uses is the RELATUS system.[106] In lieu of other credible lexicalist implementations that use referentially-integrated semantic representation in international relations applications, the discussion will be limited to the RELATUS system.[107]

5.3.1 Lexical-Interpretive Semantics

Lexical-interpretive semantics is an approach to natural-language semantics that constructs semantic representations from canonical grammatical relations and the original lexical items.[108] It incorporates a new theory of *semantic perception* – the process of mapping from a syntactic representation into a semantic representation – based on constraint posting (Mallery & Duffy, 1987). According to lexical-interpretive semantics, semantic representations are canonicalized only syntactically, not semantically or pragmatically. Instead of relying on static equivalences determined in advance, lexical-interpretive semantics requires meaning equivalences to be determined at their time of use, reference time. This requirement is met by the concept of a *meaning congruence class*, the set of syntactically-normalized semantic representations conforming to the linguistic experience of specific language users and satisfying their utterance-specific intentions. Lexical-interpretive semantics differs from approaches relying on semantic universals because meaning equivalences are determined dynamically at reference time for specific language users with individual histories rather than statically in advance for an idealized language user with a general but unspecific background knowledge.

In principle, lexical-interpretive semantics can avoid the distorting pitfalls incurred by exclusive reliance on a static analysis of meaning equivalence precisely because it

[106] The author is most interested in any counter-examples.

[107] It follows from the absence of competing lexicalist implementations in AI that AI models in international relations and foreign-policy decision-making have been semantic universalist in their implementation.

[108] The approach finds support in the available experimental psycholinguistic evidence (Fodor, *et al.*, 1980; Gentner & Landers, 1985).

can select equivalences from synonym or paraphrase congruence classes determined on the basis of dynamically changing, intentional contexts of language use. The major assumption underlying lexical-interpretive semantics is that meaning equivalences arise because alternative lexical realizations accomplish sufficiently similar speaker goals to allow substitution. A practical argument for dynamically determining meaning congruences is the intractability of a sufficiently detailed and nuanced static analysis that could capture subtle differences in speaker goals. This follows from the need to predict in advance all potential utterance situations and combinations of language-user effective-histories. [109] Although semantic canonicalization based on a general "semantic and pragmatic competence" renders static analyses of infinite language-user combinations tractable by fiat, it also reduces nuances so dramatically that intentional analysis and individual linguistic histories play a drastically diminished role. Thus, semantic canonicalization eliminates the need for effective history in determining an individual's word sense inductions and their application to recovering the speaker goals motivating specific choices of diction.

Lexical-interpretive semantics is hermeneutic because it emphasizes interpretation based on the individual effective-history of language users and the specific intentional structure of communicative situations. This is accomplished by actually representing the linguistic histories of individuals. By virtue of its emphasis on innovation in language and polysemy, lexical-interpretive semantics is perhaps most closely aligned with the phenomenological hermeneutics of Ricoeur (1975). Interpretation builds from an *eidetic* level of representation. The *eidetic* level of representation is a picture-like representation that is syntactically normalized but semantically variegated. The determination of meaning congruence classes becomes an early level of a more general and open-ended hermeneutic interpretation. This hermeneutic theory of meaning finds meaning through constructive classifications based on the experience of the interpreter rather than decomposition into atomic units (primitives).

This theory is implemented, up to the level of eidetic representation, in the RELATUS Natural Language System,[110] and research on subsequent levels is in progress. In the present system, syntactic analyses of sentences yield parse graphs. These are converted to declarative constraints and passed to a graph matcher that merges them into a semantic memory representing the beliefs of a single believer. In the process of merging sentence descriptions into memory, existing representations are found and new representations are created for new sentential objects. This process of *intersentential reference* yields a referentially integrated semantic representation that can be used for various kinds of analyses.

The main application of RELATUS to date is semantic content analysis of sequential prisoner's dilemma (SPD) games protocols and real-world conflict narratives drawn from Butterworth (1974). The hypothesis under investigation is that SPD players rely more on normative models of legitimate social interactions than on "rational" optimization strategies. Various papers discuss representing and analyzing over 150 pages of narrative protocols for game play in SPD games (Alker 1985; Alker, Duffy, & Mallery, 1985; Hurwitz, Mallery,

---

[109] A language user's *effective history* is his personal experience including the cultural and linguistic traditions inherited according to his position in society.

[110] The RELATUS Natural Language System is implemented by about 250 files, each containing around 1000 lines of LISP code each, and runs on Symbolics 3600 class Lisp machines. While the syntactic parser, the lexicon, and related utilities were designed and implemented by Duffy, the remainder of the system (including the representation, reference system, sentential constraint poster, question answering system, and editor mode) was designed and implemented by Mallery. Mallery (1987c) provides a more technical overview of the implementation than (Duffy & Mallery, 1986).

Alker & Duffy, 1986; Mallery, Hurwitz, Alker, Duffy, 1986; Mallery, 1987b). The protocol data originally was acquired from a series of experiments conducted by Alker and his graduate students at M.I.T. during the late 1970s and early 1980s (Alker & Hurwitz, 1980; Alker & Tanaka 1981). These experiments were significant because in additional to the standard behavioral traces they collected narratives accounts of undergraduates playing SPD. Following the earlier work of Emshoff and Ackoff (1970), SPD players wrote accounts of their interpretations of the game play both during and after the games. Thus, the represented protocols contain narrative accounts of game play augmented by natural language statements of player expectations and player beliefs about the course of play. A set of SPD extensions to the RELATUS editor mode provide an interface to an analysis package that performs an analysis of represented protocols according to phases of cooperation and defection. Another tool in the package shows phases of expectations. Other facilities allow an analyst to extract the beliefs, intentions, and causal attributions of players by phase. Studying linguistic expressions of norms, beliefs, and intentions demands a modeling technique capable of representing beliefs, and more generally, propositional attitudes. The RELATUS approach provides a new way of formalizing belief that is more consistent with common-sense reasoning by people than formalistic accounts of belief (Mallery, 1987b).

This protocol analysis is part of new analysis of sequential prisoner's dilemma that requires natural language modeling to demonstrate its arguments because it focuses on the interpretive and reflective aspects of play. A lexicalist/hermeneutic approach is required to test the hypothesis that the normative models and interpretive systems of players differ and to investigate the differences. Semantic universalist approaches would be unsuited to this modeling task because they would impose the category system of the analyst. Thus, this research represents the earliest results achieved in game-theoretic international relations through computational modeling of text. Continuing research on these protocols aims to induce archetypical normative models, fit normative models to players, and ascertain the role of precedent-based reasoning in game decisions.

6 Evaluating Computational Models

Although a literature of computational politics exists and continues to grow, there is no evaluative literature to distinguish sound models from spurious ones. This section adduces methodological concerns to guide model building and to warn against potential pitfalls. The specific issues discussed are:

- Various ways in which the felicitous representation of phenomena can be compromised;
- Assessment of decision models;
- Learning from the success or failure of models;
- Ways to reports results that improve scientific cumulation;
- Issues in determining the reliability and validity of models.

There is no attempt to apply these standards to existing work because computational politics has not yet advanced far enough to offer models with much practical use. It is better to encourage more research in the field to stimulate discussion before assessing its ability to actually model political phenomena. Nevertheless, work on evaluative criteria is important because it sensitizes researchers to the methodological issues that their models need to address. It also provides desiderata for determining the wisdom of promoting specific AI models for use in demonstrating scientific hypotheses or in critical foreign-policy applications.

Ideally, computational models should fit some facets of reality. But, precise correspondence to the empirical world is not the only reason for pursuing computational models of politics. The discipline of formulating a model requires the researcher to make specific commitments to a theory explaining the structure and function of components pieced together into a dynamic representation of the phenomena. This process alone, independent of the performance of the model, can help clarify and refine theories about the phenomena. Once operational, the model may not function well enough for predictive purposes; yet it may still yield heuristic insights into the motive structure of the phenomena. Political gaming (Guetzkow, *et al.*, 1963; Bloomfield & Gearin, 1969; Bloomfield, Gearin, & Foster, 1970; Bloomfield, 1982: 193-221), for example, does not pretend to simulate reality in every detail in order to provide valid insights and expose previously unrecognized issues. Similarly, some symbolic modeling efforts may succeed insofar as they expose previously unnoticed aspects of foreign-policy processes. The general point, then, is not to introduce positivistic criteria for judging scientific merit but rather to offer a set of considerations that help guide researchers toward better modeling practices. But the contrary should hold for AI applications to critical areas of national defense: Stringent standards of performance can minimize the chances of catastrophic failures.

6.1 Representational Felicity

Representational felicity in symbolic models is analogous to specification error in multivariate statistics. When the modeling problem is mispecified, the results are spurious. Characterizing felicitous specification is more involved and difficult for symbolic modeling because of the open-ended nature of the undertaking. This section attempts to identify various ways representational felicity can be compromised, but it is not exhaustive. Many of the felicity concerns discussed in this section, particularly the coding and meaning problems, constitute a revision of the operational assumptions of traditional content analysis and more generally political science.

6.1.1 The Coding Problem

The coding problem[111] arises when natural language statements must be converted into a target formal language or representation system by a human coder or a natural language system. Since formal languages, such as logic and even rule systems, are designed to represent single meanings unambiguously, the human coder must commit to a similarly narrow and precise interpretation of the original natural language. This practice raises a major difficulty; a single interpretation may not capture the significances attributed by different political actors or the same political actors under different situational circumstances. Interpretation

[111] An earlier version of this discussion appears in (Duffy & Mallery, 1986.

depends critically upon the situation and the knowledge of the interpreter (Mallery, Hurwitz, Duffy, 1987). Thus, in the coding process, the naive coder brings his own experience, categories, and knowledge to bear in the interpretation, rather than those of political agent or agents who are the object of the modeling effort.

This dilemma is reified by approaches which rely upon fixed-field codings for features (feature vectors[112]) and natural language paradigms that reduces meaning to combinations of fixed primitives.[113] The representational formalisms of these approaches (feature vectors and primitives) inherently reify the categories because they provide no mechanisms for representing *alternative* category systems. The modeler must first select a single set of features or primitives, and then, proceed to encode the data accordingly. One way to side-step the limitations of these mechanisms is to use different features or different primitives for each actor modeled.

To avoid the coding problem the social scientist needs to reconstruct the interpretive processes of political actors. First, phenomenal or raw data must be collected. Next, the interpretive processes of the political actors must be simulated, either automatically using a natural language system or manually using the social scientist's empathic understanding (George, 1959; Gadamer, 1960). In short, the coding problem may be avoided by making explicit the process of coding of data and by critically assessing the interpretations the coding process imputes to each political actor.

6.1.2 The Meaning Problem

The meaning problem[114] is related to the coding problem but concerns the nature and use of a representation rather than the preparation of textual data. There are two components to the problem: the nature of the representational formalism and the interpretive operations that it supports. It is known that the meaning of an utterance depends on the goal a speaker wishes to accomplish and the context in which it is uttered, in short, the intentional context (Appelt, 1985). Speakers may use flouting, metaphors or other pragmatic devices to make their points (Levinson, 1983). What they mean must therefore be determined *dynamically* according to the intentional context of the discourse. Another aspect of meaning depends on connotations of specific lexical items within a language. Speakers may chose specific words, rather than other words with equivalent dictionary definitions, because they wish to exploit the connotations of the chosen words.

These considerations about meaning are grave problems for reductionist or semantic universalist approaches to meaning representation, such as semantic universalism, rule systems or other formalist encodings, because they:

- perform static meaning analyses;

[112]See the discussion of the differences between feature vector matching and structural matching in section 4.4.

[113]See the discussion of semantic universalism in section 5.2

[114]Mallery and Duffy (1987) provide an extended discussion of the the meaning problem and include a number of examples. An earlier version of this discussion appears in (Duffy & Mallery, 1986) and is itself based on earlier drafts of (Mallery & Duffy, 1987).

- cannot support reinterpretation (because interpretation is done on encoding) according to different intentional contexts, practical situations, or presuppositional histories;[115]
- impose a single category system due to their ahistoricity;
- lose connotation as they dispense with the original lexical items
- assume a literalist theory of language that precludes interpretation of metaphor (and other tropes), and thus, eliminates the main source of semantic innovation.

Although a committed formalist might reply that all possible combinations could be encoded, there is clearly no hope for redeeming reductionism in this way. Even if all possible combinations could be coded, their exhaustive identification is clearly impossible.[116] The great empirical variability in human nature, individual backgrounds, organizational cultures, the sheer numbers of individuals and organizations, languages, cultures, and historico-technico epochs rules out the possibility of successful reductionist strategies. Of course, there are areas of "sparseness" (Minsky, 1985), such as descriptions of physical locations and physical actions, which tend to produce wide agreement across socio-linguistic variations. Here, the reductionist approach might seem to represent adequately. However, the introduction of social or political organization instantly shatters this meaning consensus. Thus, the only good representation is a semantic graph that retains the original lexical items, supports reinterpretation on the basis of constructivist classification of utterances, and allows for the possibility of metaphor interpretation.[117]

6.1.3 The Opacity Problem

Strategic language is the linguistic expression of belief and intentions as they relate to the plan-based pursuit of goals (Mallery, 1987b). The difference stems from greater structural

[115] To support reinterpretation, it is necessary to retain the original lexical items. But, retaining the original lexical items is more parsimoniously accomplished by a lexical-constructivist approach. Note that teleological focus operators in semantic universalist representations (*e.g.,* Carbonell, 1981b) can produce different problem-solving patterns and generalizations of interpretations but not fundamentally different interpretations.

[116] After presenting his doctoral research on parsing into a conceptual dependency representation and on translating between five languages, a recent doctoral dissertation on machine translation explains the difficulty as follows: "Before we can design a foolproof natural language system which produces the correct representation for a large class of texts, we must first know how to represent all of the different sorts of conceptual entities which the texts can be about" (Lytinen, 1984: 161).

[117] The RELATUS system (Duffy & Mallery, 1986) is one effort along these lines. There are some other lexicalist semantic network systems but as a rule their design criteria are not hermeneutically informed. It is important to note that this criteria does not necessarily rule out formal logic approaches to natural language understanding. Logic approaches *can* be used to create informal representations and associated inferential procedures for it. But, logic approaches typically orient its adherents away from the task of devising informal representations and weak inference methods. The general point is that any Turing equivalent formalism can be used to construct a system with any desired properties. Criticisms of logic-based approaches should be tempered by this caveat. Moreover, the meaning problem does not mean that rule systems and other earlier knowledge encodings are useless. Rather, it means that the social scientist must be aware of the important limitations and distorting propensities of these systems.

complexity and its heavy reliance on linguistic constructions known as opaque contexts (sometimes also referred to as propositional attitudes or referential opacity/transparency).[118] Such linguistic constructions require suspension of the facticity of their contents and explicit deliberation about their admissibility for inferential purposes. Typical examples include future or subjunctive verb tenses, belief statements, and affective expressions. The ability to represent and accurately (with respect to some cognitive agent's "ideologic"[119]) is a central issue for models of strategic action and planning.

Based on the implementation in RELATUS, Mallery (1987b) explains that for a representation system to successfully handle opacity it needs to be intensional[120] with referential opacity the norm. This converts referential opacity[121] to inferential opacity[122] because it makes the following conditions hold:

- **No Existential Generalization:** Creation of a representational entity does not necessarily imply existential generalization (the existence of the corresponding entity in the real world);
- **No Substitution of Equals:** As new items are input to (perceived by) the representation system, no substitution of equals is allowed.

When these two conditions hold, it is possible to add new items to the representation without determining in advance if they involve opacity. This means that the normal inferential mechanisms can draw on opaque representations but they must check the admissibility of the opaque propositions into the current belief context. If referential opacity were not converted to inferential opacity in this way, it would be necessary to determine in advance the opacity of items to be represented using *ad hoc*, pre-representational datastructures and procedures.[123] The conclusions of this line of reasoning are:

- Representational systems relying on substitution of equals or existential generalization at input time, e.g., semantic universalism, knowledge-based systems, and many

[118] See Levinson (1983) for a general discussion from a linguistic viewpoint and Fawcett (1986) for a specific discussion of opaque constructs from a logic programming perspective.

[119] An *ideologic* is, roughly, the collection non-standard inference rules that account for ideologically or "psycho-logically" distorted reasoning (Alker, 1979b).

[120] An *intensional representation* contains objects that do not directly refer to the real world. Any such correspondence needs to be established by inference. An *extensional representation* contains only objects that correspond to the real world. Thus, the ability to represent counterfactual objects and relations requires an intensional representation.

[121] *Referential opacity* is the situation in which a believer may know two different terms but not know that they corefer. The standard example due to Frege is that of the morning star and the evening star. Unless a believer knows that they are both Venus, the believer does not know they are the same. Thus, not knowing they both refer to Venus makes their reference opaque. Mallery (1987b) provides further examples, discussion, and pointers to literature.

[122] Instead of resolving the opacity during the reference process, it is resolved later by inferences to determine the equality (identity) of different denotational descriptions.

[123] Of course, this is not a major problem for small cases in which everything is already handled with *ad hoc* methods.

logic-based approaches (but not the more sophisticated versions), face serious (perhaps insurmountable) difficulties in representing the central components of strategic action and inference;

- Lexicalist semantic theories provide the only obvious means of representing strategic inference.

Thus, referential opacity provides a powerful desiderata for determining the adequacy of AI representations and inference methods for models of planning and counterplanning. Any system that purports reason about belief needs to explain how it handles referential opacity.[124]

6.1.4 The Planning Problem

Recent results in complexity theory confirm the intuitions underlying notions of "bounded rationality" (Simon, 1957, 1969, 1982) and "muddling through" (Lindbloom, 1959, 1965). These results suggest important complexity theoretic constraints on how people and organizations plan and also how computers might model their planning behavior. The results are for a complete and general form of planning, known as *non-linear conjunctive planning.* Conjunctive planning is formulating a series of actions that when executed achieve several goals. Non-linear planning refers to plans that express dependencies between actions as partial orders, deferring commitment to a final execution order until all actions are specified, that is, the plan is complete.

Chapman (1985) proves that non-linear conjunctive planning belongs to the class of np-complete algorithms, a class which is believed to be exponential and therefore computationally intractable.[125] The main source of difficulty is the interaction of actions in ways that defeat the goals they are intended to achieve. Determining in advance whether one action defeats another goal is the primary source of computational intractability. An interesting but omitted point is that the source of difficulty is a version of the "frame problem" (McCarthy & Hayes, 1969), or determining whether the state of the world remains the same across observations and actions.[126] Chapman also shows that planning is undecidable, *i.e.*, that it may be impossible to predict if a plan achieving the goals does or does not exist.[127]

Since these results are for non-linear conjunctive planning, there may be domain-specific planning strategies with better computational properties. The important

[124] Mallery (1987b) provides a more detailed discussion of referential opacity and its implementation in RELATUS. An interesting consequence developed there, is that communication assumes referential transparency at its base. This argues that people understand other cognitive agents by projecting their self-models onto other, and then, they incrementally construct any deviations they impute to other agent. This is a testable psychological hypothesis.

[125] If an algorithm requires exponential time to execute on a serial computer, it will require exponential processors to execute on a parallel computer. Thus, the computational intractability of plan formation remains a problem even for the parallel cases of multiple individuals or multiple organizations.

[126] For discussions of the frame problem as it has been formulated in AI see, McDermott (1982) and Shoham (1986). A standard interpretation of the "frame problem" is the temporal persistence of states of achieved by actions. Mallery, (1987b) discusses in greater detail the relationship of the "frame problem" to Chapman's result.

[127] See section 6.5.1 for further discussion of undecidability and intractability.

point is that non-linear conjunctive planning provides an abstract and general characterization of the task with a precise complexity characterization. Of course, counterplanning, the situation in which an agent formulates plans to thwart or assist the plans of other agents, is more complex than planning without competitors. Moreover, the frame problem is very real and is not simply a problem of the temporal persistence of states of affairs in the world; a competitor may undo actions while an agent is doing something else.

The results for planning are applicable to decision-making and problem-solving. The main conclusion for human and organizational problem-solving is that people must perform these tasks using different methods, probably grounded in common-sense reasoning. Massive parallelism in the human mind (multiplied by the size of an organization) can allow domain specific strategies to be devised for each problem domain. Alternatively, memory-intensive strategies can obviate the need for much planning from first principles by allowing an agent to replay his own past solutions to problems (precedents) or to copy and adapt the past solutions of other agents. Analogical transfer of solutions from different domains provide another potential source of ready-made plans.

Although Chapman's results may not be directly transferable to the methods used by people and organizations, it seems reasonable to assume that they constitute a real constraint on the planning methods actually used. The full range of implications for AI models of organizational behavior may take some time to uncover. But, there are some clear implications.

- "Bounded rationality" is a provable constraint on problem-solving.
- Humans and organizations cannot be assured or accused of producing in finite time correct plans that achieve their goals. Complex domains increase the likelihood of producing incorrect plans.
- Since humans and organizations cannot be certain of formulating effective plans at all, let alone in reasonable time, it is better to save time and increase confidence in a plan's working correctly by recycling old plans (applying precedents) which have succeeded in the past. Rederiving plans from first principles seems quite implausible.
- Any domain-independent AI program attempting to model human and organizational planning will certainly face Chapman's result. Therefore, AI systems will perform better if they finesse the main difficulties of planning with precedent-based planning.
- Large plans can be made more tractable by decomposing them into smaller independent pieces whenever possible because formulating the plan is exponential in the length of the plan.[128] Division of labor, or functional specialization, in societies, organizations, and the human mind inherently performs such a decomposition, although independence may not be guaranteed.[129]

[128]Because interaction of plan components was the main source of computational intractability, increasing the independence of plan components reduces the possible interactions, *ergo*, the computational difficulty. Reasoning from Chapman's complexity result, therefore, supports Simon's (1962) argument for the necessity of "near decomposibility" in complex systems.

[129]Task decomposition, however, requires communication to coordinate the combination of smaller tasks into larger collective actions. Language, or conversations, become an essential communicative medium for action coordination. If conversations are required to formulate and execute social or organizational plans, felicitous modeling of such processes requires modeling the conversations.

- Since the interaction of actions which might defeat goals is hard to predict, short plans utilizing the environment as the domain representation can side-step the problem in practice.[130] Although this strategy may be fine for humans and organizations in situations amenable to incremental solutions, it is not helpful for modeling planning in computational politics.

Planning, or goal-directed behavior, is a key aspect of human cognition and organizational problem-solving – even though it may not always be easily identified by the naive observer. For organization theory, as it has been formulated for foreign-policy decision-making, the intractability of formal planning introduces some difficulties in the specific explanations of problem-solving but these difficulties seem like they can be addressed by precedent logics. The consequences for each of the major organization theories are as follows:

- **Rational Actor Model:** (Allison, 1971: 10-38; Steinbruner, 1974; 25-46) The rational actor model assumes the possibility of optimizing decisions; but the undecidability of non-linear planning suggests an optimal solution may not be exist. Moreover, the exponential nature of planning suggests that the solutions selected will just "satisfice." Of course, rational actor approaches typically rely on quantitative approaches, deemed cognitively implausible in section 3.1.4, which are not formally subject to Chapman's result. However, if their quantitative formalisms do not predict in ways similar to symbolic ones (*i.e.*, there is no counterpart to the planning problem in the formalism), then this may be another cognitive plausibility problem for the approach.
- **Organizational Process Model:** The organizational process model (Allison, 1971: 67-96), building from the organization theory (March & Simon, 1958; Cyert & March, 1963; Crecine, 1969) inspired by the generalized problem-solver, avoids the difficulties of undecidability and intractability to the extent that organizational behavior is governed by standard operating procedures (SOPs). However, the need to formulate new plans, *i.e.* devise new SOPs, brings out the planning problem.
- **Bureaucratic Politics Model:** The bureaucratic politics model (Allison, 1971: 147-181) fares quite well because it emphasizes competing factions and the structure of the organization. However, to the extent that the units making decisions (individuals and factions) are considered to be maximizing their goals or interests. the model comes into doubt. Considering the decision-making elements to be responding to their subjective perceptions, or perhaps suboptimal formulation and pursuit of goals, handles the difficulties of the non-linear planning result. But this casts no light on the cognitive processes at work and their aggregations in organizations without treating planning and bringing out the planning problem.
- **Cybernetic Model:** Steinbruner (1974: 47-87) applies organizational cybernetics (Ashby, 1956; Beer, 1959, 1966, 1975; Deutsch, 1963) to foreign policy decision-making.[131] A pure formulation of the cybernetic model does not treat planning *per se*

[130] Brooks (1987) uses this strategy in his mobile robot project at the M.I.T. Artificial Intelligence Laboratory.

[131] Although Steinbruner makes some claims for affinity with the general-problem solving approach, the primary thrust of his theory is cybernetic because emphasizes control and communication rather than problem solving. Steinbruner's claim notwithstanding, this makes Steinbruner's model a different type of model from Allison's organizational process model, which emphasizes goal-directed problem-solving.

because planning falls out of homeostatic maintenance: When an undesirable internal state is registered by the organization, a appropriate action is selected to remove the undesirable state and return to homeostasis. Because planning is not directly addressed, the planning problem has no direct consequence for the theory.[132] The planning problem does have relevance for a revised organizational theory incorporating cybernetic insights: The planning problem is a modern complexity-theoretic result with greater specific relevance for problem-solving in organizations than Ashby's (1956) notion of *requisite variety*.

The planning problem is considerably worse in *strategic situations*, where organizational or individual planning depends on predicting the plans and counterplans of other actors.[133] Whether the relationship is adversarial or cooperative, the planning problem arises recursively with each level of reflection ("I will plan to ... because he is planning to ... because he thinks that I am planning to ..."). A case in point is the interdependence problematique (Cooper, 1968; Keohane & Nye, 1977): Planning to cooperate with interdependent countries is an instance of the strategic situation. Designing control strategies to attenuate dysfunctional consequences (*e.g.*, fluctuations in exchange rates) requires consideration of the decisions (planning) of individual corporations and countries as well as the subdivisions whose debates yield policy and action outcomes for the collectivities to which they belong. Thus, the planning problem needs to inform not just theories of individual and organizational problem-solving but also theories of strategic situations, involving multiple political actors that make plans to cooperate or compete.

A major use of AI models is to simulate planning, problem-solving or decision-making. If this cannot be done tractably, the models will be of limited value. The complexity results for non-linear conjunctive planning suggest that memory-intensive precedent logics (see section 4.4) offer the most plausible strategies for humans, organizations, and AI models to pursue. The ability to recycle old plans and to maintain the independence of plan elements also mitigates the deleterious consequences.

6.1.5 Inferential Poverty

The cognitive plausibility of inference mechanisms presently employed in AI models is entirely suspect. It is clear that individuals have much richer inferential repertoires. AI and cognitive science have a considerable distance to go before artificial intelligence systems will begin to simulate inferential mechanisms as rich as those ordinary people use everyday. The cognitive status of organizational inference or decision-making is considerably less clear. Although organizations may be composed of individuals, it may not be necessary to simulate the cognitive processing of each individual. To the extent that organizational decision-making is adequately paraphrased by higher level abstractions such as precedents, "operational codes," or other clearly articulated inference mechanisms, AI models may be able to provide reasonably felicitous approximations. Since the adequacy of inferential mechanism depends on the specifics of a model, each AI model needs to be assessed individually. The important point is that inferential poverty requires attention as a potential source of distortions.

[132]The theory suffers from the absence of a coherent account of problem-solving.

[133]See Mallery (1987b) for a more detailed discussion of the strategic situation, planning and its manifestations in language.

Since many approaches rely on heuristic rules as their inference mechanism, a few words on their cognitive plausibility are needed. Rules are essentially disembodied concepts. As such, they have no systematic grounding in phenomena. Since they are ungrounded, knowledge-based systems will encounter difficulties simulating the learning that cognitive agents might do when faced with failures in their domain concepts. Rules express a formalization of a domain. They require considerable effort to distill from domain experts. This suggests that they represent, at best, *a post hoc* reconstruction of the expert's domain theory. The experts typically rely on the kind of informal knowledge that rules cannot adequately express, *i.e.*, historical cases. An even more serious problem is the resort to theoretically and empirically unmotivated *ad hoc* rules in order to make case-specific models work. Even if rules have shortcomings or present possible sources of implausibility, they offer powerful and well-developed tools for constructing AI models that cannot be ignored until better alternatives are available. In short, some model of inference is better than none. In the meantime, constant vigilance to their shortcomings can yield better model specifications and more plausible interpretations of results.

6.1.6 Data Poverty

The main unifying characteristic of AI models of political phenomena is that they require large amounts of phenomenal data. Like the situation for cognitive mapping, it is usually difficult or impossible to collect sufficient data for many cognitive modeling problems. The absence of sufficient data may preclude the development of plausible models in many areas. Thus, an important question for research designs is to show that the appropriate amounts and types of data are available. Good models will be those that can exploit the available data to show interesting results. Systems that operate at a fairly high level of abstraction or show logical consequences inherent in *a priori* formulations will probably be most successful in the near term.

Even if sufficient data is available to construct a model, there may be significant difficulties in converting it into a form suitable for computational processing. Schemes that rely on large numbers of coders to translate data into a formal language are likely to be costly and suffer from intercoder reliability problems (see section 6.5.4). Natural language processing, even if restricted to certain machine-parsable constructions, promises a means of coding large sets in ways that make the coding process formal and reproducible. The major benefit of automatic natural language coding is that the target language need not fully interpret the data. That is, important components of the interpretation process may been performed by the AI system or the system user. Freeing the coder from the requirement of producing "objective" interpretations allows the model to simulate aspects of the interpretation process. Even with natural language processing, human coders performed important cognitive tasks as they convert the initial text to a machine parsable form.[134] However, as the natural language processing improves. this source of unreliability will recede.

[134]Because a full solution to the natural language processing problem requires a full solution to the artificial intelligence problem, it is AI-complete. Between the present and the solution of the AI problem, there are a series of ever more sophisticated natural language processing models. The human coders will have to reformulate free-form texts according to the contemporary capabilities of their system. But, as their system advances, the amount of reformulation decreases.

6.1.7 Strategic Simplification

Any political model, whether computational or not, presupposes a theory. The facts and theory elements incorporated into the model are determined by the theory. The utility of any theory rests in its ability to simplify a phenomena in ways that expose its core causal (structural and functional) and teleological components. The more effective the simplification, the higher the explanatory power of the theory or model. Conversely, failure to treat relevant aspects of the phenomena reduces the explanatory power for all cases in which the omissions are relevant. The ability to simplify at all, however, requires explanation of generalizations that hold for the phenomena. Thus the task of theory is to identify a small set of regularities which, through repeated application, can account for the phenomenal behavior. These are precisely the kinds of principles that can make a for good AI simulation in the social sciences. If the explanation cannot be distilled to a few principles and facts, the data will never be collected and the simulation never performed, or else, the simulation will have little to do with the empirical phenomena. To extent that the model performs adequately, the embodied theory simplifies the phenomena without introducing distortions.

6.1.8 Equifinality Problem

The equifinality problem occurs when a model predicts the same outcomes as observed in historical cases but does so on the basis of *different* factors and processes. Of course, all artificial models, in so far as they are a simplification of reality, must suffer from the equifinality problem at some level – the level at which theoretically unmotivated computational devices substitute or simulate phenomenal objects. As an explanatory tool, an AI model should be useful for filling in aspects of political processes for which direct data is not available. Use in this capacity requires establishing that the unobservable aspects actually correspond to the processes being modeled. This correspondence can never be fully established because of the problem is fundamentally inductive. However, confidence in these unobservables can be increased by inferring their presence on the basis of correlated deviations in the behavior observables as the model is applied to many cases.[135] This is not very helpful for case-specific models. Since application to many cases may not always be feasible, a heuristic method can also increase confidence in the unobservables. In the heuristic method, the first step is to clearly delineate the equifinal level and distinguish higher levels building from it. Given this, two general heuristics can minimize the likihood of drawing spurious inferences due the equifinality problem:

- All imputations from the model must be derived from at least one level higher than the base level, where the model interfaces to theoretically unmotivated computational devices.
- The AI modeler must be able to construct a credible theoretical account of phenomena that supports the imputed unobservables.[136]

[135] The analogy underlying this deductive detection of unobservables is the detection of Pluto and Neptune by astronomers. They deduced their existence on the basis of a deductive mathematical theory the explained observed orbital deviations of other planets by postulating the gravitational pull of the unknown planets.

[136] Statistical polimetrics has employed this technique to avoid equifinality problems that can occasionally arise from model specification errors.

In general, the further removed the predictive level from the equifinal level, the greater the confidence in unobservables. The main implication for model specification is to clearly distinguish the equifinal level. This may, however, prove more difficult in practice because most or all of the model may be at the equifinal level. Thus, confidence must ultimately rely on theoretical coherence, agreement with complementary theories, and the absence of competing accounts with comparable explanatory power.

6.1.9 Depth Versus Breadth of Coverage

There is a basic trade-off in AI systems between the depth of coverage and the breadth of coverage. The success of knowledge-based systems and systems operating on microworlds illustrates how narrow coverage provides the constraint that allows depth to be achieved. This can be understood by noting that scientists and their computers are capable of handling some specific order of complexity in a historico-technico epoch. Thus, depth of coverage follows from narrowing breadth. If the same complexity management capacity is applied to broad phenomenal coverage, lower depth will be achieved. The key question, however, is which approach can best benefit from increases in scientific manpower, cumulation. For the depth-first approach, cumulation means expanding breadth based on depth. For the breadth-first approach, cumulation means expanding depth based on breadth. The history of domain-specific research and microworld AI systems indicates that expanding from depth is more difficult than expanding from breadth. The reasons are:

- Achieving depth in domain-specific applications relies on case-specific, *i.e., ad hoc*, methods;
- The absence of a general, domain-independent foundation means that new case-specific knowledge and heuristics are required for each new domain;
- The focus is on obtaining quick results in narrow domains rather than developing general principles.

Although there are few breadth-first research efforts to provide positive examples, there are general reasons why this approach should ultimately prove more effective than domain-specific approaches:

- General capabilities are only developed on the basis of general principles with wide application;
- The general approach must address the *a priori* cognitive structures required for systems to function rather than case-specific mechanisms;
- Investigating the *a priori* structure of cognition is a research strategy that proceeds from ontology and epistemology;
- This pursuit of unified *a apriori* cognitive theories can exploit sources of constraint from different parts of the same theories to constrain possible mechanisms, leading to an overdetermination of the possible solutions;

To the extent that international relations models are devised using domain-independent technology, there is less likelihood of introduction of *ad hoc* structures to make the models work, which in turn, degrades the generality of the embodied political theory. In fact, the presence of general, cognitively plausible underlying mechanism channels modeling efforts in directions that are better grounded in a substantive sense. Thus, the cognitive plausibility of the modeling tool has important consequences for the theoretical credibility of political theories that might build from it.

6.1.10 Mechanism Artifacts

Artifacts of specific mechanisms used in a computational model may account for important characteristics of the observed results. For example, a counterfactual simulation may not incorporate learning mechanisms that would yield novel strategies for a decisionmaker. Or, a simulation may be performed on a computer of limited power, and therefore, be unable to process enough information or search a large enough space to provide plausible results. Note that the computational unit in the phenomena is a human being with 10^{10} neurons operating in parallel and these are combined in even more powerful computational devices called organizations, governments, and countries. Neither an IBM PC, a Cray supercomputer, nor a Connection Machine is much of a match in terms of raw computational power. Another source of distortion is the quantity of information contained in a model relative to the phenomena. The political modeler always has inadequate inferential mechanisms, computational crunch, and information. This demands reliance on clever theoretical formulations of the modeling task that minimize the distortions in the model. Since not all distortions can be eliminated (by definition), modelers need to normalize results for known defects in the computational tool and mismatches to the phenomena. Critiquing simulation results in this way can yield a list of methodological problems for improvement in subsequent models.

6.2 Assessing Decision Models

Much of the AI modeling literature in international relations and foreign policy emphasizes models of decision-making. This section outlines a few general issues beyond the preceding discussion of representational felicity (see section 6.1).

6.2.1 Assumptions

The possibility of using symbolic modeling techniques to simulate organizational decision-making depends critically on several assumptions:

- The presence of identifiable problem-solving (Sylvan, 1987b);
- The availability of sufficient information to test a model's account of problem-solving (Sylvan, 1987b);

- The processing of information, the assessment of situations, and the selection of actions must be governed by rules, *e.g.*, operational codes (Job and Johnson, 1986);
- Policies and choices (options) must not be subject to frequent revision (Job and Johnson, 1986);
- Memory of past actions of a political actor as well as other actors allows learning to be incorporated into subsequent actions (Job and Johnson, 1986);
- The successful reproduction of the organization's identity, which could be characterized as relative stability in high level goals and methods for achieving them.

Several factors can make the modeling effort easier or harder.

- Hierarchical organizations should be easier to model because top-down strategies require less data and less powerful computers; By modeling the top-echalons of the organization first and incrementally extending the model to lower levels as the higher levels fail to explain the phenomena, the top-down strategy can achieve the best return per unit of modeling effort.
- Anarchical or distributed organizations should be more difficult to model because the use of top-down strategies to model them will miss more essential elements leading to decisions.
- More bureaucratized organizations evince more regular behavior that is easier to model.
- Organizations with tighter environmental constraints will have less latitude for decisional variation.
- More specialized organizations will be easier to model because the sources of variation are less.

These factors are not exhaustive but the provide some indications of issues that will effect the chances of developing good decision models. Further indicators of the feasibility of developing specific models follow from computational complexity considerations and the availability of data.

6.2.2 Levels of Analysis

Sylvan (1987b) notes the importance of selecting the correct level of analysis for the modeling problem.

- **Causal-functional simulations** of international political systems assume that organizational units simply respond to the state of the environment according to fixed-function strategies;

- **Unitary actor simulations** perform cognitive processing to select actions without internal disputes and pursue goals according to criteria framed at the actor level;

- **Internally differentiated actors** have multiple loci of cognitive processing that focus on multi-level (from the collective actor down to the decisional unit), partially conflicting goals and compete for the action selected by the organization;

  1. The cognitive processing of a single top decision is modeled on the assumption that significant decisional output depends on the "great man;"
  2. The cognitive processing and communicational interaction of a small elite of top decision makers (Leites, 1951) is modeled on the grounds that the group controls the output of the organization;
  3. The competing interests and information processing of different organizational sectors determine organizational decisions in the model;
  4. A model of an internally differentiated actor is juxtaposed with similar models of other actors in the environment that can influence the course of decision either directly (pressure groups) or indirectly by creating environmental constraints or enablements (meta-power).

- **Complete bureaucratization** holds that standard operating procedures, operational codes, "bureaucratic culture" dominates the decision of organizations independent of the individuals filling positions within them;

- **Precedent-guided decisions** characterize organizations that exploit collective memory to select, or combine fragments of, similar responses used on past occasions;

- **Mythical or ideological decision-making** expects organizations to act on the basis of strong social myths or ideologies and that there may be little connection with reality (Hurwitz, 1987).

The existence of many different levels and dimensions of analysis suggests the use of computational models to keep track of them and to test their internal and external consistency.

6.2.3 Third Order Analysis

Constructing models of foreign-policy decision-making is significantly different in substance from studying foreign-policy decision-making *per se*. A comparative study of foreign policy has two major components:

- Collect substantive case descriptions;

- Develop a set of abstractions or a descriptive theory covering the cases.

Computational models of foreign policy decision-making go beyond the study of comparative foreign policy in five ways because the investigator must:

- Formulate a theory of the decision-making process;
- Implement the theory using a suitable computational formalism;
- Code the case descriptions according to the descriptive theory for the chosen formalism;
- Test the implemented theory on the coded case data;
- Evaluate the performance of the model in light of actual or hypothetical cases.[137]

The ways in which computational models exceed comparative foreign policy amount to adding two orders of analysis:

- The theory of the decision process is a second order analysis;
- The implementation of the theory is a third order analysis.

In general, the discipline of computational modeling makes explicit the theory of decision underlying case descriptions as well as the computational specification of the decision process.

6.3 Learning from Models

6.3.1 Success Analysis

The desirable outcome for a symbolic model is that it successfully predicts substantive behavior. When this happy result obtains, it is necessary to identify the operative components of the model that are responsible for its predictive capabilities. Other components will presumably have only marginal effects. Thus, the scientific documentation of the research needs to report the key predictive principles and assess their relative importance, perhaps through sensitivity analysis.

6.3.2 Failure Analysis

Once an investigator realizes that his AI model fails to adequately capture the relevant facets of the phenomenal world, the failure can be turned to his advantage through "fault analysis techniques." The failure pattern of computer models provides a tool for identifying what aspect of the phenomena have not been modeled and what components of the methodology are not up to snuff. For example, Job and Johnson (1986) note that their model of

[137] Evaluation of postdictive uses of models is simpler than evaluation of other uses, such as predictive or heuristics uses.

U.S. decision-making concerning the Dominican Republic failed to predict certain decisions because it did not incorporate a mechanism for representing delayed responses to information. This is a problem both for the computational tools and data acquisition. On the one hand, they need a symbolic analog to lag functions in multiple regression models. On the other hand, they also need information about what determines delayed responses by the U.S. decision processes. In general, failure analysis provides a powerful tool for improving both political theories and symbolic modeling methodology. Although some omissions from theories could be identified without resort to computational models, there should be types of omissions which are too subtle or involve such complex interactions that human analysts may not easily identify them. Computational modeling provides a method for systematically identifying gaps in theories, and therefore, offers itself as a tool to assist theory formation and theory refinement.

6.4 Reporting Results

A major problem with the existing literature on computational modeling is that it is usually very difficult to understand how a model is implemented or what it does on the basis of published papers.[138] Explicit identification of key methodological components including citations to the AI literature where appropriate, can help tremendously. For comparability, replicability and scientific cumulation, authors should always report the following:

- The principles that make the simulation work.
- Flowchart overviewing processing stages and associated datastructures.
- The general representational approach and its specific use, *e.g.* frames, conceptual dependency used to represent actors and actions.
- The programming language or AI package used, *e.g.* Common LISP, PROLOG.
- Rough statement of system size, *e.g.* number of lines of code.
- Rough statement of data size, *e.g.* number of knowledge representation units.
- Rough statement of inferential mechanism complexity, *e.g.* number of rules.
- Assessment of system case-specificity, *i.e.* the *ad hocness* coefficient.
- Sources for data, *e.g.* the *New York Times*, interviews.
- Preparation of data for input to the model, *e.g.* rephrasing in terms of rules.
- Mapping from input to the representation, *e.g.* event statements are converted into PROLOG expressions.
- All rules, axioms, and other data should be provided whenever feasible so that others can replicate the results. Otherwise, this material should be available from the author.

[138] This problem is also prevalent in the AI literature because systems are complex. It is mitigated where the technology is generally known.

The advancing power of computers available to political scientists will lead to programs and system that are too large to fully treat in published papers and too large to be useful in the form of a source-code listing. Thus, arrangements should be made for exchanging machine readable versions of programs, systems, and data. In conjunction with a published literature documenting the theory and practice of particular AI modeling systems, the availability of machine readable versions of the system provide the best vehicle for scientific cumulation. This can lead to cooperative work by multiple researchers and greater aggregate achievements.

6.5 Reliability and Validity

6.5.1 Computational Limitations

Reliability and validity questions concerning substantive political issues depend on the computational foundation from which they build. This section considers these foundational issues.

- **Halting Problem:** A foundational theorem of computer science, known as the *halting problem*, proves that in the general case it is not possible to prove whether a universal Turing machine will halt (terminate). Because the universal Turing machine provides the theoretical foundation (the computational model) for work in computer science, this result effects computational models of politics implemented on general-purpose computers. It follows that if it cannot be shown that an AI system will halt (terminate), the correctness of the system cannot be established because there is never any result whose correctness can be checked![139]
- **Undecidability:** A problem is undecidable if, given an instance of the problem, no algorithms exists for determining whether or not a solution exists.[140] As noted in section 6.1.4, planning for conjunctive goals is undecidable. Thus, practical modeling efforts need to pay careful attention to whether or not parts of the simulation are decidable and determine the consequences for the overall simulation.
- **Intractability:** Exponential (NP-complete) algorithms are believed to be computationally intractable. To the extent that cognitive models require simulating exponential processes (*e.g.*, planning), it may be impossible to produce practical simulations. However, cognitive agents have finite computational capacities, and therefore, it may be possible to simulate their ability to handle exponential computations, failing just as they do.[141] In general, the complexity of practical political simulations can increase

[139] For discussion of the halting problem, see any basic textbook on the theory of computation, for example (Hopcroft & Ullman, 1979; Boolos & Jeffrey, 1980; Lewis & Papadimitriou, 1981).

[140] See Hopcroft & Ullman (1979: 177-216) for a more detailed discussion of undecidability.

[141] Progress in parallel processing promises to bring computers closer to the capacities required to simulate cognitive agents. Specifically, if the size of the memory of a computer is equal to the number of processors (as in a connection machine), memory limitations prevent requirements for exponential processors from taking effect. The human brain has on the order of 10^{10} neurons. There are about 5 billion people on the planet. In the worst case, modeling international relations could require cognitive models of all people

as computer hardware and software technologies advance Thus, while some problems may be practically intractable today, they may become feasible tomorrow.

- **Program Correctness:** Contemporary technology for proving the correctness of programs can handle only simple programs, which are orders of magnitude simpler than the computational mechanisms demanded by computational politics. To the extent that the halting problem or undecidability are issues – and they certainly seem relevant for cognitive models – there can be no hope of showing the correctness of program or knowledge-based systems simulating cognitive processing.[142]

- **Hardware Reliability:** Assuming that none of the above computational issues cause trouble for a political simulation, small hardware failures (such as dropping bits) may produce spurious results. Although this is not a problem for the social scientist who can just rerun the simulation, it presents a potentially serious concern for those would who have AI systems that make critical decisions, such as those required in SDI applications.

The general conclusions for the underlying computational mechanisms is that no *a priori* demonstrations of reliable operation or valid conclusions are possible. This means that the performance of AI models must be judged in terms of their empirical behavior.

6.5.2 Empirical Induction

In highly constrained domains, the output of AI systems, given specific inputs, can be compared to results known to be true or derived by other methods. The key here is to develop new verification procedures for AI models and combine them with the existing knowledge of empirical validation in political science, *e.g.* (Eckstein, 1975; Hermann, Philips, & Thorson, 1978; Cook & Campbell, 1979). As long as the complexity of the output is small and the mapping to empirical benchmarks straight forward, exhaustive comparisons of input sets to output sets provides an effective means of verifying the correct operation of an AI system. However, as problem size increases or the domain becomes unconstrained (as in interesting political models), exhaustive comparison is no longer feasible. For these models, partial comparisons of input and output sets to known domain fact provides the basis for inductively inferring the correctness of the model. As more comparisons are done, confidence in the model increases. However, the possibility always remains that the model will fail on some as yet unexamined input and output set. Moreover, it is usually not possible for models in international relations and foreign-policy decision-making to obtain domain facts for comparison.[143] In sum, empirical induction provides a fallible but useful means of verifying AI models.

---

and organizations. That would require a parallel computer with about 10^{19} processors. Although this is substantially beyond the 64,000 processors available in today's Connection Machines, our current technology is only 10^{14} short of the worst case requirement!

[142]See, for example, Sedgewick (1983) for an introduction to algorithms and issues of correctness.

[143]Indeed, the unavailability of domain facts and the impracticability of laboratory experiments are important reasons for developing computational model of politics.

6.5.3 Logical Coherence

Empirical induction clearly fails in counterfactual analyses. No simple one-to-one mapping from AI model to the real world can be used to test the model. For similar analyses where no simple empirical mapping is possible, analysis of the logical coherence of the model (the argument that it makes) provides another weak verification method.[144] This amounts to verifying the system's internal consistency and logical structure. Verification of predictions requires establishing the correctness of all datapoints and verifying the inferential mechanisms conform to the modeled domain. Since human reasoning (and therefore, organizational reasoning) is not well-understood, it cannot be verified. Thus, AI models can only help us check our own thinking about how social actors reason.

6.5.4 Lessons from Content Analysis

Content analysis has already faced the issues of reliability and validity.[145] Content analysis interprets results by considering the reliability of a model and the validity of the components of the model. In general, reliability refers to consistency and accuracy of knowledge encodings for a single coder over time or across many coders. In the AI context, a coder should be understood as the combination of a knowledge engineer and a domain expert. This may or may not be the same person. If the AI model is an expert system, the coding task is culling and preparing rules. If the AI system uses a natural language front-end, the coding task is converting the text to a format which the natural language system can process. Drawing on the analogy from content analysis, reliability of an AI model has several aspects:

- **Stability:** Stability refers to the stability of coding of data for analysis over time. When the *same* coder prepares input data for representations, whether the rules, logical formulae, or English text, stability can be ascertained by examining the differences in the resulting representation. For the case of AI models, stability needs to encompass the stability of inferences that can be derived from the representation. Thus, the concern in *inferential stability* is how inferences are effected by variations in the knowledge encoded in the representation and how the codings vary for the same coder over time.

- **Reproducibility:** Although stability measures the consistency of private understandings, reproducibility, also known as *intercoder reliability*, expresses the similarities and differences between different coders. Reproducibility problems can occur when coding rules are ambiguous or inconsistent, when errors in data exist, when differences in situational factor prompt different codings, or simply when cognitive differences between the coders lead to different classifications. As in stability, the concern here is with the impact of coding on the inferences produced by the AI system.

[144] The method is weak due to the general difficulties of proving correctness of programs. Nevertheless, logical analysis of an AI system, often referred to as a "model theory" by logicians, provides a basis for arguing whether its behavior is correct or incorrect. Ultimately, logical analysis founders to the extent that human reasoning is not understood.

[145] This section draws on Weber (1985: 16-21) and Krippendorff (1980: 130-154) but reformulates their criteria for the case of computational politics.

- **Accuracy:** Accuracy of coding show the extent to which knowledge encoding corresponds to a standard or norm. This is the strongest test of reliability which can be applied to the performance of human coders once coding standards are available. For an expert system, coding accuracy refers to application of the classificational criteria used in the model to code source materials into rules. For a natural-language based system, accuracy refers to the correct conversion of raw text to the linguistic model handled by the AI system.

The analogy with the notion of validity in content analysis is not as clear for AI models with the exception of those using natural language as their source of knowledge.

- **Construct Validity:** Construct validity refers to the correlation of different measure for the same attribute in content analysis. Two types of correlation are considered:

  1. *Convergent constructs* use different indicators that express same underlying concept;
  2. *Divergent constructs* uses opposing indicators that must not be present if their opposed indicator is present.

  In knowledge-based systems, construct validity can naturally arise in the coding of rules because of the need to map different surface expressions of propositions into single canonical rules. In natural language systems, construct validity can span both the coding process and the interpretations performed by the AI system. For inferences based on knowledge representations, the notion of construct validity can guide alternative sources for the same inferences and suggest contradictory knowledge that would prevent inferences from going through.

- **Hypothesis Validity:** Hypthothesis validity is the extent to which the knowledge representation yields inferences conforming to the substantive theory and the procedural theory that the AI model purports to implement. For example, given a theory of organizational decision-making and data on a specific organization, the AI model should produce inferences consistent with both. To the extent that it does not, there may be problems in the validity of the theory of decision-making or the data (assuming the implementation accurately reflects these).

- **Predictive Validity:** Predictive validity is the extent to which the knowledge representation yields inferences that predict actual outcomes in the phenomenal world. Either the future, the past (postdiction), or concurrent events may be the object of prediction. This is a particularly effective test of validity because it encourages the use of generalizable methods and depends on data not directly under the control of the investigator.

The present literature on computational politics does not treat the important questions of reliability and validity of models. Although this is understandable because current research strives to devise workable models, it is nevertheless important to develop designs that facilitate showing the reliability and validity of the models. Only in this way will computational politics develop the scientific techniques necessary to rigorously demonstrate results.

7 Conclusions

7.1 Responsible Uses

AI modeling in international relations and foreign-policy decision-making is an area of basic research. The field of inquiry is just beginning to take form and is little understood. However, one thing is clear: *AI modeling in this field is presently suitable only for descriptive research, generation of heuristic insights, and theory clarification but not prescriptive policy-making.* Scientific responsibility, common-sense, and the discussion of the myriad ways in which symbolic models can fail, demands foregoing applications to real-world decision problems in foreign affairs. The infeasibility of demonstrating the correctness of computational systems and the lack of knowledge about the reliability and validity of AI models mean that any application to critical areas of national defense could produce catastrophic failures. Considering the grave consequences of system failure, the proponents of AI system in critical national defense applications need to *prove* that failure cannot produce dire consequences. These proofs should be redundantly checked by competing teams of scientists, who seek to show how it can fail. Even then, a good design, like a good design for nuclear reactors, should plan for failure and make failure modes tolerable. Ultimately, though, the responsibility for a catastrophic failure of an AI system in a critical application will rest not with the designers or the reviewing scientists but with the public officials whose signatures bring the system on line.

7.2 The Inadequacy of Formal Planning

Formal planning cannot explain problem-solving or planning as found in the substance and process of foreign policy. Thus, strategic policy in the military and intelligence spheres is not effectively computable. The formal complexity-theoretic results for non-linear planning for conjunctive goals (see section 6.1.4) show that general, domain-independent planning is computationally intractable and also undecidable. The planning referred to here is involves composing a series of actions that achieve multiple goals, roughly a means-ends analysis more sophisticated than generalized problem solving (GPS) (see section 4.1). This means that simulating the planning behavior of a single cognitive agent without resort to domain-specific special-case planning methods is not feasible for a computer, whether serial or parallel, or humans, whether individually or collectively. The foreign policy case is worse. It is necessary not just to simulate or predict the planning behavior of multiple political actors in international arenas but also to simulate and predict the counterplanning of the actors vis-a-vis other actors in the system. Because formal planning, specifically generalized problem-solving figures so prominently in the organization theories deployed to understand foreign policy decision-making (see section 6.1.4), there is a pressing need to revise these theories in light of the formal results on planning. In short, formal planning does not provide an adequate basis for understanding the substance or process of foreign policy or simulating foreign policy decision-making.

7.3 Precedent Logics as a Computable Alternative

Precedent logics (see section 4.4) are the most promising alternative to formal planning as an account of the substance and process of foreign policy. The results on formal planning notwithstanding, people and organizations do manage to anticipate, to varying degrees, the strategies, plans, and counterplans of competitors and allies. This constitutes an existence proof for a computationally tractable, albeit partial, means of simulating the planning and counterplanning of political actors. The virtue of precedent logics is that they allow old plans known to work for some situations to be recycled in applications to new but similar situations – without attempting to rederive new plans from first principles. Instead of determining whether the precedent will work, it is only necessary to ascertain how it might fail, *i.e.*, how the present situation differs from the past one. Thus, by recycling knowledge of the past, and perhaps transferring it across domains, precedent logics gain desirable computational properties that make them plausible in principle. For this reason, they seem capable of providing a computationally adequate account of situations with multiple political actors engaged in planning and counterplanning. They also suggest revisions for the organization theories used to understand and simulate foreign policy decision-making. But, beyond mere computational complexity considerations, precedent logics evince properties not found in formal planning systems, specifically:

- They strategically simplify;
- They capture the "lessons" of history (May, 1973);
- They have high cognitive plausibility, especially due to their grounding in informal, common-sense knowledge;
- They, in conjunction with related metaphorical processes, provide a source for semantic innovation or learning;
- They paraphrase the essential ingredients of decision and inference;
- They simulate the prestructuring or "seeing as" found in problem framing;

In sum, precedent logics offer a computationally tractable and cognitive plausible approach to theorizing about and simulating the substance and process of foreign policy.

7.4 Effective History, Natural Language, and Historical Method

Precedent logics are just one component of a hermeneutical approach to understanding and modeling political action. If the cognitive processing, whether individual or collective, is memory intensive, then the contents of memory, the experience or *effective history* of political actors provides the stock of potential precedents that could be applied to new situations. Furthermore, the taxonomic categories that guide the precedent-based reasoning, or more general cognitive processing, of political actors are largely induced from their linguistic experience or transferred to them through language. Because the predominant experiential mode of political actors is linguistic, natural language representation must figure equally prominently in AI models of international relations or foreign policy decision-making. Quite apart from various technical difficulties, semantic universalism (see section 5.2) is unsuited to the task of representing different effective histories because it imposes a single classificational scheme to interpret utterances of political actors with very different effective histories, and therefore, very different taxonomic categories. In contrast, lexicalist approaches to natural language representation (see section 5.3.1) are expressly suited to the problems of interpreting and representing the utterances of political actors with divergent effective histories. In sum, natural language processing using a lexicalist-hermeneutical approach allow:

- Simulation of the subjective interpretation processes of cognitive agents, whether individuals or organizations;
- Construction of informal representations suited to precedent reasoning, analogy, and metaphor;
- Quick construction of large datastructures from text;
- Simulation of the classificational, learning, and understanding processes of cognitive agents;
- Representation of conversational communications between language-users.

Thus, natural language modeling, precedent logics, and other forms of experiential learning form the basis for a symbolic modeling grounded in historical method. This, then, comprises a computational method that is closely allied with the traditional historical method most often found in international relations and foreign policy studies.

These considerations, and the fact that other formalisms such as knowledge-based systems are subsumed by a successful natural language approach, argue that natural language modeling targeted at common-sense reasoning and learning promise the greatest potential payoff for symbolic modeling. Their potential vastly outstrips that of knowledge-based systems because, instead of being inherently domain-specific, precedent-based natural language systems promise multiple domain generality.

7.5 Importing Computational Insights into Political Theory

Just as cybernetics and systems theory enriched the descriptive vocabulary of international relations modelers and theorists, such as Deutsch (1963), Forrester (1971), Choucri and North (1975), Meadows *et al.* (1972), distinctly computational constructs like "bounded rationality" and "satisficing" have been introduced by first generation AI researchers such as Herbert Simon. The information processing view of social systems concerns itself primarily with issues revolving around search and its computational requirements. A new wave of concepts from learning theory and common-sense reasoning is now coming into international relations. Analogical and precedent-based reasoning as well as inductive concept formation are important ideas that color the analysis of organizations and individuals as history-based learners. Complexity theoretic results for planning and issues from parallel processing, such as database consistency, can also help guide theory formation and pose empirical questions. Computational hermeneutics (Mallery, Hurwitz, Duffy, 1987) illustrates the non-reductionist nature of meaning and proposes metaphor as a major source of innovation. In general, the introduction of new descriptive vocabularies into international relations opens up new ways of seeing traditional problems, and presumably, new more penetrating analysis of system function. For the foreign policy scholar, basic knowledge of computational modeling can help not only understand new simulation techniques as they emerge from the laboratory but also inspire computationally grounded explanations of phenomena. Thus, if nothing else, AI modeling efforts will hasten the incorporation of these new vocabularies and may perhaps return some notions of organization, culture, and ideology to AI.

7.6 Importing Social Science into AI

The directionality of scientific exchange is not one way. Although AI maintains some interchange with cognitive science, there is a need for greater input from social scientists. In particular, the problem of rationality – understood as the *apriori* structure of cognition and the development of cognition within social communities – requires more than an asocial monological approach (see sections 3.3 and 3.4). Once the role of social embedding becomes a focal problem, the social scientists who investigate it become AI researchers. The role of ethnomethodology and social psychology as well as other levels of social analysis are gaining currency among AI researchers (Winograd, 1980; Winograd & Flores, 1986; Mallery, Hurwitz, Duffy, 1987). The main topic of current interest include small group interaction, socialization, the family structure, organizational psychology, interpersonal psychology, ideational formation, and linguistic development. Understanding the ways in which organizations make decisions and learn seems to be an area with great potential contributions to AI, mostly because it is more readily observable than the inner workings of an asocial, monological individual. Given the tremendous amount of information acquired through socialization within linguistic communities, it is unlikely that a true asocial individual would evince much of what is considered cognitive processing.

8 Acknowledgments

An earlier version of this report was submitted as a Master's Thesis (February, 1988) to the M. I. T. Department of Political Science. Hayward R. Alker, Jr. was the thesis supervisor and Lincoln P. Bloomfield was the second reader.

Hayward R. Alker, Jr. provided extensive comments, endless encouragement, and the opportunity to pursue this research at M.I.T. Lincoln P. Bloomfield provided useful comments and pointed out how to sharpen the focus. Roger Hurwitz provided many useful comments on this and earlier drafts. Conversations with John Batali, Chris Stacy, Paul Davis, Gavan Duffy, Ken W. Haas, Roger Hurwitz, Robert J. P. Ingria, David Kirsh, Rick Lathrop, Mike Leavitt, Jintae Lee, David McAllester, Marvin L. Minsky, Fanya S. Montalvo, Alan Newell, Ronald Rivest, Herbert Simon, Richard Stallman, and David V. Wallace, Terry Winograd, Patrick H. Winston provided part of the discursive background reflected in this document. These are just some of the names that come to mind; many others too numerous to identify also contributed to the discourse background. The open-mindedness and forbearance of the Department of Political Science contributed in no small measure to this research. Any errors or shortcomings are due to the author or time constraints.

This report began during 1980 as an effort to apply Winston's analogical reasoning system to international relations texts. After understanding the limitations of the contemporary analogy system, the effort turned to developing an independent natural-language system for large text applications during July, 1983. By the summer of 1984, the system, RELATUS,[146] had progressed to the point where Gavan Duffy and I were able to demonstrate it at the 1984 National Conference on Artificial Intelligence. After many improvements and significant reimplementations of the reference system and the semantic inverter, it came time to write up a Master's Thesis. Unfortunately, the research had become too extensive to treat adequately in a Master's Thesis for Political Science. Consequently, the problem was abstracted and I wrote this document during the Spring of 1987 to overview the symbolic modeling literature in international relations and foreign-policy decision-making and to delineate the field of computational politics.

The RELATUS system would not have been developed were it not for the collaboration of Gavan Duffy, who developed an excellent natural language parser from which semantic representations could be constructed. The moral and practical support of J. Michael Brady, Berthold P. K. Horn, Tom Knight, Marvin L. Minsky, Gerald J. Sussman, and Patrick H. Winston is gratefully acknowledged. Patrick H. Winston provided to me the initial opportunity to work, between January, 1980 and January, 1983, with his analogy system. This enabled me to learn the practical tools of AI, and later, to develop my own system. Boris Katz helped me grasp the essential concepts required to build large systems and provided the impetus to develop RELATUS. But, Richard Greenblatt, Robert W. Kerns and Kent M. Pitman taught me how to implement large systems. Lloyd S. Etheredge helped instill the confidence required to tackle difficult problems. The unique environment of the M.I.T. Artificial Intelligence Laboratory, constituted by its many past and present members, provided a stimulating intellectual atmosphere. The absence of impediments to research, openness, freedom, equality, and sponteneity of the M.I.T. Artificial Intelligence Laboratory reflects the imprimatur of its founders, particularly Marvin L. Minsky. Perpetuating the tradition of creative excellence in the face of creeping bureaucracy, corporate competition, and financial constraints falls to their successors.

The author was supported in part from September, 1987 through May, 1988 by a John D. and Catherine R. MacAurthur Foundation grant for research on international security and arms control to the Center for International Studies of the Massachusetts Institute of Technology. Most of the writing for this report was not funded due to the vagaries of DARPA funding and the politics of resource allocation within the AI laboratory. Financial support from September, 1984 through January, 1987 as well as ever present computational support from the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology made this work possible. Support for the M.I.T. Artificial Intelligence Laboratory's

[146]The RELATUS Natural Language System owned by John C. Mallery and Gavan Duffy, each holding copyrights (1983, 1984, 1985, 1986, 1987) to their respective components. Gavan Duffy designed and implemented the syntactic parser, categorial disambiguator, lexicon, and related utilities. John C. Mallery designed and implemented the semantic representation system, the sentential constraint poster, the noun-phrase quantification system, the reference system, the semantic inverter, the question answering system, the prisoner's dilemma analysis system, the lexical classifier, the precedential reasoning system, and related utilities, including the RELATUS ZMACS mode.

artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract number N00014-85-K-0124. American Microsystems, Inc., and Robert S. Kirk, supported RELATUS research during the summer of 1985. Symbolics, Inc., and John L. Kulp, supported RELATUS research during the summer of 1986. Finally, the longstanding financial and moral support of Bayard M. Mallery and Virginia R. Mallery is ultimately responsible for the possibility of this work. Of course, the views expressed in this document are those of the author and not of any institutions that may have supported aspects of the research, including the United States Government.

9 References

R. P. Abelson and M. J. Rosenberg, [1958], "Symbolic Psychologic: A Model of Attitudinal Cognition," *Behavioral Science*, 1958, 3: 1-13.

R. P. Abelson, [1959], "Modes of Resolution of Belief Dilemmas," *Journal of Conflict Resolution*, 1959, 3: 343-352.

R. P. Abelson and J. D. Carroll, [1965], "Computer Simulation of Individual Belief Systems," *American Behavioral Scientist*, 1965, 8: 24-30.

R. P. Abelson, [1968], "Psychological Implication," in *Theories of Cognitive Consistency: A Sourcebook*, R. P. Abelson, *et al.*, editors, Chicago: Rand McNally, 1968: 112-139.

R. P. Abelson and C. M. Reich, [1969], "Implication Molecules: A Method for Extracting Meaning from Input Sentences," in D. E. Walker and L. M. Norton, editors, *Proceedings of The International Joint Conference on Artificial Intelligence*, Boston: The Mitre Corporation, 1969.

R. P. Abelson, [1973], "The Structure of Belief Systems," in R. C. Schank and K. M. Colby, editors, *Computer Models of Thought and Language*, San Francisco: W. H. Freeman, 1973.

R. P. Abelson, [1979], "Differences Between Belief and Knowledge Systems," *Cognitive Science*, 1979, 3: 335-366.

D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, [1985], "A Learning Algorithm for Boltzmann Machines," *Cognitive Science*, 1985, 9(1): 147-169.

H. R. Alker Jr., [1970], "Directive Behavior: A Desirable Orientation for Mathematical Social Science," *Revue Francais de Sociologie*, special issue entitled, *Analyse de Systemes en Sciences Sociales (I)*, 1970: 99-122.

H. R. Alker Jr., [1971], "Assessing the Impact of The U.N. Collective Security System: An Operational Multicultural Approach," *Proceedings of The American Society of International Law*, 1971, 65: 33-39.

H. R. Alker Jr., [1975], "Polimetrics: Its Descriptive Foundations," in (Greenstein & Polsby, 1975: 140-210).

H. R. Alker Jr., [1976], "On Cybernetic Hierarchies Within Sociocultural Change," paper presented at *The Conference on Indicators of Social and Cultural Change within Computer Models of World Development*, Moscow, U.S.S.R., June, 1976.

H. R. Alker Jr., [1978], "From Political Cybernetics to Global Modeling," Paper delivered at *The 1978 Meeting of the American Political Science Association*, New York City, 1978.

H. R. Alker Jr., [1979a], "From Information Processing Research to The Sciences of Human Communication," *Informatique et Sciences Humains*, mars-juin, 1979, 40-41: 407-420.

H. R. Alker Jr., [1979b], "Some Dynamics of Ideo-logical Thought," paper presented at The 1979 Meeting of the International Political Science Association, Rio de Janeiro, 1979.

H. R. Alker Jr., [1980], "The Dialectical Logic of Thucydides Melian Dialogue," paper presented at *The 1980 Meeting of The International Studies Association*, March, 1980.

H. R. Alker Jr. [1981], "Logic, Dialectics, Politics," in H. R. Alker, Jr., editor *Dialectical Logics for The Political Sciences*, a special issue of *Poznan Studies*, 1981, 7. A more extensive unpublished manuscript is entitled "Learning How to Do Analysis in Political and Social Science," Cambridge: Department of Political Science, M.I.T., mimeo, circa 1977.

H. R. Alker Jr., [1984], "Historical Argumentation and Statistical Inference: Towards More Appropriate Logics for Historical Research," *Historical Methods*, Summer 1984, 17 (3): 164-173.

H. R. Alker, Jr., [1985], "Reflective Resolution of Sequential Prisoner's Dilemmas," paper presented at *The 1985 Annual Meeting of Society for General Systems Research*, March 1985.

H. R. Alker, Jr., [1986], "Bit Flows, Rewrites, Social Talk: Toward More Adequate Informational Ontologies," *Proceedings of the Tokyo University Conference on Information and Its Functions*, 1986.

H. R. Alker, Jr. [1987], "The Presumption of Anarchy in World Politics," in H. R. Alker, Jr., and R. Ashley, editors, *After Neorealism: Anarchy, Power, and International Collaboration*, forthcoming.

H. R. Alker Jr., J. Bennett, and D. Mefford, [1980], "Generalized Precedent Logics for Resolving Insecurity Dilemmas," *International Interactions*, 1980, 7: 165-206.

H. R. Alker Jr., and C. Christensen, [1972], "From Causal Modeling to Artificial Intelligence: The Evolution of a UN Peace-Making Simulation," in J. La Ponce and P. Smoker, editors, *Experimentation and Simulation in Political Science*, Toronto: University of Toronto Press, 1972.

H. R. Alker Jr., G. Duffy, and J. C. Mallery, [1985], "From Quantity to Quality: A New Research Program in Sequential Prisoner's Dilemmas," paper presented at *The 1985 Meeting of The American Political Science Association*, New Orleans, 1985.

H. R. Alker, Jr., and W. Greenberg, [1971], "The U.N. Charter: Alternate Pasts and Alternate Futures," in E. H. Fedder, editor, *The United Nations: Problems and Prospects*, St Louis: Center for International Studies, University of Missouri, 1971.

H. R. Alker, Jr., and W. Greenberg, [1976], On Simulating Collective Security Regime Alternatives," in G. M. Bonham and M. J. Shapiro, editors, *Thought and Action in Foreign Policy*, Basel: Birkhauser Verlag 1976: 263-306.

H. R. Alker, Jr. and R. Hurwitz, [1980], *Resolving Prisoner's Dilemmas*, Washington, D. C.: American Political Science Association, 1980.

H. R. Alker Jr., W. G. Lehnert, and D. Schneider, [1985], "Two Reinterpretations of Toynbee's Jesus: Explorations in Computational Hermeneutics," in (Tonfoni, 1985: 49-94).

H. R. Alker, Jr. and D. J. Sylvan, [1986], "Political Discourse Analysis," paper presented at *The 1986 Meeting of The American Political Science Association*, Washington, D. C., September, 1986.

H. R. Alker, Jr. and A. Tanaka, [1981], "Resolutional Possibilities in 'Historical' Prisoner's Dilemmas," paper presented at the *1981 Annual Meeting of the International Studies Association*, Philadelphia, March, 1981.

G. T. Allison, [1971], *The Essence of Decision*, Boston: Little, Brown, & Co., 1971.

S. J. Alvarado, M. G. Dyer, M. Flowers, [1985], "Memory Representation and Retrieval for Editorial Comprehension," *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*, 1985: 228-235.

S. J. Alvarado, M. G. Dyer, M. Flowers, [1986], "Editorial Comprehension in OpEd Through Argument Units," *Proceedings of the 1986 National Conference on Artificial Intelligence*: 250-256.

P. A. Anderson and S. J. Thorson, [1982], "Artificial Intelligence Based Simulations of Foreign Policy Decision Making," *Behavioral Science*, April, 1982, 27(2): 176-193.

S. J. Andriole, [1987], "Artificial Intelligence and Strategic Intelligence Analysis and Production: Some Opportunities and the Limits to Growth," in (Cimbala, 1987a: 21-34).

K. O. Apel, [1980], "Noam Chomsky's Theory of Language and Contemporary Philosophy: A Case Study in The Philosophy of Science," in K. O. Apel, *Towards a Transformation of Philosophy*, G. Adey and D. Frisby, translators, London: Routledge, Kegan Paul, 1980.

D. E. Appelt, [1985], "Planning English Referring Expressions," *Artificial Intelligence*, 1985, 26: 1-33.

W. R. Ashby, [1956], *An Introduction to Cybernetics*, New York: Wiley, 1956.

K. D. Ashley and E. L. Rissland, [1986], "Toward Modeling Legal Argument," in A. A. Martino and F. S. Natali, eds, *Automated Analysis of Legal Texts, Logic, Informatics, Law,* New York: Elsevier, 1986: 19-30.

K. D. Ashley and E. L. Rissland, [1987], "But, See, Accord: Generating "Blue Book" Citations in HYPO," *Proceedings of The First International Conference on Artificial Intelligence and Law, Boston, MA, May 27-29, 1987,*, New York: Association for Computing Machinery, 1987: 67-74.

R. Axelrod, [1976], *The Structure of Decision: The Cognitive Maps of Political Elites*, Princeton: Princeton University Press, 1976.

R. Axelrod, [1984], *The Evolution of Cooperation*, New York: Basic Books, 1984.

R. Axelrod, [1986], "An Evolutionary Approach to Norms," *American Political Science Review*, December, 1986, 80(4): 1095-1111.

R. Axelrod, [1986], "Prisoner's Dilemma and The Genetic Algorithm," Cambridge: Talk at Thinking Machines Corporation, February 24, 1987.

E. E. Azar, [1982], *The Codebook of The Conflict and Peace Data Bank*, College Park: Center for International Development, University of Maryland, 1982.

S. Banerjee, [1986a], "Reproduction of Social Structures: An Artificial Intelligence Model," *Journal of Conflict Resolution*, June, 1986, 30(2): 221-252.

S. Banerjee, [1986b], "The Representation of Technology," paper presented at *The 1986 Meeting of the International Studies Association*, March, 1986.

S. Banerjee, [1986c], "The Reproduction of Perceptions and Decisions in The Early Cold War Period," New York: Department of Political Science, Graduate Center and Baruch College, CUNY, mimeo, 1986.

L. W. Barsalou and G. H. Bower, [1984], "Discrimination Nets as Psychological Models," *Cognitive Science*, 1984, 8: 1-26.

W. W. Bartley, III, [1987], "Philosophy of Biology versus Philosophy of Physics," in (Radnitzky & Bartley, 1987: 7-45).

J. Barwise and J. Perry, [1983], *Situations and Attitudes*, Cambridge: M.I.T. Press, 1983.

J. Batali, [1987], *Praxis: An Introspective Theory of Rationality*, Cambridge: Doctoral Dissertation, Department of Electrical Engineering and Computer Science, M.I.T., forthcoming, 1987.

S. Beer, [1959], *Cybernetics and Management*, New York: Wiley, 1959.

S. Beer, [1966], *Decision and Control: The Meaning of Operational Research and Management Cybernetics*, New York: Wiley, 1966.

S. Beer, [1975], *Platform for Change*, New York: Wiley, 1975.

J. P. Bennett, [1978], *Perpetuating Failure: Security Practices and System Transformation of South Pacific Transnational Politics, 1830-1905*, Cambridge: Ph.D. Thesis, Department of Political Science, M.I.T., 1978.

J. P. Bennett, [1987], "Deterrence in Contemporary Policy Debates: The Expression of Strategic Doctrines in Political Arguments about SDI," paper presented at *The 1987 Meeting of the International Studies Associations*, Washington, D. C., April, 1987.

J. P. Bennett and H. R. Alker, Jr., [1977], "When National Security Politics Bred Collective Insecurity: The War of The Pacific in a World Politics Simulation," in K. W. Deutsch, editor, *Problems of World Modeling*, Cambridge: Ballinger, 1977: 215-299.

W. L. Bennett, [1981], "Perception and Cognition: An Information-Processing Framework for Politics," in S. L. Long, editor, *The Handbook of Political Behavior*, volume 1, New York: Plenum Press, 1981: 69-193.

L. P. Bloomfield, [1987], "Nuclear Crisis and Human Frailty," in H. Cleveland and L. P. Bloomfield, editors, *Prospects for Peacemaking: A Citizen's Guide to a Safer Nuclear Strategy*, Cambridge: M.I.T. Press, 1987.

L. P. Bloomfield, [1986], "Analytical Rigor and Real-World Conflicts: The Death and Life of CASCON," paper presented at *The 1986 Meeting of the International Studies Association*, Anaheim, March, 1986.

L. P. Bloomfield, [1982], **The Foreign Policy Process: A Modern Primer**, Englewood Cliffs, NJ: Prentice-Hall, 1982.

L. P. Bloomfield and R. Beattie, [1971], "Computers and Policy-Making: The CASCON Experiment," *Journal of Conflict Resolution*, Fall, 1971.

L. P. Bloomfield and C. J. Gearin, [1969], "Games Foreign Policy Experts Play: The Political Exercise Comes of Age," *Orbis*, Winter 1973: 1008-1031.

L. P. Bloomfield, C. J. Gearin, and J. L. Foster, [1970], *Anticipating Conflict-Control Policies: The CONEX Games as a Planning Tool*, Cambridge: Center for International Studies, M.I.T., 1970.

L. P. Bloomfield and A. Leiss [1969], *Controlling Small Wars: A Strategy for The 1970s*, New York: Knopf, 1969.

D. B. Bobrow, D. A. Sylvan, and B. Ripley, [1986], "Japanese Supply Security: A Computational Model," paper presented at *The 1986 Meeting of The International Studies Association*, Anaheim, March, 1986.

M. A. Boden, [1977], *Artificial Intelligence and Natural Man*, New York: Basic Books, 1977.

M. Bonham and M. J. Shapiro, [1975], editors, *Thought and Action in Foreign Policy*, Basel: Birkhauser Verlag, 1975.

G. S. Boolos and R. C. Jeffrey, [1980], *Computability and Logic*, Cambridge: Cambridge University Press, 1980.

R. J. Brachman and H. J. Levesque, [1984], "The Tractability of Subsumption in Frame-Based Description Languages," *Proceedings of The 1984 National Conference on Artificial Intelligence*, 1984: 34-37.

R. J. Brachman and J. G. Schmolze, [1985], "An Overview of The KL-ONE Knowledge Representation System," *Cognitive Science*, 1985, 69: 171-215.

P. Bracken, [1983], *The Command and Control of Nuclear Forces*, New Haven: Yale University Press, 1983.

S. J. Brams, [1985], *Superpower Games: Applying Game Theory to Superpower Conflicts*, New Haven: Yale University Press, 1985.

R. A. Brooks, [1981], "Symbolic Reasoning Among 3-D Models and 2-D Images," *Artificial*

Intelligence, 1981, 17: 285-348.

R. A. Brooks, [1987], "Intelligence Without Representation," paper presented at *The Workshop on Foundations of Artificial Intelligence*, Endicott House, M.I.T., Cambridge, June, 1987.

L. Brownston, R. Farrell, E. Kant, and N. Martin, [1985], *Programming Expert Systems in OPS5: An Introduction to Rule-Based Programming*, Reading: Addison-Wesley, 1985.

B. G. Buchanan and E. H. Shortliffe, [1984], *Rule-Based Expert Systems*, Reading: Addison-Wesley, 1984.

B. Bueno de Mesquita, [1981], *The War Trap*, New Haven: Yale University Press, 1981.

M. H. Burstein, [1986], "Concept Formation by Incremental Analogical Reasoning and Debugging," in (Michalski, Carbonell, and Mitchell, 1986: 351-370).

D. T. Campbell, [1960], "Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes," in (Radnitzky & Bartley, 1987: 91-114). Originally published in *The Psychological Review*, 1960, 67: 380-400.

D. T. Campbell, [1974], "Evolutionary Epistemology," in (Radnitzky & Bartley, 1987: 47-89). The essay was originally published in P. A. Schilpp, editor, *The Philosophy of Karl Popper*, La Salle: Open Court, 1974, 413-463.

J. G. Carbonell, [1978], "POLITICS: Automated Ideological Reasoning," *Cognitive Science*, 1978, 2: 1-15.

J. G. Carbonell, [1981a], "Counterplanning: A Strategy-Based Model of Adversary Planning in Real-World Situations," *Artificial Intelligence*, 1981, 16: 295-329.

J. G. Carbonell, [1981b], *Subjective Understanding: Computer Models of Belief Systems*, Ann Arbor: UMI Research Press, 1981.

J. G. Carbonell, [1986], "Derivational Analogy: A Theory of Reconstructive Problem Solving and Expertise Acquisition," (Michalski, Carbonell, and Mitchell, 1986: 371-392).

D. Chapman, [1985], David Chapman, "Planning for Conjunctive Goals," *Artificial Intelligence*, July, 1987 32: 333-377. This is a revised version of a master's thesis submitted to the Department of Electrical Engineering and Computer Science, M.I.T. An earlier version was published under the same title as Technical Report No. 802, Artificial Intelligence Laboratory, M.I.T., November, 1985.

N. Chomsky, [1965], *Aspects of a Theory of Syntax*, Cambridge: M.I.T. Press, 1965.

N. Chomsky, [1981], *Lectures on Government and Binding*, Dordrecht: Foris, 1981.

N. Choucri, [1974], *Population Dynamics and International Violence: Propositions, Insights, and Evidence*, Lexington: D. C. Heath, 1974.

N. Choucri and R. C. North, [1975], *Nations in Conflict: National Growth and International Violence*, San Francisco: W. H. Freeman, 1975.

S. J. Cimbala [1987a], editor, *Artificial Intelligence and National Security*, Lexington: Lexington Books, 1987.

S. J. Cimbala [1987b], "Artificial Intelligence and SDI: Corollaries or Compatriots?," in (Cimbala, 1987a: 203-214).

C. Von Clausewitz, [1832], *On War*, New York: Penguin, 1981. Originally published in 1832.

P. R. Cohen and H. J. Levesque, [1985], "Speech Acts and Rationality," *Proceedings of the 1985 Meeting of The Association For Computational Linguistics*, June, 1985: 49-59.

P. R. Cohen and H. J. Levesque, [1987a], "Persistence, Intention, and Commitment," Menlo Park: SRI International, Technical Note 415, February, 1987.

P. R. Cohen and H. J. Levesque, [1987b], "Intention = Choice + Commitment," *Proceedings of The 1987 National Conference on Artificial Intelligence*, 1987: 410-415.

R. Cohen, [1983], *A Computational Model For The Analysis of Arguments*, Toronto: Doctoral Dissertation, Computer Systems Research Group, University of Toronto, Technical Report No. CSRG-151, October, 1983.

R. Cohen, [1987a], "Interpreting Clues in Conjunction With Processing Restrictions in Arguments and Discourse," *Proceedings of The 1987 National Conference on Artificial Intelligence*, 1987: 528-533.

R. Cohen, [1987b], "Analyzing the Structure of Argumentative Discourse," *Computational Linguistics*, 1987.

T. D. Cook and D. T. Campbell, [1979], *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Chicago: Rand-McNally, 1979.

R. N. Cooper, [1968], *The Economics of Interdependence: Economic Policy in the Atlantic Community*, New York: McGraw-Hill, 1968.

CACI, [1979], *Analysis of Chinese Crisis Management Experience: Summary Report*, Arlington, VA: Consolidated Analysis Centers, Inc. (CACI), 1979.

J. P. Crecine, [1969], *Governmental Problem Solving: A Computer Simulation of Municipal Budgeting*, Chicago: Rand McNally, 1969.

R. Cyert & J. March, [1963], *A Behavioral Theory of the Firm*, Englewood Cliffs: Prentice-Hall, 1963.

F. R. Dallmayr, [1984], *Language and Politics: Why Does Language Matter to Political Philosophy?*, Notre Dame: University of Notre Dame Press, 1984.

P. K. Davis, [1987], "A New Analytic Technique for The study of Deterrence, Escalation, Control, and War Termination," in (Cimbala, 1987a: 35-60).

P. K. Davis, S. C. Bankes, and J. P. Kahan, [1986], *A New Methodology for Modeling National Command Level Decision-Making in War Games and Simulations*, Santa Monica: Strategy Assessment Center, RAND Corporation, July, 1986.

P. Diesing, [1962], *Reason in Society: Five Types of Decisions and Their Social Conditions*, Westport, CT: Greenwood Press, 1976. Originally published in 1962.

T. G. Dietterich and R. S. Michalski, [1983], "A Comparative Overview of Selected Methods For Learning from Examples," in (Michalski, Carbonell & Mitchell, 1983: 41-82).

K. W. Deutsch, [1953], *Nationalism and Social Communication: Am Inquiry into the Foundations of Nationality*, second edition, Cambridge: M.I.T. Press, 1966. The first edition was originally published in 1953.

K. W. Deutsch, [1963], *The Nerves of Government*, New York: Free Press, 1963.

E. A. Devereux, [1985a], "Processing Political Debate: A Methodology for Data Production with Special Application to the Lincoln-Douglas Debates," Cambridge: Bachelor's thesis, Department of Political Science, M.I.T., June, 1985.

E. A. Devereux, [1985b], "Game Strategy as Argument: Modeling Generative Role Logics in Dramaturgical SPD Data," a revised version of a paper presented at *The 1985 Meeting of the American Political Science Association*, New Orleans, September, 1985.

G. Duffy, [1987], *Language, Politics, and Method: Grounding a Computational Hermeneutic*, Cambridge: Doctoral Dissertation, Department of Political Science, M.I.T., August, 1987.

G. Duffy and J. C. Mallery, [1984], "Referential Determinism and Computational Efficiency: Posting Constraints from Deep Structure," *Proceedings of the 1984 National Conference on Artificial Intelligence*, August, 1984, 101-105.

G. Duffy and J. C. Mallery, [1986], "RELATUS: An Artificial Intelligence Tool for Natural Language Modeling," paper presented to the Panel on "Artificial Intelligence Models of International Politics" at *The 1986 Meeting of The International Studies Association*, Anaheim, March, 1986.

M. G. Dyer, [1983], *In-Depth Understanding*, Cambridge: M.I.T. Press, 1983.

H. Eckstein, [1975], "Case Study and Theory in Political Science," in (Greenstein & Polsby, 1975: 79-137).

F. H. van Eemeren and R. Grootendorst, [1984], *Speech Acts in Argumentative Discussions: A Theoretical Model for The Analysis of Discussion Directed Towards Solving Conflict of Opinion*, Dordrecht: Foris, 1984.

F. Emery and E. Trist,[1974], *Futures We're In*, Canberra: National University of Australia, 1974.

J. Emshoff and R. Ackoff, [1970], "Explanatory Models of Interactive Choice Behavior," *Journal of Conflict Resolution*, 1970, 14: 77-89.

M. M. Ensign, [1985a], "Artificial Intelligence and International Relations: A Review of Recent Research," College Park: Political Science Department, University of Maryland, mimeo, 1985.

M. M. Ensign, [1985b], *An Analysis of Private Bank Loans to Developing Countries: The Relationship Between Bankers' Images and Their Lending Policies*, New York: Gordon and Breech, 1985.

M. M. Ensign, [1986], "Artificial Intelligence-Based Simulations in International Relations," Paper presented at *The 1986 Meeting of The International Studies Association*, Anaheim, March, 1986.

L. S. Etheredge, [1985], *Can Governments Learn? American Foreign Policy and the Central American revolutions*, New York: Pergamon Press, 1985.

B. Falkenhainer, K. D. Forbus, D. Gentner, [1986], "The Structure Mapping Engine," *Proceedings of the 1986 National Conference on Artificial Intelligence*, 1986: 272-277.

B. Fawcett, [1986], "The Representation of Ambiguity in Opaque Contexts," Toronto: Computer Systems Research Institute, T.R. CSRI-178, April, 1986.

E. A. Feigenbaum and H. A. Simon, [1984], "EPAM-like Models of Recognition and Learning," *Cognitive Science*, 1984, 8: 305-336.

C. F. Flores, [1981], *Management and Communication in the Office of the Future*, Berkeley: Doctoral Dissertation, University of California at Berkeley, 1981.

M. Flowers, R. McGuire, L. Birnbaum, [1982], "Adversary Arguments and the Logic of Personal Attacks," in (Lehnert & Ringle, 1982: 275-294).

J. A. Fodor, M. F. Garrett, E. C. T. Walker, and C. H. Parkes, [1980], "Against Definitions," *Cognition*, 1980, 8: 263-367.

J. W. Forrester, [1971], *World Dynamics*, Cambridge: Wright-Allen Press, 1971.

K. S. Fu, [1974], *Syntactic Methods in Pattern Recognition*, New York: Academic Press, 1974.

K. S. Fu, [1982], *Syntactic Pattern Recognition and Applications*, Englewood Cliffs: Prentice-Hall, 1982.

H. G. Gadamer, [1960], *Truth and Method*, New York: Continuum, 1975. Originally published as *Wahrheit und Methode*, Tubingen: 1960.

H. Gardner, [1985], *The Mind's New Science: A History of the Cognitive Revolution*, New York: Basic Books, 1985.

A. L. Gardner, [1984], *An Artificial Intelligence Approach to Legal Reasoning*, Palo Alto:

Doctoral Dissertation, Department of Computer Science, Stanford University, 1984.

A. L. George, [1959], *Propaganda Analysis: A Study of Inferences Made from Nazi Propaganda in World War II*, White Plains, NY: Row, Peterson, & Co., 1959.

A. L. George, [1969], "The Operational Code: A Neglected Approach to The Study of International Political Leaders and Decision-Making," *International Studies Quarterly*, 1969, 13: 190-222.

A. L. George, [1979], "Case Studies and Theory Development: The Method of Structured Focused Comparisons," in P. G. Lauren, editor, *Diplomacy: New Approaches in History, Theory and Policy*, New York: Free Press, 1979: 43-68.

C. Geertz, [1973], *Interpretation of Cultures*, New York: Basic Books, 1973.

D. Gentner, [1983], "Structure Mapping: A Theoretical Framework for Analogy," *Cognitive Science*, 7: 155-170.

D. Gentner and R. Landers, [1985], "Analogical Reminding: A Good Match is Hard to Find," paper presented to the panel on "Common-sense Reasoning" at *The International Conference on Systems, Man and Cybernetics*, Tucson, November, 1985.

P. J. Giers, [1986], "The Sylvan-Majeski "Rostow" Model: An Organizational Perspective," Department of Political Science, M.I.T., mimeo, May, 1986.

G. N. Gilbert & C. Heath, [1985], *Social Action and Artificial Intelligence: Surrey Conferences on Sociological Theory*, Brookfield, Vermont: Gower, 1985.

T. Gilovich, [1981], "Seeing the Past in The Present: The Effect of Associations to Familiar Events on Judgements and Decisions," *Journal of Personality and Social Psychology*, 1981, 40: 797-808.

D. Goodman, [1987], *The Complete Hypercard Handbook*, New York: Bantam Books, 1987.

F. Greenstein and N. Polsby, [1975], editors, *Handbook of Political Science, volume 7*, Reading: Addison-Wesley, 1975.

W. F. Grunbaum, [1986], "Using Artificial Intelligence Techniques to Predict Supreme Court Decision Making," paper presented at *The 1986 Meeting of The American Political Science Association*, Washington, D.C., 1986.

H. Guetzkow, *et al.*, **[1963]**, *Simulation in International Relations*, Englewood Cliffs, NJ: Prentice-Hall, 1963.

K. W. Haas, [1987], *Typical: An Implemented Approach to Type Specification and Inference*, Cambridge: Artificial Intelligence Laboratory, M.I.T., Technical Report, 1987.

J. Habermas, [1979], "What is Universal Pragmatics?," in J. Habermas, *Communication and The Evolution of Society*, T. McCarthy, translator, Boston: Beacon Press, 1979.

J. Habermas, [1981], *The Theory of Communicative Action: Reason and the Rationalization of Society*, Thomas McCarthy, translator, Boston: Beacon Press, 1981. German edition published in 1981.

J. Handle and T. S. Whitaker, [1983], *Zmail Concepts and Techniques*, Cambridge: Symbolics, Inc, 1983.

R. Harre, [1985], *Personal Being*, Cambridge: Harvard University Press, 1985.

D. G. Hays, [1969], 'Linguistic Foundations for a Theory of Content Analysis," in G. Gerbner, O. R. Holsti, K. Krippendorf, W. J. Paisley, and P. J. Stone, editors, *The Analysis of Communication Content*, New York: Wiley, 1969: 57-67.

F. Hayes-Roth, [1987], "Expert Systems," in (Shapiro, 1987: 287-298).

F. Hayes-Roth, D. A. Waterman, and D. B. Lenat, [1983], *Building Expert Systems*, Reading: Addison-Wesley, 1983.

F. Heider, [1946], "Attitudes and Cognitive Organization," *Journal of Psychology*, 1946, 21: 107-112.

M. Heidegger, [1927], *Being and Time*, J. Macqarrie and E. Robinson, translators, New York: Harper and Row, 1962. Originally published as *Sein und Zeit*, Tubingen: Neomarius Verlag, 1927.

C. F. Hermann, W. R. Phillips, and S. J. Thorson, [1978], "Validating International Relations Forecasts to Develop Theory," in N. Choucri & T. W. Robinson, editor, *Forecasting in International Relations: Theory, Methods, Problems, Prospects*, San Francisco: W. H. Freeman, 1978: 69-78.

C. Hewitt, [1986], "Offices Are Open Systems," *Proceedings of The 1986 ACM Conference on Office Information Systems*, Providence, RI, October 6-8, 1986.

D. Hillis, [1986], *The Connection Machine*, Cambridge: M.I.T. Press, 1986.

G. E. Hinton, T. J. Sejnowski, and D. H. Ackley, [1984], "Boltzmann Machines: Constraint Satisfaction Networks that Learn," Pittsburg: Department of Computer Science, Carnegie Mellon University, Technical Report Number CMU-CS84-119, 1984.

J. H. Holland, [1975], *Adaptation in Natural and Artificial Systems*, Ann Arbor: University of Michigan Press, 1975.

J. H. Holland, [1986], "Escaping Brittleness: The Possibilities of General Purpose Algorithms Applied to Parallel Rule-Based Systems," in (Michalski, Carbonell, & Mitchell, 1986: 593-624).

O. Holsti, [1968], "Content Analysis," in G. Lindzey and E. Aronson, editors, *The Handbook of Social Psychology, Volume 2*, Reading: Addison-Wesley, 1968.

J. E. Hopcroft and J. D. Ullman, [1979], *Introduction to Automata Theory, Languages, and Computation*, Reading: Addison-Wesley, 1979.

R. Hurwitz, [1987], "The Fiction in the Prisoner's Dilemma," paper presented at *The 1987 Meeting of the International Studies Association*, Washington, D. C., April, 1987.

R. Hurwitz, J. C. Mallery, H. R. Alker, Jr., and G. Duffy, [1986], "Anarchy or Community: A Study of Developmental Patterns in SPD Games," paper presented to the Panel on "Artificial Intelligence Models of International Politics" at *The 1986 Meeting of The International Studies Association*, Anaheim, March, 1986.

R. Hurwitz, J. C. Mallery, H. R. Alker, Jr., and G. Duffy, [1987], "Analyzing Natural Language Protocols of Simulated Strategic and Social Interactions," *Journal of Applied Artificial Intelligence*, forthcoming.

I. L. Janis and L. Mann, [1977], *Decision Making*, New York: The Free Press, 1977.

R. Jervis, [1976], *Perception and Misperception in International Politics*, Princeton: Princeton University Press, 1976.

B. L. Job & D. Johnson, [1986], "A Model of US Foreign Policy Decision Making: The US and The Dominican Republic, 1961-1965," paper presented at *The 1986 Meeting of The International Studies Association*, Anaheim, March, 1986.

D. Kahneman, P. Slovic, and A. Tversky, [1982], editors, *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press, 1982.

J. J. Katz and J. A. Fodor, [1963], "The Structure of a Semantic Theory," *Language*, 1963, 39 (2): 170-210.

R. L. Keeny and H. Raiffa, [1976], *Decisions With Multiple Objectives: Preferences and Value Tradeoffs*, New York: Wiley, 1976.

R. O. Keohane, [1984], *Hegemony: Cooperation and Discord in The World Political Economy*, Princeton: Princeton University Press, 1984.

R. O. Keohane and J. S. Nye, [1977], *Power and Interdependence: World Politics in Transition*, Boston: Little, Brown, 1977.

J. Klockner and W. L. Bennett, [1987], "Inferring Belief System Structure and Operation from Resource Planning Requirements," paper presented to the panel on "Computing Strategic Rationality: Structural Models of Practical Belief" at the at *The 1987 Meeting of The International Studies Association*, April, 1987.

D. A. Knuth, [1983], "The WEB System of Structured Documentation," Stanford: Department of Computer Science, Stanford University, 1983.

L. Kohlberg, [1969], *Stages in the Development of Moral Thought and Action*, New York: Holt, Rinehart, and Winston, 1969.

J. L. Kolodner, [1980], *Organization and Retrieval in a Conceptual Memory*, Ph.D. Thesis, Department of Computer Science, Yale University, 1980. Published as CS Technical Report, No. 187.

J. L. Kolodner, [1983a], "Maintaining Organization is a Dynamic Long-Term Memory," *Cognitive Science*, 1983, 7: 243-280.

J. L. Kolodner, [1983b], "Reconstructive Memory: A Computer Model," *Cognitive Science*, 1983a, 7: 280-328.

K. Krippendorff, [1980], *Content Analysis: An Introduction to its Methodology*, Beverly Hills: Sage, 1980.

T. S. Kuhn, [1962], *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press, 1962.

J. E. Laird, A. Newell, and P. S. Rosenbloom, [1987], "SOAR: An Architecture for General Intelligence," *Artificial Intelligence*, 1987, 33: 1-64.

H. D. Lasswell, N. Leites, *et al.*, [1949], *Language of Politics: Studies of Quantitative Semantics*, Cambridge: M.I.T. Press, 1949.

H. D. Lasswell, D. Lerner, & I. D. S. Pool, [1952], *The Comparative Study of Symbols*, Stanford: Stanford University Press, 1952.

R. R. Lau and D. O. Sears, [1986], *Political Cognition: The 19th Annual Carnegie Symposium on Cognition*, Hillsdale: Lawrence Erlbaum, 1986.

M. Lebowitz, [1986], "Concept Learning in a Rich Input Domain: Generalization-based Memory," in (Michalski, Carbonell, & Mitchell, 1986: 193-214).

W. G. Lehnert, [1981], "Plot Units and Narrative Summarization," *Cognitive Science*, 1981 4:293-331.

W. G. Lehnert, [1982], "Plot Units: A Narrative Summarization Strategy," in (Lehnert & Ringle, 1982: 375-412).

W. G. Lehnert and C. L. Loiselle, [1985], "Plot Unit Recognition for Narratives," in (Tonfoni, 1985: 9-47).

W. G. Lehnert and M. H. Ringle, [1982], editors, *Strategies for Natural Language Processing*, Hillsdale: Lawrence Erlbaum, 1982.

N. Leites, [1951], *The Operational Code of The Politburo*, New York: McGraw-Hill, 1951.

D. B. Lenat, [1983], "Eurisko: A Program that Learns New Heuristics and Domain Concepts," *Artificial Intelligence*, March, 1983: 61-98.

D. B. Lenat, A. Clarkson, [1986], "Artificial Intelligence and C^3I," *Signal*, June, 1986: 115-119.

D. B. Lenat, A. Clarkson, and G. Kiremidjian, [1983], "An Expert System for indications & Warning Analysis," *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, West Germany, August, 1983, 259-262.

D. B. Lenat, M. Prakash, and M. Shepherd, [1986], "CYC: Using Common-Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks," *The AI Magazine*, Winter, 1986, 65-85.

H. R. Lewis and C. H. Papadimitriou, [1981], *Elements of The Theory of Computation*, Englewood Cliffs: Prentice-Hall, 1981.

J. Loevinger, [1976], *Ego Development: Conceptions and Theories*, San Francisco: Josey-Bass, 1976.

S. G. Levinson, [1983], *Pragmatics*, Cambridge: Cambridge University Press, 1983.

C. E. Lindbloom, [1959], "The Science of Muddling Through," *Public Administration Review*, 1959, 19: 79-88.

C. E. Lindbloom, [1965], *The Intelligence of Democracy*, New York: The Free Press, 1965.

A. R. Luria, [1974], *Cognitive Development: Its Cultural and Social Foundations*, M. Lopez-Morillas and L. Solotaroff, translators, M. Cole, editor, Cambridge: Harvard University Press, 1976. First published in Russian as *Ob istoricheskom razvitii poznavatel'nykh protsessov*, Moscow: Nauka, 1974.

U. Luterbacher and M. D. Ward [1985], editors, *Dynamic Models of International Conflict*, Boulder: Lynne Rienner, 1985.

S. L. Lytinen, [1984], *The Organization of Knowledge In a Multi-lingual, Integrated Parser*, New Haven: Doctoral Dissertation, Department of Computer Science, Yale University, Report No. YALE/CSD/RR-340, November, 1984.

S. J. Majeski, [1985], "Recommendation Heuristics for Intervention: Korea in 1950 and Vietnam in 1961, paper presented at *The Annual Meeting Of The American Political Science Association*, New Orleans, 1985.

S. J. Majeski, [1987], "A Recommendation Model of War Initiation: The Plausibility and Generalizability of General Cultural Rules," in (Cimbala, 1987a: 61-86.

S. J. Majeski and D. J. Sylvan, [1985], "Decision and Recommendation Heuristics for War Involvement: The 1961 Vietnam Commitments, A Progress Report" paper presented at *The 1985 Meeting of the International Studies Association*, 1985.

J. C. Mallery, [1987a], "Constraint-Interpreting Reference," Cambridge: Artificial Intelligence Laboratory, M.I.T., AI Memo No. 827, 1987.

J. C. Mallery, [1987b], "Computing Strategic Language: Natural Language Models of Belief and Intention," paper presented to the panel on "Computing Strategic Rationality: Structural Models of Practical Belief" at the at *The 1987 Meeting of The International Studies Association*, Washington, D. C., April, 1987.

J. C. Mallery, [1987c], "Overview of the RELATUS Natural Language System," Cambridge: M.I.T. AI Laboratory, mimeo, February, 1987.

J. C. Mallery and G. Duffy, [1987] "A Computational Model of Semantic Perception," Cambridge: Artificial Intelligence Laboratory, M.I.T., AI Memo No. 799, 1987.

J. C. Mallery and R. Hurwitz, [1987], "Analogy and Precedent in Strategic Decision-Making: A Computational Approach," paper presented at *The 1987 Meeting of the American Political Science Association*, Chicago, September, 1987.

J. C. Mallery, R. Hurwitz, and G. Duffy, [1987a], "Hermeneutics," in (Shapiro, 1987: 362-376).

J. C. Mallery, R. Hurwitz, with H. R. Alker, Jr., and G. Duffy, [1986], "Analyzing Natural Language Protocols of Sequential PD Games," paper presented to the Microcomputer Group at *The 1986 Meeting of the American Political Science Association*," Washington, D.C., August, 1986.

J. March & H. Simon, [1958], *Organizations*, New York: John Wiley, 1958.

C. Marshall, [1986], *Notecards Release 1.2 Reference Manual*, Palo Alto: Xerox Special Information Systems, Intelligent Systems Laboratory, Xerox Palo Alto Research Center, April 3, 1986.

H. R. Maturana, [1970], "Biology of Cognition," in (Maturana & Varela, 1980: 2-62).

H. R. Maturana, [1977], "Biology of Knowledge," in R. W. Reiber, editor, *The Neurophysiology of Language*, New York: Plenum Press, 1977.

H. R. Maturana and F. Varela, [1980], *Autopoeisis and Cognition: The Realization of the Living*, Dordrecht: Reidel, 1980.

E. May, [1973], *"Lessons" of the Past*, New York: Oxford University Press, 1973.

J. M. McCarthy and P. Hayes, [1969], "Some Philosophical Problems from The Standpoint of Artificial Intelligence," *Machine Intelligence*, 1969, 4: 463-502.

L. T. McCarthy, [1977], "Reflections on TAXMAN: An Experiment in Artificial Intelligence and Legal Reasoning," *Harvard Law Review*, 1977, 90: 837-893.

L. T. McCarthy, N. S. Sridharan & B. C. Sangster, [1979], *The Implementation of TAXMAN II: An Experiment in Artificial Intelligence and Legal Reasoning*, Laboratory for Computer Science Research, LRP-TR-2, January, 1979.

P. McCorduck, [1979], *Machines Who Think*, San Francisco: W. H. Freeman, 1979.

D. V. McDermott,[1982], "A Temporal Logic for Reasoning about Processes and Plans," *Cognitive Science*, 1982, 6: 101-155.

D. H. Meadows, D. L. Meadows, J. Randers, and W. K. Behrens, [1972], *The Limits to Growth*, New York: Universe Books, 1972.

D. Mefford, [1979a], "From Static to Transformational Models of Political Reasoning," paper presented to the panel on "International Learning and Non-Learning," within the section on “Quantitative Mathematic Approaches to Politics,” at *The XIth triennial World Congress of the International Political Science Association*, Moscow, August, 1979.

D. Mefford, [1979b], “The Grammar of Change in Political Beliefs: A Constructive Critique of Cognitive Modeling,” paper presented to the panel on “International Learning and Non-Learning,” within the section on “Quantitative Mathematic Approaches to Politics,” at *The XIth Triennial World Congress of the International Political Science Association*, Moscow, August, 1979.

D. Mefford, [1979c], “Simulating The Practical Logic of Deterrence: An Application for Artificial Intelligence,” paper presented at *The 1979 Meeting of the International Studies Association*, March, 1979.

D. Mefford [1984], “Formulating Foreign Policy on The Basis of Historical Analogies: An Application of Developments in Artificial Intelligence,” paper presented at *The 1984 Meeting of The International Studies Association*, March, 1984.

D. Mefford [1985], “Changes in Foreign Policy Across Time: The Logical Analysis of a Succession of Decision Problems Using Logic Programming,” in (Luterbacher and Ward, 1985: 401-423). An earlier version of the paper was presented at *The Conference on Dynamical Models of International Conflict*, Boulder, CO, October 31 to November, 3, 1984.

D. Mefford, [1986a], “Analogical Reasoning and the Definition of the Situation: Back to Snyder for Concepts and Forward to Artificial Intelligence for Method,” in C. F. Hermann, C. W. Kegley, Jr., and J. N. Rosenau, editors, *New Directions in the Comparative Study of Foreign Policy*, New York: Allen and Unwin, 1986.

D. Mefford, [1986b], “Using Political Narratives to Structure Decisions and Games: The Design for an Expert System Shell,” paper presented at *The 1986 Meeting of the International Studies Association*, Anaheim, March, 1986.

D. Mefford, [1986c], “Proposing Models of Evidence While Constructing Scenarios,” Los Angeles: Social Science Research Institute, University of Southern California, 1986.

D. Mefford, [1986d], “What Morton Kaplan Wanted To Do, And Should Have Done: Balance of Power as a Rule-Based System,” paper presented at *The 1986 Meeting of the International Studies Association*, Anaheim, March, 1986.

J. A. Meldman, [1975], “A Preliminary Study in Computer Aided Legal Analysis,” Cambridge: Project MAC, M.I.T., Technical Report No. MAC-TR-157, 1975.

J. A. Meldman, [1977], “A Structural Model for Computer Aided Legal Analyst,” *Rutgers Journal of Computers and Law*, 1977, 6(27).

M. Mesarovic and E. Prestel, [1974], *Mankind at The Turning Point*, New York: Dutton, 1974.

R. S. Michalski, [1987], “Concept Learning,” in (Shapiro, 1987: 185-194).

R. S. Michalski, J. G. Carbonell, T. M. Mitchell, [1983], editors, *Machine Learning: An Artificial Intelligence Approach*, Palo Alto: Tioga, 1983.

R. S. Michalski, J. G. Carbonell, T. M. Mitchell, [1986], *Machine Learning: An Artificial Intelligence Approach, Volume II*, Los Altos: Morgan Kaufman, 1986.

M. Minsky, [1975], "A Framework for Representing Knowledge," in (Winston, 1975).

M. Minsky, [1985], "Communication with Alien Intelligence," *Byte: The Small Systems Journal*, April, 1985, 10(4): 127-138.

M. Minsky, [1987], *The Society of Mind*, New York: Simon and Schuster, 1987.

T.M. Mitchell, R.M. Keller, S.T. Kedar-Cabelli, [1985], "Explanation-Based Learning: A Unifying View," New Brunswick: Department of Computer Science, Rutgers University, Technical Report No. ML-TR-2, 1985.

E. T. Mueller, [1987], *Daydreaming and Computation: A Computer Model of Everyday Creativity, Learning, and Emotions in the Human Stream of Thought*, Los Angeles: Artificial Intelligence Laboratory, University of California, Los Angeles, Technical Report No. UCLA-AI-87-8, February, 1987.

J. Z. Namenwirth, [1973], "Wheels of Time and the Interdependence of Value Change in America," *Journal of Interdisciplinary History*, 1973, 4: 649-683.

J. Z. Namenwirth and R. P. Weber, [1986], *Culture Indicators Research: A Content Analytic Perspective*, forthcoming, 1986.

A. Newell, [1980], "Physical Symbol Systems," *Cognitive Science*, 1980, 4: 135-183.

A. Newell, [1987], "Foundations of Cognitive Systems," William James Lecture at Harvard University, Cambridge, February 25, 1987.

A. Newell, J. C. Shaw, and H. A. Simon, [1957], "Preliminary Description of a General Problem Solving Program-I (GPS-I)," Pittsburg: Carnegie Institute of Technology, CIP Working Paper No. 7, 1957.

A. Newell, J. C. Shaw, and H. A. Simon, [1959], *Report on a General Problem Solving Program*, Santa Monica: RAND Corporation, Report No. P1584, January, 1959.

A. Newell and H. A. Simon, [1972], *Human Problem Solving*, Englewood Cliffs, NJ: Prentice-Hall, 1972.

G. O'Donnell, [1973], *Bureaucratic Authoritarianism and Modernization*, Berkeley: University of California Press, 1973.

M. Olson, [1965], *The Logic of Collective Action: Public Goods and The Theory of Groups*, Cambridge: Harvard University Press, 1965.

N. G. Onuf, [1985], "Do Rules Say What They Do? From Ordinary Language Philosophy to International Law," *Harvard International Law Journal*, Spring, 1985, 26(2): 385-410.

N. G. Onuf, [1987], "Rules in Sociomoral Development," *Human Development*, forthcoming, 1987.

R. Pfeifer, [1982], *Cognition and Emotion: An Information Processing Approach*, Pittsburg: Department of Psychology, Carnegie-Mellon University, CIP Working Paper No. 436, 1982.

J. Piaget, [1972], "Intellectual Evolution from Adolescence to Adulthood," *Human Development*, 1972, 15: 1-12.

J. Piaget, [1977], *The Development of Thought: Equilibration of Cognitive Structures*, New York: Viking, 1977.

W. R. Phillips, [1987], "Alternative Futures in the Middle East: The Results from Three Simulations," in (Cimbala, 1987a: 161-186).

W. R. Phillips and M. M. Ensign, [1982], "Self-Reliance and Political Development: Simulations of Inner Environments," *International Political Science Review*, 1982, 3(4): 455-478.

K. M. Pitman, [1985], "CREF: An Editing Facility for Managing Structured Text," Cambridge: Artificial Intelligence Laboratory, M.I.T., AI Memo Number 829, February, 1985.

D. Pruitt and K. M. Kimmel, [1977], "Twenty Years of Experimental Gaming: Critique, Synthesis, and Suggestions for The Future," *Annual Review of Psychology*, 1977, 28: 363-392.

M. Polanyi, [1958], *Personal Knowledge*, Chicago: University of Chicago Press, 1958.

I. S. Pool, [1959], editor, *Trends in Content Analysis*, Urbana: University of Illinois Press, 1959.

K. R. Popper, [1974], "Campbell on the Evolutionary Theory of Knowledge," in (Radnitzky & Bartley, 1987: 115-120). Originally published in P. A. Schlipp, editor, *The Philosophy of Karl Popper*, La Salle: Open Court, 1974.

K. R. Popper, [1977], "Natural Selection and the Emergence of Mind," (Radnitzky & Bartley, 1987: 139-153). Originally delivered as the first Darwin Lecture at Darwin College, Cambridge, on November 8th, 1977. The lecture was published in *Dialectica*, 1978, 22(3): 339-55.

G. Radnitzky and W. W. Bartley, III, [1987], editors, *Evolutionary Epistemology, Theory of Rationality, and the Sociology of Knowledge*, La Salle: Open Court, 1987.

H. Raiffa, [1970], *Decision Analysis*, 2nd edition, Reading: Addison-Wesley, 1970.

N. Rescher, [1977], *Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge*, Albany: State University of New York Press, 1977.

N. Rescher, [1978], *Peirce's Philosophy of Science: Critical Studies in His Theory of Induction and Scientific Method*, Notre Dame: University of Notre Dame Press, 1978.

L. F. Richardson, [1960], *Statistics of Deadly Quarrels*, Pittsburg: Boxwood, 1960.

P. Ricoeur, [1971], "The Model of Text: Meaningful Action Considered as Text," *Social*

Research, 1971, 38: 529-62. Reprinted in J. B. Thompson, editor and translator, *Paul Ricoeur: Hermeneutics and The Human Sciences*, Cambridge: Cambridge University Press, 1981.

P. Ricoeur, [1973], "Creativity in Language," *Philosophy Today*, (1973), 17: 97-111. Reprinted in C. E. Reagan and D. Stewart, editors, *The Philosophy of Paul Ricoeur: An Anthology of His Work*, Boston: Beacon Press, 1978: 120-133.

E. L. Rissland and K. D. Ashley, [1987], "A Case-Based System for Trade Secrets Law" *Proceedings of The First International Conference on Artificial Intelligence and Law, Boston, MA, May 27-29, 1987,*, New York: Association for Computing Machinery, 1987: 60-66.

C. Reiger, [1975], "The Common-Sense Algorithm as a Basis for Computer Models of Human Memory, Inference, Belief Systems, and Contextual Language Comprehension," College Park: Department of Computer Science, University of Maryland, Technical Report No. TR-373, 1975.

C. Reiger, [1976], "The Causal Representation and Simulation of Physical Mechanisms," College Park: Department of Computer Science, University of Maryland, Technical Report No. TR-495, 1976.

R. C. Schank, [1972], "Conceptual Dependency: A Theory Of Natural Language," *Cognitive Psychology*, 3, 552-631 (1972).

R. C. Schank, [1982], *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*, Cambridge: Cambridge University Press, 1982.

R. C. Schank and R.P. Abelson, [1977], *Scripts, Plans, Goals, and Understanding*, Hillsdale: Lawrence Erlbaum, 1977.

P. A. Schrodt, [1984a], "Artificial Intelligence and International Crisis: The Application of Pattern Recognition to The Analysis of Event Sequences," paper presented at *The 1985 Meeting of the American Political Science Association*, August, 1984.

P. A. Schrodt, [1984b], "Artificial Intelligence and The State of Mathematical Modeling in International Relations: A Survey," paper delivered at *The U.S. and Swiss National Science Foundation Conference on Dynamic Models of International Conflict*, Boulder, October 31 to November 3, 1984.

P. A. Schrodt, [1985], "Adaptive Precedent-based Logic and Rational Choice: A Comparison of Two Approaches to The Modeling of International Behavior," in (Luterbacher and Ward, 1985: 373-400).

P. A. Schrodt, [1986], "Set Prediction of International Behavior Using a Holland Classifier," paper presented at *The 1986 Meeting of The International Studies Association*, Anaheim, March, 1986.

P. A. Schrodt, [1987], "Pattern Matching, Set Prediction, and Foreign Policy Analysis," in (Cimbala, 1987a: 89-107).

W. Schwabe and L. M. Jamison, [1982], *A Rule-based Policy-Level Model of Non-superpower Behavior in Strategic Conflicts*, Santa Monica: RAND Corporation, Report No. R-2962-DNA, 1982.

J. R. Searle, [1969], "Speech Acts: An Essay in the Philosophy of Language," Cambridge: Cambridge University Press, 1969.

J. R. Searle, [1983], *Intentionality: An Essay in The Philosophy of Mind*, Cambridge: Cambridge University Press, 1983.

R. Sedgewick, [1983], *Algorithms*, Reading: Addison-Wesley, 1983.

V. M. Sergeev, [1987a], "Advanced Computer Technology and Strategic Stability Problems," Stockholm: SIPRI, mimeo, 1987. Forthcoming in a SIPRI conference proceedings.

V. M. Sergeev, [1987b], "The Types of Conceptual Interpretation of Political Processes," in *Models of Dialogue in Systems of Artificial Intelligence*, *Acta Comintationes Universitatis Tartuensis*, Tartu, 1987.

M. J. Shapiro, [1984], *Language and Politics*, New York: New York University Press, 1984.

M. J. Shapiro, [1988], *The Politics of Representation*, Madison: University of Wisconsin Press, 1988.

S. C. Shapiro, [1987], *The Encyclopedia of Artificial Intelligence*, New York: John Wiley, 1987.

Y. Shoham, [1986], *Reasoning about Change: Time and Causation from The Standpoint of Artificial Intelligence*, New Haven: Computer Science Department, Yale University, 1986.

H. Simon, [1957], *Models of Men: Social and Rational*, New York: Wiley, 1957.

H. A. Simon, [1962], "The Architecture of Complexity," *Proceedings of The American Philosophical Society*, December, 1962, 106:467-482. Reprinted in (Simon, 1969).

H. A. Simon, [1969], *The Sciences of The Artificial*, second edition, Cambridge: M.I.T. Press, 1981. Originally published in 1969.

H. A. Simon, [1982], *Models of Bounded Rationality: Behavioral Economics and Business Organization*, Cambridge: M.I.T. Press, 1982.

H. A. Simon, [1985], "Human Nature in Politics: The Dialogue of Psychology with Political Science," *The American Political Science Review*, 1985, 79: 293-304.

T. Skocpol, [1979], *States and Social Revolutions*, Cambridge: Cambridge University Press, 1979.

C. Stanfill and B. Kahle, [1986], "Parallel Free Text Search on The Connection Machine," *CACM*, December, 1986.

C. Stanfill and D. Waltz, [1986],"Memory-Based Reasoning," Cambridge: Thinking Machines Corporation, Technical Report No. 86-7, March, 1986.

J. D. Steinbruner, [1974], *The Cybernetic Theory of Decision: New Dimensions of Political Analysis*, Princeton: Princeton University Press, 1974.

P. J. Stone, D. C. Dunphy, M. S. Smith, D. M. Ogilvie, with associates, [1966], *The General Inquirer: A Computer Approach to Content Analysis*, Cambridge: M.I.T. Press, 1966.

W. B. Schwartz, R. S. Patil, & P. Szolovits, [1987], "Artificial Intelligence in Medicine: Where do We Stand?," *New England Journal of Medicine*, March 12, 1987.

R. Stallman, [1981], "EMACS: The Extensible, Customizable, Self-Documenting Display Editor," Cambridge: Artificial Intelligence Laboratory, M.I.T., AI Memo Number 519a, March, 1981.

M. E. Stickel, [1985], "An Introduction to Automated Deduction," in G. Goos and J. Hartmanis, editors, *Fundamentals of Artificial Intelligence: An Advanced Course*, New York: Springer-Verlag, 1985: 75-133.

D. J. Sylvan and S. J. Majeski, [1983], "A Formalization of Decision and Measurement Heuristics for War Initiation: The 1961 Vietnam Commitments," paper presented at *The 1983 Meeting Of The American Political Science Association*, 1983.

D. A. Sylvan, [1987a], "Political Limits in The Energy Area," in (Cimbala, 1987a: 189-201.

D. A. Sylvan, [1987b], "Supplementing Global Models With Computational Models: An Assessment and an Energy Example," *Behavioral Science*, July, 1987, 32(3): 212-231.

D. A. Sylvan and S. Chan, [1984], editors, *Foreign Policy Decision Making*, New York: Praeger, 1984.

D. A. Sylvan, B. Ripley, D. B. Bobrow, [1986], "Economic Security as Managed Interdependence: A Computational Model of Japanese Energy Policy," paper presented at *The 1986 Meeting of the North American Regional Science Association*, November, 1986.

Symbolics, Inc., [1986], "Using The Online Documentation System," in *User's Guide to Symbolics Computers*, Cambridge: Symbolics, Inc., 1986: 43-64.

A. Tanaka, [1981], *Chinese International Conflict Behavior, 1949-1978*, Cambridge: Doctoral Dissertation, Department of Political Science, M.I.T., 1981.

A. Tanaka, [1984], "China, China Watching, and CHINA_WATCHER," in (Sylvan & Chan, 1984: 310-344).

S. T. Thorson, [1984], "Intentional Inferencing in Foreign Policy: An AI Approach," in (Sylvan & Chan, 1984, 280-309).

S. T. Thorson and D. A. Sylvan, [1982], "Counterfactuals and the Cuban Missile

Crisis," *International Studies Quarterly*, December, 1982, 26(4): 539-571.

G. Tonfoni, [1985], editor, *Artificial Intelligence and Text Understanding, Quaderni di Ricerca Linguistica*, 1985, 6.

S. Toulmin, [1964], *The Uses of Argument*, Cambridge: Cambridge University Press, 1964.

R. H. Trigg & M. Weiser, [1984], "Textnet: A Network-Based Approach to Text Handling," College Park: Computer Science Department, University of Maryland, 1984.

A. M. Turing,[1950], "Computing Machinery and Intelligence," *Mind*, 1950, 54: 443-460.

C. H. Waddington, [1976], "Evolution in The Sub-Human World," in E. Jantsch and C. H. Waddington, editors, *Evolution and Consciousness: Human Systems in Transition*, Reading: Addison-Wesley, 1976: 11-15.

I. Wallerstein, [1974], *The Modern World System: Capitalist Agriculture and the Origins of the European World Economy in The Sixteenth Century*," New York: Academic Press, 1974.

I. Wallerstein, [1980], *The Modern World System II: Mercantilism and The consolidation of the European World Economy, 1600-1750*, New York: Academic Press, 1980.

D. Waltz, [1985], "Scientific Datalink's Artificial Intelligence Classification Scheme," *The AI Magazine*, 1985, 6(1): 58-63.

D. A. Waterman & M. A. Peterson, [1981], "Models of Legal Decision-Making," Santa Monica: RAND, Technical Report No. R-2717-ICJ, 1981.

R. P. Weber, [1978], *The Dynamics of Value Change, Transformations and Cycles: British Speeches from the Throne, 1689-1972*, Ph.D. Thesis, University of Connecticut, 1978.

R. P. Weber, [1985], *Basic Content Analysis*, Beverly Hills: Sage, 1985.

S. K. White, [1986], "Foucault's Challenge to Critical Theory," *American Political Science Review*, June, 1986, 80(2): 419-432.

Y. Wilks, [1987], "Primitives," in Shapiro (1987: 759-761).

T. Winograd, [1980], "What Does It Mean to Understand Natural Language," *Cognitive Science* 1980, 4: 209-241.

T. Winograd and F. Flores, [1986], *Understanding Computers and Cognition: A New Foundation for Design*, Norwood, NJ: Ablex, 1986.

P. H. Winston, [1975], "Learning Structural Descriptions from Examples," in P. H. Winston, editor, *The Psychology of Computer Vision*, New York: McGraw-Hill, 1975: 157-210.

P. H. Winston, [1977], "Learning to Identify Toy Block Structures," in P. N. Johnson-Laird and P. C. Wason, editors, *Thinking: Readings in Cognitive Science*, Cambridge: Cambridge University Press, 1977, 199-211.

P. H. Winston, [1980], "Learning and Reasoning by Analogy," *Communications of the ACM*, December, 1980, 23(12).

P. H. Winston, [1984], *Artificial Intelligence*, Reading: Addison-Wesley, 1984.

L. Wittgenstein, [1953], *Philosophical Investigations*, New York: MacMillan, 1958. First edition published in 1953.