MASSACHUSETTS INSTITUTE
OF TECHNOLOGY
CENTER FOR INTERNATIONAL STUDIES

PROJECT WORKING PAPER

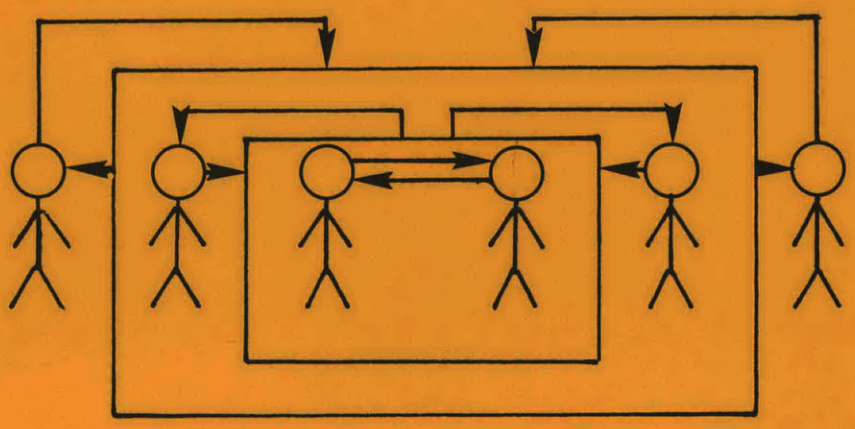
SIMULATING THE PRACTICAL LOGIC OF DETERRENCE: AN APPLICATION FOR ARTIFICIAL INTELLIGENCE

by

Dwain Mefford

Massachusetts Institute of Technology

REFLECTIVE LOGICS FOR RESOLVING INSECURITY DILEMMAS



SIMULATING THE PRACTICAL LOGIC OF DETERRENCE:

AN APPLICATION FOR ARTIFICIAL INTELLIGENCE

by

Dwain Mefford
Massachusetts Institute of Technology

Prepared for delivery at the Annual Convention of the International Studies Association, the Royal York Hotel, March 21-24, 1979. Work reported in this paper has been supported by N.S.F. Grant No. 7806707 to the Center for International Studies, Massachusetts Institute of Technology, Hayward R. Alker, Jr., chief investigator.

Abstract:

This paper describes the design of a simulation of strategic problem-solving. Both in the conception of what transpires in strategic perception and planning, and in the implementation of these procedures, we draw on techniques and approaches which have been developed in several of the areas of work in artificial intelligence. We introduce a notion of resonance or mapping between a surface logic of argument and richly structured memories of the past. This combines conventionally structured argumentation with conviction-instilling analogizing. To implement this we employ "frames" and frame manipulation. Using Schelling, we expand on his thesis regarding the tacit dimensions of strategic interaction, stressing the complexity of achieving and maintaining a mutually accessible practical logic sufficient for parallel problem-solving. Deterrence is an example of such a practical logic. It is rooted in jointly experienced episodes, and is reaffirmed and restructured through time as a function of the contingent contexts in which the parties are forced to act, and through acting to define themselves to one another. As a mechanism for didactically showing another party the logic of one's actions, we sketch the procedure of "glass boxing", using McNamara's official reflections as an example. In addition to glass boxing the major process for the transformation of strategic perception and reasoning involves the alteration and respecification of frame "entry conditionals". Entry conditionals, like gates on which experience is deemed relevant to a present situation, are the principle determinants of what effectively constrains the field of choice out of which actual courses of action emerge.

In 1953 J. Robert Oppenheimer voiced a sane man's incredulity at one of the foremost theorems of deterrence: the proposition that the defense of populations leads not to peace and stability, but to the opposite. To Oppenheimer the thinking which drives this argument is so uncanny that it could "occur only when even the men who know the facts can find no one to talk about them, when the facts are too secrete for discussion, and thus for thought."[1] To Oppenheimer the proposition approaches the limits of what can be conceived and believed. From our vantage point twenty-five years later, Oppenhemier's wonder is itself hard to fathom. The basic axioms of deterrence and derivative propositions such as this one have become comfortable fixtures of debate and policy. In these matters all seem to talk the same special language and assume the same special logic, the leadership, the bureaucracies, the leading voices in the media, the public at large -- in short, the people and agencies who define, carry out, or simply observe the affairs of state. All who count have become remarkably proficient in exercising that peculiar reasoning which was astounding a generation ago. It has become the norm.

Deterrence is a vehicle for a particular kind of problem-solving. It serves to identify the prerequisites of "security" through the application of categories and criteria, such as "counterforce" and "first strike" -- criteria and rules of inference which permit cognoscenti and laymen alike to build arguments relating arms decisions to intentions, to reciprocal fears, and ultimately to the prospect of war. Since there is a clear affinity between the

---

[1] J. Robert Oppenheimer, "Atomic Weapons and American Policy," Bulletin of Atomic Scientists IX (July, 1953):202-05. Cited in George H. Quester, Nuclear Diplomacy. The First Twenty-Five Years, New York: Dunellen, 1970, p. 78.

practice of deterrence thinking and the performance of other types of general and specialized problem-solving which have been the principle object of study in Artificial Intelligence since its inception,[2] it seems likely that the perspectives and techniques of AI, particularly those expressly developed for replicating common sense reasoning as enacted within practical contexts, might be usefully applied to describe the logic of strategic thinking as we witness and practice it.

But deterrence as a medium for the common articulation of the nature of strategic relations is substantially more than an inference system to be captured in axiomatic form. It has emerged and evolved through concrete historical episodes and represents, in a very real sense, the precipitate of thirty years of threats and force. The point in citing Oppenheimer is to draw attention to the fact that there can be, and has been, an epic change in the canons underlying our thinking and convictions in matters of security. In understanding such transformation we are faced with the question of how to articulate and plot an elusive but real reorientation. If we are to comprehend how this body of rules and perceptions is reproduced and modified through time, then we need to devise a means for rendering the manner in which actual events, and the concrete problems of interpretation and practical response they pose, become embedded into the shared sensitivities which orient our sense for what is politically real and feasible. We require a theoretical perspective which can adequately characterize

---

[2] Allen Newell's "Artificial Intelligence and the Concept of Mind," in Roger C. Schank and Kenneth Mark Colby, eds., Computer Models of Thought and Language, San Francisco: Freeman, 1973, pp. 1-60 is an authoritative reconstruction of the history of AI in terms of the domains of intelligent behavior which have been incorporated into the field as it expanded. A more recent selective survey with an emphasis on programming conception and implementation is Patrick Winston, Artificial Intelligence, Reading, MA, Addison-Wesley, 1977.

the nature of what catalyzes or inhibits the self-changing "intelligence" we see operating in the episodic history of deterrence.

A line of research that we might adapt to the requirements of our subject matter involves the design of systems equipped with large memory and capable of bootstrapping their problem-solving performance by modifying and innovating upon their store of heuristics.[3] But some of the most visible and critical dynamics at work within strategic interactions appear to have no immediate analogue in these systems, nor for that matter in any of the exemplary systems which have been constructed to date. The lack is not in the technical facilities, but in the restricted field of phenomena to which they have been applied. One such factor is that of "prominence" of a solution.[4] With or without communication, the very possiblity of resolving a competitive or conflict situation of any complexity rests in the capacity of each of the actors to identify and solve the matter in what they know to be essentially the same way. They operate with some common algorithm or some reference point such as an exemplary case. If the competitive relationship recurs through time and is bound up with a diversity of issues in varying contexts, then the success of the interactions must depend on a practical logic, what Schelling terms a "tradition."[5] This practical

---

[3]Two such systems, and foremost examples of the Carnegie-Mellon perspective are: D.A. Waterman, "Generalization Learning Techniques for Automating the Learning of Heuristics," Artificial Intelligence, 1 (1970): 121-170; H.A. Simon and G. Lea, "Problem Solving and Rule Induction: A Unified View," in L.W. Gregg, ed., Knowledge and Cognition, Potomac, MD.: Erlbaum, 1974, pp. 105-27. For the basic tenets of the production system approach for which these are examples, see Herbert A. Simon, "Information Processing Models of Cognition," Annual Review of Psychology, 30(1979): 363-96.

[4]The thesis in chapter 3 of Thomas Schelling, The Strategy of Conflict, Cambridge, MA:Harvard University Press, 1960.

logic is a set of instructions or pointers, undoubtedly tied to episodes in the past, which have the cardinal property of being salient to all involved.

Using Herbert Simon's terminology, the most fundamental problem we face in representing an evolving system of strategic interactions is the problem of capturing the processes of "search" and "evaluation" which prepare and channel the act of choice. Problem-solving in parallel requires that the participants achieve some minimal competence in (1) co-defining the space of the problem in which they are both engaged, and (2)jointly and demonstrably exercising similar algorithms for searching that space for solutions which are viable. As Schelling argues with demonstrations in practical settings, the strategic task involves reading the same message, constructing the same problem, and solving it in concert via some shared casuistry or common perception. The fact that parallelisms of various sorts must emerge and be constantly maintained if the strategic interaction is not to spring apart due to the centrifugal forces of misunderstanding and bad information, makes problem-solving in this context an intrinsically more complicated achievement than problem-solving as it is generally protrayed in the literature. The inclination is to assume that the act of solving problems, exemplified in the puzzles and tasks Herbert Simon has investigated at such length,[6] is an act performed by a single agent acting in essential isolation. The problem solver performs without regard to the performance of others.

---

[5] ibid., pp. 106-7.

Strategic problem-solving, by definition, imposes an entirely new order of constraints which must be satisfied at each point in time. In point of fact, as political history demonstrates repeatedly, the "solution" to the strategic problem is often not more nor less than the solving of the problem of finding a mutually intelligible definition of what the issues of contention really are. Technical questions, like those concluded in the protocols of the SALT treaty, are circumscribed at every point by a larger exercise in shaping a language which will support analysis and argument that is transparent to all parties.

The work in proof theory, analogical reasoning, rule-induction, and so forth, has generally assumed a monolithic image of the problem-solver. But irrespective of this, these directions within Artificial Intelligence can be made compatible with a notion of problem-solving as a distributed and interactive affair. The basic components already engineered can be assembled into a system capable of portraying the issues of problem-solving in parallel.

At M.I.T. we have been experimenting with a system consisting of several "problem-solvers" or actors which, as initialized, differ not only in their knowledge bases, but in the heuristics which direct the logic of their calculations and actions. Through time, as a complex function of their collective response to crisis-like contingencies, these actors generate internal models of themselves, each other, and

---

[6] Allan Newell and Herbert A. Simon, Human Problem Solving, Englewood Cliffs, N.J.: Prentice-Hall, 1972.

the history of their "world." This individual capacity to monitor one's own performance and the performance of others against the backdrop of previous outcomes permits the actors to generate working relationships. These relations, or rather the practical problem-solving they embody, evolve through time undergoing incremental and innovative change and evidencing an array of circularities and other pathologies.

The system as a whole is constructed around an essential claim which is the basis of not only its programming design, but the essential object of the investigation. It is intended to be as a genuine contribution to the conception of what transpires in the practical logic of strategic argument and conviction as that logic is enacted. The notion is that this form of reasoning, as it is displayed in Congressional hearings, or the syndicated columns in the New York Times, or the talk shows on AM radio, -- or all and sundry taken as an aggregate -- functions at two levels simultaneously. The first, and more superficial, obeys the conventions of proper argument, listing or alluding to premisses and evidence and drawing reasoned conclusions. For example, Paul Nitze's case for the MX is structured as a practical syllogism: given Russian missile capacity, and given MIRVs, the American land-based ICBM force will be vulnerable in the eighties, which requires that we... The second level, and the one from which conviction appears to flow up into the brief and incomplete syllogisms proferred at the more superficial level, involves

a process of marshalling entire assemblies of concrete evidence impacted in the complex memory which blends images of the real or imagined past. This densely structured ensemble of evidence, which we will term a "frame" for reasons to be given, is rich in connotations and serves to fill out and otherwise redeem the laconic claims and deductions constructed in the course of argument. A frame might be the complex image retained of the Second World War, and all of the personal and public fears and rationales woven into such an event. An alternative frame might be the package of instantiated "lessons" associated with Vietnam. The point is that there is a complex mapping operation which links the fragile and information-poor syllogisms of debate to the descriptively thick and compelling practical knowledge of what is believed to have been the case in the past. We will call this mapping relation resonance, a process of filling-out and amplifying which redeems the logical ellipses which characterize the surface logic of strategic thinking. This is the backup which explains why the omission of premisses and connectives in such arguments rarely disturbs the cadence of reasoning: such ommissions are functionally "bound" within the network of connections and unquestioned associations which a resonating frame supplies. What this mapping or resonance entails can be made clearer after a discussion of the philosophy of data structure which "frame" signals.

Since in the selection and design of the medium for our simulation we are departing from the formats which are conventionally

used in the computer-based modelling of conflict and strategic interaction,[7] it seems imperative to justify our choice in terms of what it purports to achieve and in terms of its plausibility as an extension of demonstrated techniques and conceptions. The resources we are drawing upon are the approaches and instruments which have been developed in automatic programming (planning), natural language understanding, and other principle subfields within Artificial Intelligence. To sketch and justify this application is the object of the present paper. We avoid a detailed account of the engineering of the system since listings and flowcharting is available elsewhere.[8]

The principle questions which we pose and seek to embody in the operation of the system can be grouped under three headings:

(1) Problem-solving per se: the fitting and manipulation of real world knowledge within an evolving representation -- issues pertinent to the history of strategic thinking as a changing medium;

(2) Problem-solving in parallel: the syncretization of autonomous actors which are capable of achieving coordinated analysis and behavior via adjustment and restructuring -- questions directed to the issue of the tacit dimensions of strategic interaction;

---

[7] For instance, Harold Guetzkow's INS or Stuart Bremer's SIPER.

[8] As memoranda for the NSF project #86648, Hayward R. Alker, Jr., chief investigator.

(3) The reproduction and expansion of the capacity to problem-solve in the face of new contingencies -- the ultimate determinant of the robustness of a working relationship which channels and forecloses lethal conflict.

Problem-solving per se

The areas of research in AI which are for us most applicable or suggestive belong to what has been termed the knowledge or representational perspective on the design of problem-solving systems. This contrasts with the power orientation which seeks to achieve full generality in its design of a computational mechanism. The distinction is between a commitment to the representational and inferential issues intrinsic to highly specific, task-oriented performance (such as that involved in running an office or an airline), versus a commitment to completely general forms of reasoning exemplified in systems based on the predicate calculus. The tradeoff pits specificity and diverse forms of knowledge and inference against computational power, in the mathematician's sense, operating on a uniformly constructed and ordered knowledge base.[9]

To represent and organize knowledge which we encounter in concrete situations, the designers of a number of AI systems have developed large, flexible data structures, variously termed "frames," "schema," and "scripts." The principle features of these data structures which make them appropriate to the organization and manipulation of real world knowledge (the features of "instantiation," "procedural embedding," "defaults," etc.) are presented and illus-

---

[9] Ira Goldstein and Seymour Papert, "Artificial Intelligence, Language, and the Study of Knowledge," Cognitive Science 1(1977):84-123. Marvin Minsky and Seymour Papert, Artificial Intelligence, Condon Lectures, Oregon State System of Higher Education, Eugene, Oregon, 1974.

strated in Marvin Minsky's account of frames.[10] As an important supplement to the Minsky use and interpretation of these structures, principally to questions arising in such areas as robotic vision, Robert Abelson and Roger Schank's development of formally analogous configurations is of very special importance for our purposes. Their concern has been with that ordering of information which makes possible the construction and understanding of plots and plans and other causally and temporarily organized sequences.[11] Their notion of "script" capitalized on the manner in which episodes are chained together into accounts or stories or histories or biographies.

---

[10] Marvin Minsky, "A Framework of Representing Knowledge," in Patrick Winston, ed., The Psychology of Computer Vision, New York: McGraw-Hill, 1975, pp 211-277; An abridged version is: Marvin Minsky, "Minsky's Frame System Theory," in R. Schank and B. Nash-Webber, eds., Theoretical Issues in Natural Language Processing, An Interdisciplinary Workshop, Cambridge, MA: M.I.T., 1975.

[11] The concept of script has undergone a series of amplifications since it was first introduced by Abelson in his simulation of right wing foreign policy thinking (reported in Robert Abelson, "The Structure of Belief Systems," in R. Schank and K. Colby, eds., Computer Models of Thought and Language, San Francisco: Freeman, 1973). Abelson's current and most general definition of "script" presents it as "a coherent sequence of events expected by an individual either as a participant or an observer." (Robert Abelson, "Script Processing in Attitude Formation and Decision Making," in J.S. Carroll and J.W. Payne, eds., Cognition and Social Behavior, Hillsdale, N.J., Erlbaum, 1976. For the relation of scripts to the Rumelhart and Norman notion of "schema" with which it closely resembles, see Robert Abelson, "Scripts," Address to the Midwest Psychological Association, May, 1978. A recent account of schemas is D.E. Rumelhart and A. Ortony, "The Representation of Knowledge in Memory," in R.C. Anderson, R.J. Spiro and W.E. Montague, eds., Schooling and the Acquisition of Knowledge, Hillsdale, N.J., Erlbaum, 1976.

Both frames and scripts expand the earlier AI strategy for structuring knowledge in memory in the form of nodes and links comprising networks of facts and relations.[12] Frames are essentially ensembles of such nodes and links, and are studded with instructions as to the use of the information represented. That is, procedural rules and cues are attached which direct attention to requirements, consequences, etc., which must hold if and when the frame comes into play. Used as a vehicle for representing situations,

> [t]he "top levels" ... are fixed, and represent things that are always true about the supposed situation. The lower levels have many terminals -- "slots" that must be filled by specific instances or data. Each terminal can specify conditions its assignments must meet. (The assignments themselves are usually smaller "subframes.") Simple conditions are specified by markers that might require a terminal assignment to be a person, an object of sufficient value, or a pointer to a subframe of a certain type. More complex conditions can specify relations among the things assigned to several terminals. Collections of related frames are linked together into frame-systems..." [13]

---

[12] Ross Quillian is credited with the application of the node and link format to semantically rich knowledge bases: R. Quillian, "Word Concepts: A Theory and Simulation of Some Basic Semantic Capabilities," Behavioral Science 12(1967): 410-430; R. Quillian, "Semantic Memory," in Marvin Minsky, ed., Semantic Information Processing, Cambridge, M.I.T. Press, 1968, pp. 227-270.

[13] Winston, Artificial Intelligence, op. cit. p. 180.

Problem-solving on the basis of frames can be construed as a process of fitting incoming information into the configuration of types of facts and related actions which have been preassembled to constitute the frames in the system's repertoire. Identifying the course of action which, for example, a crisis or negotiating ploy holds out for an actor is a process of selecting and imposing a frame or analogue onto the situation. In so doing, the actor becomes aware of a specified array of plausible dangers and opportunities -- the assembled information which will enter into subsequent calculations as to options and their consequences. In a fundamental sense, the choice of action is structured by the initial, possibly inadvertent selection of the frame or script which is to orient the definition of the situation. Choice is circumscribed by the criteria and limits which the orienting frame imposes on the field of conceivable options. To stress, with the concept of frame, this pre-structuring of choice is to emphasize that strategic problem-solving involves substantially more than the conventional notion of the analysis and evaluation of all possible coursesof action or interpretations. As Abelson argues, the deeper dimension of decision making is almost binary in character: a matter of invoking or not invoking a particular definition. Once the definition or frame is in place, for instance, that riots in a foreign capital is the work of Communist agents, the universe of response is essentially fixed. The crucial initial step is finding the applicable reference, a step which can be thought of as "matching"

---

[14] Robert Abelson, "Scripts," op. cit., p. 14.

new information against what one or more frames require if they are to be invoked. This matching involves fulfilling the frame's "entry conditionals," the crucial initial step in the application of a frame or script. This entails satisfying the "front end to the script, mediating between presentation of the defining context for the script and participation in it. This is a little decision program ... learned along with the script, containing "action rules" or criteria for participation."[14]

If we lack in our repertoire of experiences or within the scope of our imagination anything which closely resembles some new situation we confront, we are forced to adapt existent frames or, in the rare case, innovate new ones. In such circumstances, the resources at our disposal are the attributes of the frames we can call up and modify, extend and qualify, and otherwise reshape. This process of restructuring, the basic activity of learning, belongs to the third group of issues concerned with the behavior of an intelligent system engaged in self-modification.

The basic question for frame-based problem-solving are the questions of how frames or scripts are first acquired. Acquisition might be accounted for as a process of the imprinting of a particular situation, an automatic generalizating or stereotyping which encompasses an entire class of future situations. What triggers this imprinting on individual and collective memory, and what order of experience must subsequently transpire if what has been acquired is to be qualified or supplanted, are open questions.

It seems intuitive that the lived experience of war has such an imprinting character. An alternative mechanism for the acquisition of frames is the notion of assimilating patterns of information by "working through" examples in the manner that skills and theoretical knowledge is absorbed by an individual into his working competence. The clear analogy in strategic interactions is the "learning" which can accompany the successful or disastrous resolution of a crisis. Experience, painfully assembled, is "compiled" into a fusion of rules and illustrations which are henceforth available as a prototypical for subsequent contingencies with struturally similar attributes. This general process is possibly illustrated in the "immunizing" and "catalytic" effects which the face-offs in Berlin and the Straits of Taiwan have purportedly had on the functioning of the international system.[15]

The case that can be made for adopting frames as the device for organizing and manipulating the information base of our simulation is a case which closely parallels the argument Alexander George and Richard Smoke make for adopting a methodological alternative to the level of analysis generally assumed in the study of crisis and strategic interaction. They argue for "situational analysis,"[16] a codeword for an approach which combines attention to detail, as in case studies, with a commitment to disciplined inter-period, inter-crisis comparison. The shift is to a methodological "middle way" which bridges between the parsimony of formulation prized by behavioral science and the descriptive completeness but

---

[15]Alexander George and Richard Smoke, Deterrence in American Foreign Policy. Theory and Practice, New York: Columbia University Press, 1974, p. 508.

ideosyncratic focus of the historian. Frames, since they are expressly designed to house a diversity of types of information within a single configuration, would seem a workable format for assembling richness of detail while facilitating multi-dimensional comparison. To use Minsky's terminology, a situational frame may well be unique in the arrangement of descriptive detail represented at its "terminals" or "leaves," but, in all probability, resembles other frames derived from other concrete situations, in its deeper structure. Hence, the possiblity of comparison without the sacrifice of thick description.

The frustrating task before the analyst or theorist, that of forcing the comparison of incomparables, is little more than an extension of the distortion-prone comparing and interpreting of cases which the decision maker and all of his formal and informal advisers must continually accomplish. This process means juxtaposing whole arrays of heterogeneous factors in the effort to determine if the situation of the moment belongs to one class of crises or another or to neither but both. Is the situation in Iran, and the appropriate American attitude best assimilated to events like the resurgence of Islam in the early part of this century, or is it more correct to see it in light of the French experience in Algeria, or are both of clear relevance along with a great deal more? Whatever definition emerges, even if it is ultimately unique unto itself, it will draw on and composite a number of previously compiled frames. The process of disambiguating and understanding the newly encountered situation is a process of fabricating a new configuration of characteristics via the intersection, splicing, and respecification of frames already acquired.

---

[16] Ibid., Chapter 16.

Problem-solving in the plural; problem-solving in parallel.

Referring to Schelling's discussion of the tacit dimension of strategic problem-solving we have already raised the question of dynamic coordination. In essence the design of our simulation is the embodiment of a proposal for extending Schelling. Where he talks about the perceptual syncratization which parties to a conflict must achieve if a stable basis for resolution is to come about, we want to stress the more inclusive activity of which coordinated perception is a part: the competence acquired in solving the shared problem via shared heuristics. Schelling deals with this question of the acquisition of such a shared competence through time as a fundamental by-product of repeated encounters, when he speculates on what might be involved in the building of a "tradition" for the management of conflict.[17] Deterrence, as the practical logic or body of heuristics which has emerged in the course of thirty years of confrontation is a functioning example of such a medium.

To simulate the emergence and subsequent reordering of such a medium we construct several autonomous actors or problem-solvers, each of which is opaque to the other in terms of the interpretation and planning which determines his course of action. Initially, each can assume only that there are certain basic traits which all actors share qua actors, for example, the capacity to recognize the fact of a crisis, though the fact's interpretation for each actor is assumed to differ. In short, the state of the world as a whole at the outset

---

17 Schelling, op cit., Chapter 3.

is Hobbesian. Though each actor knows that his reasoning and perceptions are complex and laced with ambiguity, he tends to assume, unless evidence emerges for thinking otherwise, that the reasoning and intentions of other actors is either unintelligible or basically simple, eg. dictated by crude principles of power maximization constrained only by fear. Each individual monitors his performance and the performance of others as enacted in specific contexts, by and large crises. Using a file of such information the actor can pose and confirm to himself hypotheses about the effectiveness of his own actions (and the calculations on which they are based), and about the logic and perception of other significant actors. In time, the actor differentiates heuristics which appear to enter into both his and the other's calculations. On the basis of these shared inference rules and the mutual acknowledgement of key episodes which take on the status of precedents, individual actors realize that their strategic planning can be safely predicated on the strategic planning of others. As a consequence, reciprocally conditioned policies become feasible, and so enter into the infrastructure on an evolving strategic relation among or between the actors.

If we stress that the problem-solving which transpires in this system is distributed among its subsystems and that the aggregate performance amounts to an order of problem-solving which proceeds in and through the subsystems, then we can claim that in the conception of this system we share in a perspective which Hewitt, Minsky, and Papert refer to with the image of an interacting "society of experts."[18] The conventional notion of problem-solving assumes that

some single integrated intelligence searches a space of possible solutions.[19] In opposition to this, the image of a "society of experts" stresses that problem-solving is a collective activity involving the, possibly conflictual, interaction of numerous special functions which complement but sometimes mask each other in their effects. Such contributing functions, like remembrances and associations of things past, vie for application to the problem at hand. The notion is one of critical and conflicting interaction of a multitude of types of knowledge which aggregate into a composite whole like the consensus which arises through the efforts of numerous researchers. The image of a "society of experts" calls up the Peircean conception of an emerging and ever evolving scientific consensus. The focus in this image of collective problem-solving is captured in Hewitt's quotation from Edward O. Wilson: "Reciprocal communication of a cooperative nature is the essential intuitive criterion of a society." Though Minsky and Papert's principle concern is to fabricate a computational theory of the brain and of personality, and though Hewitt's

---

[18] Marvin Minsky, "Plain Talk About Neurodevelopmental Epistemology," International Joint Conference on Artificial Intelligence, Cambridge, MA, Aug. 22-25, 1977; Pittsburgh: Department of Computer Science, Carnegie-Mellon University, 1977 Vol. 2, pp. 1083-1092. Carl Hewitt, "Viewing Control Structures as Patterns of Passing Messages," Artificial Intelligence 8 (1977): 323-364.

[19] Hewitt, Ibid., p. 325.

[20] Ibid.

[21] Gary Hendrix, "Encoding Knowledge in Partioned Networks," Menlo Park, California; Artificial Intelligence Center, SRI International, April 1978.

objective is to provide a radically flexible control structure compatible with parallel processing, the perspective they are advocating fits well with the character of political intelligence we want to render.

At an instrumental level, the construction of a world in which information, and the logic determining the processing of that information, is localized or scoped can be engineered using the device of "partioned networks."[21] The idea is that information is located within bounded spaces, and is accessible to only selected receivers within the system. This notion has been used to depict the shifting states of knowledge which actors command as they attempt to thwart each other and then subsequently find out about each other's plans. An example is a programmed assimilation of the tale of Hansel and Gretel in terms of the shifting content and demarcation of what is known and believed by each of the characters.[22] Given that we want to depict not only what is known by the various actors, but the reasoning procedures which these actors are applying to what they know, the "spaces" in our system contain mechanisms of inference as well as facts.

How such operating procedures come to be jointly held, how they are first consensually generated and then reaffirmed in the course of succeeding interactions, is a central question for the simulation. A principle mechanism responsible for creating and maintaining a working alignment between actors might be a didactic process that we are experimenting with. It involves what some theorists in computer aided instruction call "glass boxing:"[23] a strategy

---

[22] Bertram Bruce and Denis Newman, "Interacting Plans," Technical Report No. 88, Cambridge, MA: Bolt, Beranek, and Newman, June 1978.

[23] Goldstein and Papert, op. cit., p. 116.

for inducing a student to reconstruct for himself the reasoning which seems to account for the line of questions and responses the instructor, in this case a computer program, pursues in reacting to the student's performance. In international politics, glass boxing would be the effort to telegraph to an adversary not only the implications and credibility of one's policy, but the reasoning on which it is based. By doing so, the intended significance of a line of action is more effectively and reliably deciphered. Since each party can instruct the other in such a manner, a consensus consisting of each's understanding of the other is feasible, and once established may enable the parties to cooperate in curbing unintended and mutually disadvantageous outcomes. In the history of deterrence, Robert MacNamara's posture statements issued each year in justification of weapons programs might be considered an example of deliberate "glass boxing." "Glass boxing" vis-a-vis Soviet strategic planners appears to be one of McNamara's motives in spelling out the lesson of the "missile gap" of the early sixties: suspicion plus inadequate information, amplified by worst case analysis, can have but one impact on an arms race.

> ...in the course of hedging against what was then only a theoretically possible Soviet build-up we took decisions which have resulted in our current superiority... But the blunt fact remains that if we had had more accurate information about planned Soviet strategic forces, we simply would not have needed to build as large a nuclear arsenal as we have today... Furthermore, that decision in itself, justified as it was, in the end could not possibly have left unaffected the Soviet Union's future nuclear plans... .[24]

---

[24] Robert McNamara, Address to the United Press Editors and Publishers, San Francisco, Sept. 18, 1967, reprinted in The Essence of Security, Reflection in Office, New York: Harper & Row, 1968, pp. 57-58.

The extension and restructuring of a shared logic of strategic interaction: the question of "learning."

Deliberately displaying one's rationale and calculation, what we have termed "glass boxing," is a procedure which can alter the state of mutually shared knowledge, and so affect the strategies of actors. There are several other mechanisms for transforming the underlying capacity of the actors to tacitly coordinate and interlock their calculations and behavior. Possibly the simplest involves innovating new rules of inference on the basis of examining past outcomes. This is the essence of the Waterman program referred to earlier, an "adaptive production system" which induces heuristics. In light of the minor role that pure deduction from rules seems to play in the "intelligence" of states, as an alternative to the Waterman-Simon approach, we have pursued learning processes more in keeping with the frame and script mode of data structuring -- structuring and manipulation which is more in line with our basic conviction that precedents, thick with description, are the chief shaping force in the formulation of policy.

In essence, incremental and innovative change in political reasoning is affected by qualifications attached to previously acquired frames -- re-specifications which precipitate out of the practical learning which accrues in the course of handling major events. The "entry conditionals" of the relevant frames, that is, the prerequisites which determine the application of a frame are tightened or otherwise modified. An example might be rethinking the "lesson" of Korea in light of the policy and analytic debacle of Vietnam. The mechanisms for registering the misfit between frame and actuality, and the consequent steps

taken to revise or innovate, is the subject of AI research such as the work of Patrick Winston on inter-frame comparison.[25]

This process of correction and amplification can proceed inadvertantly or with deliberation. A political analogue to deliberate reassessment of the faulty reasoning deemed responsible for past disasters is the practice of "postaudit" which the British planning staffs are reported to carry out.[26] Such retrospective search can be likened to deviant case analysis. An actor's or group of actors capacity to control outcomes would seem to be clearly related to their capacity to retain and critically reinterpret the past. A principle objective of the simulation is to examine what compounding effects this order of individual and collective self-monitoring has on the dynamics of an international order.

---

[25]Patrick Winston, "Learning by Hypothesizing and Justifying Transfer Frames," Massachusetts Institute of Technology, Artifical Intelligence Memorandum No. 414, April 1977.

[26]Lincoln Bloomfield, "Planning Foreign Policy: Can It Be Done," Political Science Quarterly 93 (1978):369-391.

In conclusion, let us select out what we believe to be most valuable in the design of the simulation and the questions it embodies.

1. We have introduced a notion of resonance or mapping between a surface logic of argument and richly structured memories of the past. This combines conventionally structured argumentation with conviction-instilling analogizing. The implementation employs frames and frame manipulation.[27]

2. Using Schelling, we expand on the tacit dimensions of strategic interaction, stressing the complexity of achieving and maintaining a mutually accessible practical logic for parallel problem-solving. This practical logic, rooted in jointly experienced episodes, is reaffirmed and restructured through time as a function of the contingent contexts in which the parties are forced to act, and through acting to define themselves to one another.

3. As a mechanism for didactically showing another party the logic of one's actions, we have sketched the procedure of "glass boxing" using McNamara's official reflections as an example.

4. The chief mechanism for the transformation of strategic perception and reasoning is the alteration and respecification of frame "entry conditionals." Entry conditionals are the principle determinants of what effectively constrains the field of choice from which actual

decisions emerge.

---

[27] Bruce Roberts and Ira Goldstein, "The FRL Primer," Massachusetts Institute of Technology, Artificial Intelligence Laboratory Memorandum No. 408, Cambridge, MA, July, 1977; Daniel Bobrow and Terry Winograd, "An Overview of KRL, A Knowledge Representational Language," Cognitive Science 1 (1977): 3-46.

Reflective Logics for Resolving Insecurity Dilemmas - Project Working Papers

1. Hayward R. Alker, Jr., James Bennett, and Dwain Mefford, "Generalized Precedent Logics for Resolving Insecurity Dilemmas," (1979).

1a. Hayward R. Alker, Jr. and William J. Greenberg, "On Simulating Collective Security Regime Alternatives," (1976).

1b. James P. Bennett and Hayward R. Alker, Jr., "When National Security Policies Bred Collective Insecurity: The War of the Pacific in a World Politics Simulation," (1977).

2. Hayward R. Alker, Jr., "Some Dynamics of Ideo-Logical Thought," (1979).

3. Dwain Mefford, "Historical-Strategic Precedents as Procedures: A Strategy for Applying Selected Artificial Intelligence Design Features to Simulation Building in International Relations," (1978).

4. Dwain Mefford, "'Recursive' Versus 'Structural' Reflection: Dialectics as an Instrument of Analysis Applied to the Genesis of Interpersonal Relations," (1978).

5. Dwain Mefford, "Simulating the Practical Logic of Deterrence: An Application for Artificial Intelligence," (1979).

6. Joshua Epstein, "The Extended Calculus of G. Spencer Brown and Related Areas of Logic and Mathematics," (1979).

7. Hayward R. Alker, Jr., "Logic, Dialectics, Politics: Some Recent Controversies," (1979).

8. Dwain Mefford, "From Static to Transformational Models of Political Reasoning," (1979).

9. Dwain Mefford, "The Grammar of Change in Political Beliefs: A Constructive Critique of Cognitive Modeling," (1979).

10. Hayward R. Alker, Jr., "From Political Cybernetics to Global Modeling," (1979).

11. Dwain Mefford, "The Language of Crisis and Its Evolving Grammar," (1979).

12. Hayward R. Alker, Jr. and Roger Hurwitz, "Resolving Prisoner's Dilemmas," (1979).

12a. Hayward R. Alker, Jr. and Roger Hurwitz, "Teacher's Manual for Resolving Prisoner's Dilemmas," with programs by Akihiko Tanaka (1979).

13. Lee Farris, Hayward R. Alker, Jr., Kathleen Carley and Frank L. Sherman, "Phase/Actor Disaggregated Butterworth-Scranton Codebook," (1979).

The writing and/or reproduction of all of the papers in this series have been financed in whole or in part by National Science Foundation Grants #7806707 and #GS-2429 to the Center for International Studies, Massachusetts Institute of Technology. Neither of the aforementioned organizations should be held responsible for any of the views expressed herein.

As project working papers, they are being circulated in order to encourage comments on work in progress; none should be considered a final statement. Unless they have been published elsewhere as noted, please request permission to cite from the authors.