Experiments With an Artificial Mediator: A Progress Report

Charles S. Taber

Department of Political Science
SUNY at Stony Brook
Stony Brook, NY 11794-4392
chuck@datalab2.sbs.sunysb.edu
http://www.sunysb.edu/polsci/

Abstract: This paper reports on an ongoing project to develop an artificial mediator (AMed). I have adapted my earlier cognitive approach to modeling foreign policy decision making (POLI and EVIN, in particular) to the domain of mediation. I model the problem in terms of *information processing*, where mediation and negotiation are viewed as streams of messages among the negotiation participants, which the mediator must process and contribute to. To do this, the mediator must construct an understanding, or *problem representation*, of the messages, which I call the *perceptual layer*. Once a message (or set of messages) is interpreted, the mediator must decide how to respond. This is the *reasoning layer*. Information processing in the model follows well-known theoretical principles from cognitive science (with emphasis on the work of Schank & Abelson, 1977, 1995); the *content* of beliefs within the model, however, comes from theoretical work on mediation processes (with special attention to the work reported in Bercovitch & Rubin, 1992). In this paper, I develop the theoretical application to the domain, build a crude instantiation of the model, and conduct basic demonstration runs.

Prepared for the annual meeting of the International Studies Association, February 16-20, 1999, Washington, D.C. This work is part of the *Methodology in International Relations: Multiple Paths to Knowledge* project, organized by the Division of Social Sciences and the James A. Baker III Institute for Public Policy, Rice University, and the Program in Foreign Policy Decision Making, Texas A&M University. I thank these institutions and the organizers of the project for their support.

Whether human beings are prone to conflict by nature or by nurture, there is no denying the ubiquity of dispute in human interactions. An excellent indicator of the pervasiveness of conflict at all social and institutional levels is the explosion of interest in mediation.[1] Outside intervention in disputes has become a critically important component of conflict management. Moreover, mediation has become a central topic of research in the social sciences.[2] Among the most significant areas of systematic research on mediation is the area of international conflict resolution (see surveys and original theoretical treatments in Bercovitch & Rubin, 1992; Fisher, 1997; Princen, 1992; Zartman, 1978; Zartman & Berman, 1982).

Jacob Bercovitch, in summarizing and reviewing this research, notes the inherent complexity of the study of international mediation.[3] A "mediation system," in his conceptualization, includes at least four components: "(a) parties [to the dispute], (b) a mediator, (c) a process of mediation, and (d) the context of mediation" (Bercovitch, 1992: 7). Each of these components is itself potentially quite complicated. The process of mediation, for example,

---

[1]A recent web search using AltaVista (http://www.altavista.com/) turned up 263,130 pages devoted to the topic of mediation. Most of these were pages advertising mediation services, ranging from family mediation, through legal mediation, to even include international mediation.

[2]For example, "conflict resolution in mixed motive games," a field which builds largely on the work of social psychologist Morton Deutsch, was listed as one of the "major developments" of social psychology in a recent authoritative review (Jones, 1998; see also Levine & Moreland, 1998; Pruitt, 1998).

[3]Bercovitch points out that practitioners of mediation, in a classic art vs. science argument, have generally been skeptical that the processes of mediation would prove tractable for systematic analysis. This attitude is, of course, well-known to quantitative and formal IR scholars in many subfields.

involves an effort by the mediator to "*affect or influence [the] perceptions and behavior [of the disputants], without resorting to physical force or invoking the authority of the law*" (7).

This focus on *problem representation* or problem definition invites an explicitly psychological treatment, which may seem quite complex and messy, especially to scholars who prefer to simplify or even ignore the internal processes of political actors in order to focus on the situational determinants of behavior.[4] Elsewhere, I have suggested that formal analysts in international relations need to keep sight of the critical components and processes of the real-world systems they seek to model, even if those processes are dauntingly complex, like the ones that Bercovitch describes (Taber & Timpone, 1996a, 1996b). In those papers, Rich Timpone and I also discuss a class of formal methods — computational modeling — that I believe offer significantly greater expressive flexibility than other formal methods, perhaps even enough to tackle problems as complex as mediation. Following that theme, this paper will report progress on the development of a computational model of international mediation, based on theories drawn from psychology and international relations.[5]

I begin with an abstract discussion of the theory that underlies the Artificial Mediator (AMed) and then proceed to a more detailed discussion of the model itself.

---

[4]Most game theoretic approaches to negotiation and mediation would focus on the features of the situation (the payoffs and other structural features of the game) that drive behavior.

[5]Automated negotiation has been of great interest to many artificial intelligence (AI) scholars in part because of the continuing interest in cooperation in strategic games, and the strong ties AI has with game theory (see, e.g., recent work in distributed AI: Bond & Gasser, 1988; Kreifelts & Von Martial, 1990). Sarit Kraus and her colleagues have done particularly interesting work in modeling negotiation among automated agents in a multi-agent environment (Kraus, 1996, 1997; Kraus & Lehman, 1995; Kraus, Wilkenfeld, & Zlotkin, 1995), even producing an application to negotiation in an international crisis (Kraus & Wilkenfeld, 1993).

Theoretical Foundations

At the most basic level, I conceive of the mediator and all other actors within the mediation system as information processors, receiving stimuli in the form of a stream of messages[6], processing that information using stored symbols, and behaving by sending messages. *From the point of view of the mediator,* all stimuli, including contextual factors, must be interpreted on the basis of stored symbols. That is, they must be processed through a *perceptual layer* to have meaning for the mediator. A critical part of my model, then, concerns the processes of interpretation and perception — what are generally known in cognitive science as *problem representation*.

Once they have been interpreted, the mediator must decide how to respond to messages. This is the *reasoning layer*, where the mediator must decide what message, if any, to send in response.[7] My model of U.S. foreign policy belief systems, for example, followed this basic architecture. POLI interpreted Asian events on the basis of three competing sets of prestored belief systems; once events were "understood", the model produced chains of reasoning which culminated in policy recommendations (Taber, 1992; Taber & Timpone, 1994). The key point is that the perceptual layer and the reasoning layer were kept distinct.[8]

---

[6]By "message" I mean any discrete piece of information that may be transmitted or communicated among human actors. This would include verbal or written messages, physical signals (e.g., punching someone in the nose is a message), and even subtle signals like facial expressions or tone of voice or the symbolic selection of a messenger or medium.

[7]The decision not to send a message may itself be a message.

[8]In fact, these processes were probably too distinct in POLI, operating on separate structural components and using completely different inferential processes. In later models (e.g., EVIN), these layers, while conceptually distinct, share a core architecture and even similar processes (Taber, 1999).

The Perceptual Layer[9]

How might mediators (and other actors in a mediation system) construct an initial representation of the problem facing them when they receive a message? Several theoretical frameworks have been advanced to explain interpretation and perception processes, with significant effort put into testing these frameworks in the laboratory. I will selectively review this literature, with a focus on the work of Schank and Abelson (1977, 1995), while I attempt to build a cognitively plausible model of the perception and interpretation of mediation messages.[10]

My most basic assumptions follow. They are either banal or entirely unacceptable, depending on one's "philosophy of mind" and tolerance for theoretical complexity.

***Assumption 1*: Mediators (and other actors in the mediation system) are information processing systems, whose responses to stimuli are conditioned by (1) internally stored information (knowledge and beliefs) and (2) externally presented or collected information. Thinking and reasoning are decomposable into combinations and sequences of "elementary information processes."**

***Assumption 2*: The interpretations of messages are constructed "on the fly" rather than simply recalled from memory. Perception is an inferential process.**

---

[9]This section draws heavily from Taber (1999). For more general treatments from cognitive science, see Collins & Smith (1988), Eysenck & Keane (1990), and Feigenbaum & Feldman (1995); for political science applications, see Sylvan & Voss, 1999.

[10]A "complete" theory of interpretation by a mediator would have to explain several levels of information processing that I ignore in the current treatment. For example, in this work I will not consider many basic issues of language comprehension; the Artificial Mediator (AMed, as it is affectionately known) will not be able to parse and interpret raw English text. I am interested in how mediators interpret the messages they receive; the range and form of messages will be constrained for the time being to a rather limited set vocabulary. I anticipate that the final message set will be coordinated with the mediation "language set" used by Wilkenfeld et al., Geva, and Schrodt in their parallel projects. For now, I have developed my own rather primitive "language" for AMed.

Interpretation is the process of building a mental model of the external world. In the words of Riesbeck and Schank (1989: 3), "An understander of the world is an explainer of the world. In order to understand a story, a sentence, a question, or a scene you have witnessed, you have to explain to yourself exactly why the people you are hearing about or viewing are doing what they are doing." So interpretation is *developing causal stories* to explain a message (Schank and Abelson, 1995).

Assumptions 3, 4, and 5 lay out the basic theoretical structures of memory and knowledge.

***Assumption 3*: Mediators (and other human actors) have a long term, associatively-organized memory (LTM). More activated portions of LTM have a greater probability of affecting conscious information processing than less activated portions. Activation spreads through memory in response to informational stimuli according to the nature and strengths of the associations in memory. Associations in LTM decay through time. LTM contains both semantic and affective memory.**

***Assumption 4*: Mediators have a limited capacity working memory (WM), in which all conscious information processing occurs. WM is transitory and can only accommodate serial processing.**

In this theory, WM is the site of conscious processing. Information must be retrieved from LTM or from the environment to affect processing. Mediators are boundedly rational because of the bottlenecks of WM. LTM, on the other hand, is relatively permanent and unlimited and may be capable of parallel processing.

Following what has become the standard form, LTM can be expressed as a network of linked nodes, where each node represents a bit of conceptual knowledge or experiential memory. Moreover, in addition to its *semantic content*, each node carries an *affective tag* to represent the

direction and strength of positive or negative feelings about the node concept or experience.[11] There are also different types of links representing different types of relationships among associated concepts, allowing the representation of hierarchical or causal relations, for example.

Node activation represents the degree of "energy" a node currently possesses as the result of hearing, seeing, or thinking about the node concept. The probability of a node being "copied" from LTM to WM where it may affect conscious processing is a probabilistic function of its current level of activation. Activation is distributed across the network according to four basic rules. First, the *working memory rule* says that LTM nodes that are currently in WM receive some small continual activation. Second, according to the *processing activity rule*, LTM nodes in WM receive a "sharp jolt" of activation when they become part of current conscious processing. Third, the *fan rule* asserts that activated nodes in LTM spread activation to directly linked nodes, according to the strengths of the links. And finally, activation in LTM *decays* rapidly.[12]

Node strength corresponds to the "degree of belief" in the concept represented by the node. In some sense, it is the residue of prior activation. It does not decay through time, but increases and decreases as a function of its history of activation and how "useful" it has been in earlier interpretations (as indicated by the processing of feedback information in WM, discussed below). Since it is possible for memories to be strong but hard to find, there is a separate *link strength*, analogous to node strength, which is a function of the amount of activation that has spread across the given link in the past and how useful the link has been in earlier interpretations.

---

[11]Semantic content refers to *meaning*; affect refers to feelings or evaluative responses.

[12]There is one additional rule that affects the spread of activation in LTM — the partial matching rule — which is discussed below as assumption 7 because of its importance.

To summarize, a decision maker's prior knowledge is the set of nodes, each containing semantic and affective values, and each carrying a fleeting activation level and a more enduring node strength, and the links among those nodes, each representing a type of association and carrying a strength of association.[13]

***Assumption 5*: Memory objects, represented as nodes in LTM, may be singular concepts, as described above, or bundles of tightly associated knowledge, called schemas. A schema occupies the same "space" in WM as a single node.**

There are many types of schemas, including general conceptual categories (like "dictator"), particular instances of a category (like "Stalin"), general event sequences or scripts (like an "aggression script"), and actual historical cases (like "the US invasion of Panama").[14] As far as LTM processes are concerned, a schema behaves exactly like a singular node.[15] But schemas have important consequences for conscious processing in WM: (1) they greatly improve cognitive efficiency and allow one to overcome to some degree the space limitations in WM; (2) since they bundle together related information, they greatly improve the efficiency and power of inferential reasoning;[16] on the other hand, schematic information processing may lead us to make

[13]To tie this back to earlier language, LTM contains the stored symbols that the mediator can use to process input messages.

[14]For discussions of schematic processing in political cognition, see Conover & Feldman (1984), Larson (1994), Lodge & Hamill (1986), and Miller, Wattenberg, & Malanchuk (1986).

[15]At this early stage of theoretical development, only the schema label (or *header*) is visible to LTM processing. See below for clarification.

[16]Schank's famous example of a restaurant script comes to mind: upon entering a restaurant, most people know what to expect and how to behave; the entire package of expectations comes to mind together.

incorrect inferences because they may be oversimplified or inappropriate for the current message.[17]

***Assumption 6*: Processing is subject to motivational biases introduced by accuracy or directional goals.**

A classic criticism of cognitive science is that it pays insufficient attention to motivational or affective factors. Traditionally, affect has been left out of an already complex picture of human cognition. But this means that a variety of phenomena important to political information processing cannot be explained in these theories. A number of scholars have recently become interested in motivational factors (Kunda, 1990; Lodge & Taber, 1999; Taber, Lodge, & Glathar, 1999). In the current treatment, affective tags exist in LTM to store the evaluative feelings that people construct automatically as they are exposed to new information.

The theory of *on-line (OL) processing* holds that people form evaluations spontaneously upon exposure to information about a concept (e.g., "communism" or "Saddam Hussein") and then immediately integrate the affective charge into a running evaluative tally for the concept (i.e., the affective tag for the given concept node). Once the evaluative tally has been updated, the actual information may be forgotten, though it need not be. In other words, the information may not be stored in LTM, though its evaluative impact will alter the affective tag (Lodge, McGraw, & Stroh, 1993; Lodge, Steenbergen, & Brau, 1995). These affective tags or "charges" prove important in the motivated processing of subsequent information.

In the conceptualization I have adopted, motives fall into two broad categories: *accuracy goals*, which motivate people to reach a correct or otherwise normatively optimal interpretation

[17]Racial stereotypes, for example, are undoubtedly represented in LTM as schematic categories.

(sometimes labeled the "intuitive scientist" mode) and *directional goals*, which motivate the individual to justify a specific, preselected interpretation (sometimes labeled the "partisan" mode). Directional goals arise when LTM nodes are retrieved into WM, automatically bringing their affective content along with their semantic content. If this affective content is strong enough, it may create powerful biases in the interpretation process.[18] For example, an "international aggressor" schema may carry strong negative affect. If it is drawn into WM early in the process of interpreting the U.S. invasion of Panama, for example, it may establish a directional goal to construct a negative understanding of the event. A "police action" schema, by contrast, may lend a far more positive interpretation of the same event. On the other hand, if one is strongly motivated to be accurate and even-handed, these directional biases may be overcome.[19]

***Assumption 7*: In addition to the activation rules mentioned above, LTM nodes that partially match items in WM receive a jolt of activation.**

Schemas in LTM that are not identical to but share features with structures in WM may also receive activation. This is an important part of the inferential process, allowing reasoning from metaphor or analogy. For example, many people understand Saddam Hussein by reference to a Hitler schema (Spellman & Holyoak, 1992). For these people, the pattern of information or

[18]The mechanisms through which directional goals can bias the construction of interpretations operate primarily on memory search. If one is prone to interpret a message negatively, then information that reflects negatively on the message will have a better chance of retrieval; information inconsistent with one's directional goals may be unconsciously "censored" by biases in the spread of activation.

[19]A great deal of research suggests that it is very difficult to fully overcome directional bias and arrive at an "accurate" interpretation. Strong accuracy goals may lead one to overcorrect one's biases, for example.

messages they receive about Saddam Hussein match *to some degree* the pattern of features in the Hitler schema.

In this theory, pattern matching is a function of four factors. First, the degree of similarity between the semantic content in WM and schemas in LTM (e.g., the similarity between a message and a schema in memory). Second, the matcher looks first to strong potential matches — it is guided by node strength. Third, the matcher favors schemas with affective implications consistent with existing directional goals. And fourth, strong accuracy goals deepen the search for matches by allowing even weaker pattern matches to qualify.[20]

***Assumption 8*: When motivation (either directional or accuracy) is strong and WM holds "new" information, old memory structures may be modified or new memory structures may be assembled in WM and transferred to LTM.**

Four types of learning take place in this theory. First, the strengths of nodes and links change with the history of activation; activation passing through nodes and links increases their strength. Second, existing associations among nodes in LTM may be actively altered because of information in WM. Third, new associations might be learned. And fourth, new schemas might be constructed in WM and linked into LTM.[21]

These eight assumptions form the core of my theory of perceptual processing. To summarize, when a message is received, the mediator must construct an understanding of that message. It does so by making inferences about the message through the various processes that

---

[20]It may seem paradoxical that a concern for accuracy leads the decision maker to accept weaker matches in building understandings, but remember that the accuracy motive leads one to process more deeply, considering a wider range of possible interpretations.

[21]See Taber (1999) for a discussion of these learning processes. At this point in the development of AMed, learning processes have not been implemented.

spread activation to nodes in LTM, including through partial matching and motivated memory search. Ultimately the interpretation takes the form of an elaboration on the original message, so that the "sender", "target", and "contents" of the message, for example, will be understood as collections of attributes and connections from LTM. As we will see in the next section, many of the same processes and structures underlie reasoning in the model.

The Reasoning Layer

Once an interpretation (or perhaps, several competing interpretations) has been constructed for an input message, the mediator must decide how to respond. This response will take the form of a message (or no message, or several messages) and is selected on the basis of rules of strategy and goals at two levels: (1) general goals of the mediator toward the mediation situation, which change little if at all during the course of mediation (e.g., the basic goal of "getting to an agreement");[22] and (2) subgoals specific to the particular message being processed (e.g., "break this proposal down into simpler components"). The basic cognitive architecture for the reasoning layer is shared with the perceptual layer — that is, processing operates on the same LTM and WM structures — though I will now highlight several different features of memory structure and a new set of inferential processes.[23]

---

[22]It is important to note that mediators are not simply altruistic actors who help resolve disputes out of the goodness of their hearts. They are fully strategic actors with their own sets of interests, which are fully represented in LTM.

[23]In human cognition, there is almost certainly no clear divide between processes of interpretation and reasoning, with the two occurring simultaneously. I find it useful conceptually, however, to separate the two sets of processes temporally, so that the mediator waits upon an interpretation before beginning to decide how to respond.

Broadly speaking (and simplifying), perception as described above is the process of classification using *declarative knowledge*. A mediator, for example, might interpret a particular threatening message as a "bluff" if the features and structure of the input message match stored examples of earlier bluffs.[24] That is, the message is classified as an instance of a bluff because it corresponds sufficiently to the category knowledge stored in LTM. Reasoning, by contrast, is largely procedural, using *implicational beliefs* to derive a conclusion. In *expert systems* like POLI, implicational beliefs are stored in a knowledge base (roughly equivalent to LTM) as an unstructured list of IF-THEN production rules. But they can also be represented efficiently (and more plausibly in terms of research in cognitive psychology) in structured associative networks like our LTM described above. We simply need to include *implicational links* among our theoretical primitives.[25]

Here, one node or cluster of associated nodes — say "message isa bluff" — is linked to another node or cluster of nodes — say "call the bluff" — by an implicational link. These linked nodes now represent the implicational belief (or rule), IF "message isa bluff" THEN "call the bluff". Now if strong activation is sent to the first node(s) because the message has been interpreted as a bluff, it will spread across the implicational link to the conclusion, "call the bluff", and if activation rises above threshold, the conclusion will then be deposited in consciousness

[24]This type of schematic processing is called case-based reasoning.

[25]By theoretical primitives, I mean the basic components or building blocks that may compose a given model, in this case the types of links the model can operate upon.

(WM) as an implication.[26] Chains of such implications may "fire" until we have traveled along full lines of reasoning to a final conclusion (a message or messages).[27]

Another mechanism exists for such implicational reasoning within the architecture specified above. Recall that schemas (or scripts or cases) are packages of related beliefs. Such belief bundles are quite likely to contain implicational/procedural knowledge. For example, it is quite likely that a general schema for the concept "bluff" will exist in any experienced mediator's LTM, which will contain basic classification information (perhaps a listing of prior cases, e.g.) *along with* beliefs about how to respond. That is declarative and procedural beliefs are packaged together in most well-developed schemas. Just as most people know what to expect and thus what to do upon entering a restaurant (from information stored in a restaurant script), an experienced mediator will know how to respond to a bluff (from information stored in a bluff schema). Indeed, there is a great deal of evidence that sophisticates (people with domain expertise) reason on the basis of stored theories, schemas, and cases. Their reasoning, while more finely tuned to the domain, tends to be quite "recipe-driven".[28]

---

[26]This represents *forward chaining*. If the implicational link is traversed from conclusion to antecedent, it would represent *backward chaining*. See Taber & Timpone (1996b) for a discussion of this distinction.

[27]There is much complexity here that I must gloss over, but it is worth noting that since WM is of limited capacity, some of the links in the argument may be lost (forgotten) even if the final conclusion is eventually reached. This represents quite nicely the inability of human reasoners to articulate fully the "logic" behind their behavior.

[28]This is one reason experts are more able to articulate their lines of reasoning: since it is more efficiently packaged more of the information used on the chain of reasoning can be kept in WM.

Reasoning, in this theory, is the process of constructing a chain of implications using pre-stored implicational beliefs. Of course, just as the perceptual layer is subject to motivated biases, analogical reasoning through partial matches, and learning processes, so too is the knowledge used in reasoning. As an obvious example that is "hard-wired" into this two-layer conceptualization, reasoning can only proceed on the basis of an interpretation of the message; so it is in some sense "biased" even before it begins.[29]

Theories of Mediation

To this point, the theory described is not really a theory of a mediator per se. Nothing important distinguishes the processes I have described from human information processing in many domains.[30] But as I have already suggested, there is a large literature on mediation, which includes both sophisticated theoretical treatments and practical primers. I will not, in this paper, review this literature (see Bercovitch & Rubin, 1992). Rather, I will more generally discuss how these theories can be incorporated into an Artificial Mediator.

My theoretical expectation (bias, perhaps) is that there is nothing intrinsically different about the basic cognitive architecture and information processing of experts in just about any domain. The significant differences, I believe, live in the details, in the actual content of LTM.

---

[29]I have ignored here several important questions, including stopping rules. When does the mediator decide that an interpretation (or set of interpretations) is adequate and move on to the reasoning layer, and when does the mediator decide to stop the reasoning phase and spit out a response?

[30]And, indeed, the perceptual layer described is essentially the EVIN model of the interpretation of foreign policy events (Taber, 1999). And it wouldn't take much more than a global search and replace on the word mediator to turn this into a model of political candidate evaluation (see Lodge & Taber, 1999).

And the reader will note that I have not discussed content at all up to this point, except in giving simple examples. My model of an Artificial Mediator (AMed), then combines theoretical insights developed in cognitive psychology (and to a lesser extent in political cognition) with theoretical insights from the mediation literature. The former provide basic cognitive architecture and processing; the latter provide the knowledge and beliefs that will be used in processing.

For example, explicit in Bercovitch's (1992: 7) description of mediation processes — efforts to "*affect or influence [the] perceptions and behavior [of the disputants], without resorting to physical force or invoking the authority of the law*" — is the idea that mediators will try to manipulate the problem representations of the disputants. That is, mediators attempt to manage information (messages) so that each of the disputants will arrive at an agreeable (to the mediator's goals) perception of the mediation system. This can only happen by paying explicit attention to the perceptual layers of the disputants and trying to structure the flow of messages so that a favorable definition (or possibly definitions) of the situation is arrived at by all parties.[31]

It is also clear from Bercovitch's discussion that we would make a serious error in assuming that mediators are different in kind from the other participants in a mediation system. All actors, mediators included, come to the negotiation table with a full baggage cart of interests and biases. The particular interests, goals, and biases of the mediator are represented in this model within the beliefs and knowledge of LTM. That is, any biases or interests that exist vis-a-vis other relevant actors will find "expression" in the way messages and the very mediation situation are represented by the perceptual layer.

---

[31]Peter Bennett (1995) examines a related concern when he uses drama theory and the INTERACT software to "translate" between the problem representations of disputants in northern Ireland.

An important feature of the mediation task, which it shares with some but not all other domain tasks, is that *the mediator must construct interpretations on two levels simultaneously*, using the processes described above. The mediator starts with an understanding of the overall mediation system and progressively modifies that understanding on the basis of the message stream. In addition, the mediator interprets each individual message as it is received.

In the following sections, I will elaborate on my model of the above theory and then illustrate the theory by instantiating the model with a simplified set of propositions about mediation drawn from the Bercovitch and Rubin volume (1992).

The Artificial Mediator (AMed)

When one wishes to build a formal model of a complex process theory like that described above, the tools of computational modeling can be very helpful (Taber & Timpone, 1996a, 1996b). Broadly speaking, a computational model is a theory rendered symbolically as a computer program so that its behavior may be explored through simulation. In this section of the paper, I describe the on-going process of building the Artificial Mediator (AMed).

In the rarified world of the AMed, messages take on a highly stylized form,

<Initiator><Content><Target>,

where all "slots" may only take on values from a limited, pre-defined set. The <Initiator> and <Target> slots may refer to the actors participating in the current mediation situation as well as any other outside actors that the analyst anticipates will become important to the mediation situation during the process of negotiations (e.g., domestic actors within a participant's

organization or third parties with an interest in the negotiations).[32] The <Content> slot is defined by a set vocabulary similar to event data codes (e.g., the WEIS codes, which include such items as Yield, Consult, Approve, Promise, Propose, Reject, etc.) augmented by additional codes that tailor the general codes to the mediation domain (e.g.,"Propose equal split").[33] Note that a message need not include all elements. For example, a given message — "Serbs reject the Kosovan autonomy proposal" — might become

<Yugoslavia><Reject autonomy><Kosovo>.

Or, "Kosovo militants open hostilities against the Yugoslav army" would be rendered

<Kosovo militants><Military engagement><Yugoslavia>.

This research choice for a limited vocabulary forecloses some of the flexibility of natural English input, but seems reasonable given the complexity of the theory as it stands.[34]

AMed, then, is designed to take a message as input, process that message according to the theory developed above, and return a message (or set of messages) as output.

AMed consists of four major components: (1) the *knowledge base* (representing LTM), (2) *working memory*, (3) the *executive* (which makes all processing control decisions), and (4) the

---

[32]In reconstructing an actual real-world mediation, we know in advance who these will be. But if the model is used predictively (rather than postdictively) we must anticipate what actors may be relevant.

[33] Note that the specific vocabulary of a given instantiation of AMed depends to some degree on the theory of negotiation being represented.

[34]An alternative, which I am thus far resisting, is to attempt to incorporate one of the several well-developed ontologies, or symbolic systems of meaning, that linguists play with (e.g., Doug Lenat's CYC; for more information, check out http://www.cyc.com/).

user interface. The user interface holds little substantive interest and as such will not be discussed in this paper.

The Knowledge Base

AMed's knowledge base contains the sum total of AMed's prior knowledge about its environment (knowledge about all mediation participants, their history of interactions, and about mediation strategies, e.g.). Two distinct types of knowledge are recognized by the system, declarative and procedural. Declarative knowledge is "knowledge that" — one's understanding of concepts and conceptual relations. An example of declarative knowledge would be the statement that "Bosnia is in Europe". Procedural knowledge, on the other hand, is "knowledge how" — one's understanding of processes and procedures. When faced with international aggression, an example of procedural knowledge could be "to impose sanctions, get the backing of the U.N. Security Council."

Without structure, knowledge resembles a library without a catalog system. It is nearly impossible to find anything. Assumption 5 states that knowledge may be structured as either nodes of information or as bundles of information called schemas. Nodes represent singular concepts (e.g., Bosnia might be a node to someone who has not yet developed a detailed understanding of Bosnia) and contain semantic information (e.g., Bosnia) and affective information (e.g., vague negativity toward Bosnia). Nodes contain two additional variables, one indexing node strength (a product of the history of activation and utility of the node) and one indexing current activation (how "energized" the node is by current thought processes). To

operationalize the schema construct, I have adopted an associative network formalism that mixes nodes and *frames*.

Frames were first suggested by Minsky (1975) and defined as "a data structure for representing a stereotyped situation." A frame is a knowledge structure that represents some object and lists information about that object. It is an efficient way to operationalize the notion of schemas. A frame system is a linked group of frames. This allows one frame to reference information in another. Information regarding the object, either procedural or declarative, may be placed within the frame.

Each frame consists of two main constructs. First, each frame has a *header*. The frame header contains five variable definitions: the name of the frame (its semantic tag), the type of the frame (discussed below), a summary affective tag for the entire frame, a numerical index for "frame strength", and a numerical index for current "activation strength." As nodes are simply frame headers without "bodies" (that is, without a list of further information about the concept), I will henceforth use the term frame to refer to both frames and nodes. Knowledge-base (LTM) processing (spreading activation of various forms) operates only on the header, so there is no reason to distinguish nodes from frames at this level.

Frame strength is a function of how often and how much the frame has been activated in the past and how useful the frame has been in interpreting past events.[35]

$$fstr_i(t) = \varphi_i fstr_i(t-1) + u_i(t) \qquad [1]$$

[35]The following presents a modified version of a model by Boynton & Lodge (1994), which was itself based on Anderson's (1983) ACT*.

where $fstr_i(t)$ denotes the strength of frame *i* at time *t*, α_i is a parameter controlling the mediator's ability to remember,[36] and $u_i(t)$ is the current strengthening input of frame *i* at time *t*.

$$u_i(t) = f_1[act_i(t-1), learning] \quad [2]$$

where f_1 is a monotonically increasing function, act_i is defined in equation [3], and learning is still on my wish list.[37]

Activation strength is the level of "energy" of the frame due to current thought processes. For both singular nodes and frames, current activation is a function of the sources of new activation (linked frames and nodes from which activation may be passed or direct activation from related items in WM) and frame strength, less a decay factor.

$$act_i(t) = \beta_{i1}act_1(t-1)+\beta_{i2}act_2(t-1)+...+\beta_{ii}act_i(t-1)+...+\beta_{in}act_n(t-1)+v_i(t)-\gamma_i(t) \quad [3]$$

for *i*=1 to n, where n is the number of frames in the network, $act_i(t)$ is the activation level of frame *i* at time *t*, β_{ij} is the strength of the link from frame *i* to frame *j* (typically not equal to β_{ji}), $v_i(t)$ is the direct input of activation to frame *i* at time *t*, and $\gamma_i(t)$ is the decay factor.[38]

$$v_i(t) = f_2[a_i, b_i, c_i] \quad [4]$$

---

[36]This parameter is set to .9 on the basis of generalized experimental findings (Boynton & Lodge, 1994).

[37]Since learning is not instantiated in the current version of AMed, equation [1] implies that frame strength can only increase. Consequently, the actual function for frame strength also includes a decay parameter in the place of learning to keep node strength from growing unreasonably large.

[38]There are many forms that decay can take. For current purposes, I will set it as a constant parameter, which may be manipulated experimentally.

where f_2 is an increasing linear function, a_i is the direct activation of frame *i* due to conscious processing, b_i is the continuous buzz of activation to frame *i* due to its presence in WM, and c_i is the activation of frame *i* due to a partial match with a new piece of information in WM.

The probability that knowledge (nodes or frames) will be brought into WM is a direct function of its current activation.[39]

Second, a frame has *links* to other frames. Links allow activation to spread between frames; link strength is operationalized in AMed in the $\$_{ij}$ matrix, where $\$_{ij} \$ 0$.

$$\$_{ij}(t) = f_3[\$_{ij}(t\text{-}1), \$_{ij}(t\text{-}1)act_i(t\text{-}1), learning] \quad [5]$$

where f_3 is an increasing monotonic function. Note that $\$_{ij}(t\text{-}1)act_i(t\text{-}1)$ is the amount of activation that passed along this link from *i* to *j* in the prior period.[40]

A second matrix of size n×n, J_{ij}, contains entries that define the *types* of links. In the current version, these include classification links (subclass, superclass, instance), association links (direct, inverse), logical links (and, or), modifier links (mod), and implicational links (IF-THEN).

As mentioned above, frames are linked together in a system. For conceptual clarity I distinguish three different types of frames, the *class*, the *instance*, and the *event*.[41] A class is a broad description of entities that exist in the world. As an example, WORLD LEADERS would be considered a class and represent the stereotypical world leader. In the WORLD LEADERS frame, links might be provided to specific world leader frames, such as CLINTON or MAJOR.

---

[39]Most typically we use some form of logistic threshold function.

[40]Boynton and Lodge () suggest that link strength be a function of the ratio of activation passed along the given link from *i* to *j* to the sum of activation passed to *j* at *t*-1.

[41]Frame type is listed in the header.

Such specific frames would be called *instances*. *Case* frames are similar to instance frames, but instead represents specific events. Thus, the Russian occupation of Chechnya may be a case belonging to the event class CIVIL WAR. Frames that represent more general categories are superclasses while more specific categories are subclasses. Of course, important linkages that do not represent hierarchical classification knowledge may be drawn in the system. That is, an object that is an instance or class object in relation to one memory element may hold modifier or implicational relations with another memory element. To replay my earlier example, the assertion "the message is a bluff" links the message object as an instance of the class of bluff messages. This memory snippet may also be linked implicationally with the memory object "call the bluff", which itself may be an instance of a larger set of behavior types. But the two objects do not have a classification relation.[42]

Each frame is a more or less detailed description of the object it represents. So the second feature of frames is the representation of this descriptive information in slots. Slots contain three parts, the slot-name, the slot-filler, and a certainty factor (CF). Slot-names identify the type of information held in the slot. Slot-fillers are the specific pieces of information that go into each slot. Stretching conventional notions a bit, slot names are variable names (attributes) and slot fillers are values on those variables. A Hitler frame, for example, might contain a slot for who (German leader), what (dictator), where (Germany), why (Mad) when (1930s-40s). Note that more than one slot of each type can be associated with a frame and not every slot must be filled. While the exact contents of the slots in a particular frame may be unique, there will be a strong

[42]Undoubtedly, for an experienced mediator, this implicational belief would be included within the Bluff object rather than linked as a separate frame.

resemblance between the super and sub class frames as well as their related instances. In fact, when a new frame is created, the slot-names and slot fillers are copied from the relevant super class. Thus the more specific frame *inherits* many slot-names and slot-fillers, perhaps overriding "default" information with particular knowledge about the instance or case. When one first reads about Hitler, one is likely to copy some existing class frame (e.g., political dictator), only modifying information when specific knowledge is learned about Hitler that differs from defaults "borrowed" from the superclass. This notion of inheritance is of key importance. In effect, if AMed knows very little about a particular subclass or instance, it bases its decisions on the superclass frame.

Decision makers have varying levels of belief strengths depending on previous experience. AMed captures the belief strength of information in the certainty factors (CF) associated with each slot. CFs can take on values from -1 to 1 with higher numbers representing a stronger belief that the information in a given slot is true, 0 neutral, and negative values representing beliefs that the information is not true. Readers who wish a more detailed exposition of frame knowledge representation systems are directed to Taber and Timpone (1996b).

Working Memory

WM in AMed is a scratchpad for the temporary storage of knowledge objects while they are being processed. In cognitive terms, it represents a place where conscious information processing occurs. The theory requires that WM be severely constrained. It is of limited size, only holding 7 knowledge objects (nodes or frames) at any time. This requires a "control decision" for which memory objects will remain in WM and which will be displaced when new

pieces of information reach threshold levels of activation, which we discuss below. Processing in WM is serial, meaning only one processing step can occur at a time.

The Executive

Our library now has a system to catalog its holdings. It now needs some management. All management of the knowledge base is performed by the executive subsystem. The executive has six functions: (1) it is responsible for identifying a set of appropriate frames as candidates for admission into working memory; (2) it selects the information from this set that will enter WM; (3) it decides which objects in WM will remain and which will be displaced; (4) it manages the addition, deletion, and editing of frames and slots from the knowledge base;[43] (5) it controls the inference processes on WM at the reasoning layer;[44] and (6) it decides when interpretation and reasoning stop.

For identifying the appropriate frames for entry into WM, AMed uses the input message (<Initiator><Content><Target>) to initiate a search (using spreading activation). The message is placed in working memory as an uninterpreted object (the *message stimulus object* in my parlance). The three corresponding frames (if they exist) receive immediate activation from a partial match as defined in equations [3] and [4]. If this activation rises above threshold (which it typically will), all three memory objects are pulled into WM where they are linked into the stimulus object. That is, the <Initiator>, <Content>, and <Target> components of the stimulus object are "endowed" with the information contained in the corresponding memory objects. From

[43]Learning is not yet implemented.

[44]Here I this is forward chaining.

here on, they continue to receive activation from their presence in WM and from their participation in conscious processing. The higher the need for accuracy (at this point, I model this as an exogenously determined parameter), the higher the level of activation "poured" into the network.[45] Activation then spreads automatically through the knowledge base network according to the rules specified in the theory (equation [3]): (1) all objects in WM send a small "buzz" of activation to their corresponding LTM objects; (2) all objects undergoing current processing in WM (this will repeatedly include the stimulus object) receive a "sharp jolt"; (3) partial matches from the pattern matcher receive activation; (4) activation spreads to all frames and nodes linked to those that have received activation in a fan effect, according to the strength of the links; and (5) activation decays, both as it moves out from the originating nodes and through time for all nodes and frames. As activation spreads from the "points of entry" into the knowledge base, it diminishes at a constant rate determined by the decay factor. The part of the executive that keeps track of the activation level of each frame is called the *activation manager*.[46]

Once activation has ceased to spread in the current round (that is, once equation [3] has been computed for all frames), the executive must decide which (if any) frames and nodes to bring into WM. First, it checks to see how many WM positions are immediately available. WM positions may be available because we are in the early stages of processing this message, because the executive has determined that some current existing objects in WM are not useful, or because the information from some object has already been added to the growing stimulus object (see

---

[45]This activation is poured in at the frames that are now in WM.

[46]In practical terms, it is simply a data structure (a vector of size n) that contains the numerical index for $act_i(t)$, where $i = 1$ to n.

below). The executive then probabilistically selects the "new" contents of WM on the basis of the current activation levels of all possible objects, including those currently in WM (in other words, this is another opportunity for WM objects to be displaced).

Note that as this overall process continues, the message stimulus object in WM becomes progressively more detailed and provides a richer (though not necessarily more accurate) understanding of the message. The most difficult question here is how patterns are matched. In general, AMed's pattern matcher will calculate a "match score" for each object in WM based on the number of matching slots. For example, the Hitler frame may provide a match to the Hussein frame for someone trying to understand Saddam Hussein's invasion of Kuwait if "enough" of the slots are common to both. If the slot matching process does not produce a good "fit", each frame in working memory, including the original stimulus objects, may be probabilisticly pushed out of WM.

Frames that have contributed to the stimulus object receive a continuing "buzz" of activation as long as AMed continues processing the stimulus object (that is, as long as the model keeps trying to understand the message). So even after an object has been returned to LTM, it may continue to receive some "working memory activation" if some of its slots have become part of the growing stimulus object.

Frames and nodes are released from WM once all relevant information (at this point everything that is transferred to WM is automatically considered relevant) is transferred to the stimulus object. This opens up slots for future processing.

As the stream of messages is processed, AMed also constructs a representation of the overall mediation system. Here the elements that need to be elaborated are the motives and

resources of the actors, including the mediator, and any proposals that are currently on the table.[47] In effect, we have two stimulus objects in WM at all times: the interpretation of the current message and the interpretation of the larger system. Processing proceeds as described above for both stimulus objects simultaneously. That is, each processing round, activation pours into LTM triggered by *both* processing activities.[48]

Instantiating the Model

I have specified AMed's knowledge structures and cognitive processes in some detail but I have left open the question of the actual content of the knowledge base (among several open issues). Without some basic knowledge, AMed would be left to interpret its environment in a vacuum. As I suggested above, the knowledge within AMed needs to come from the domain expertise of practitioners or theorists of international conflict mediation.[49]

In this paper, I will merely illustrate in a stylized way the workings of the model. I have chosen the general description of mediation strategies from Bercovitch (1992), adapted from

[47]Below I will make a distinction between public and private motives, which is particularly important to the mediator, who will only report public motives.

[48]This is important, for it means that there will be important dependencies between the construction of interpretations on the two levels. Another consequence of this two-part task is that there are only 5 slots in WM for the executive to play with. This provides realistic constraints on the ability of the mediator to deal with the complex situation.

[49]I could, in the spirit of knowledge engineering, extract the beliefs of an individual mediator and represent them in AMed — Kissinger in a box, for example. But I have chosen to instantiate the model with theoretical knowledge expressed in systematic work on the subject.

Touval and Zartman (1985) as my theoretical source.[50]

Touval and Zartman (1985) develop a three part classification scheme for mediation strategies: (1) *Communication-facilitation strategies* (make contact with parties, gain the trust and confidence of the parties, clarify situation, supply missing information, etc.); (2) *Formulation strategies* (choose meeting site, establish protocol, suggest procedures, deal with simple issues first, etc.); and (3) *Manipulation strategies* (change parties' expectations, make substantive suggestions and proposals, supply and filter information, promise resources or threaten withdrawal, etc.). Bercovitch (1992) suggests ways in which the nature of the situation (particularly the resources available to the mediator) helps to determine the strategies that are selected. For example, the more coercive strategies will be available only to mediators that have coercive resources at their disposal. In terms of my framework, the nature of the situation is perceived as part of the problem representation established through the interpretation process. That is, the relative power of the involved actors and the resources available to them impact behavior only after passing through the mediator's perceptions.

This framework linking likely mediation strategies to different types of (subjectively perceived) situations provides at least the beginnings of knowledge content for AMed. A simple semantic network was created with three strategy frames, one for each of the three classes of strategies identified above. In addition, a mediator resource frame was created to contain the six types of resources identified in Bercovitch (1992: 20-21; based on French and Raven, 1959; Raven, 1990): (1) reward resources; (2) coercive resources; (3) referent resources; (4) legitimacy

[50]As we will see, this system, which I have elaborated somewhat to fill in detail, is rich enough to generate basic behavior, though not rich enough to provide the basis of meaningful experiments.

resources; (5) resources of expertise; and (6) informational resources. This mediator resource frame corresponds to self-perception on the narrow dimension of resources (and also motives as described below). I conceive resources, like power, to be relational, relative to other relevant actors. So the mediator's perceptions of her own resources may vary depending on the perceived resources of the other relevant actors. In one mediation context, therefore, a mediator may not feel he has sufficient coercive or reward resources to perform some of the manipulation strategies. The situational dependencies of referent resources (which stem from a "sense of mutual identity between a mediator and the disputing parties" [Bercovitch, 1992: 20]), and legitimacy, expertise, and informational resources are even clearer. In AMed mediator resources, as subjectively perceived for the current mediation system, constrain what strategies are likely to be selected in the reasoning phase in a manner generally suggested by Bercovitch (1992).[51]

Of course, this provides only the roughest of beginnings.[52] Nevertheless, this skeleton of a theoretical basis for mediation is where I will stop in the current paper.[53] I turn now to the thorny problem of providing an environment (at minimum, a virtual environment) within which AMed

---

[51]For example, Bercovitch (1992) suggests that mediators who have only referent, legitimacy, or informational resources will be limited to communication facilitation strategies. Those that have all of the types of resources (perhaps because they have the backing of a powerful state) have all strategies available to them. Unfortunately, he does not fully detail these connections, so I have been forced to draw some additional linkages between resources and strategies that seem reasonable to me. Future versions of AMed will be more carefully constructed through interaction with domain experts.

[52]For one thing, resources are likely to remain relatively unchanged during a given negotiation process, so they cannot produce changing interpretations and reasoning about particular messages.

[53]Frankly, I don't wish to take the specification of theoretical content too far until I have had a chance to consult with my colleagues on parallel projects as part of the larger project.

can act. As we will see, this also entails specifying some knowledge about a particular negotiation situation.

A Demonstration Run

One of the perplexing problems of modeling individual actors in detail is the problem of context. How does one create a meaningful environment, rich enough to "exercise" the model but constrained enough that the model (which we have a tendency to forget is still a vast simplification of human cognition) is able to interact with it. As an obvious practical matter, for example, unless the model is capable of processing natural language, all messages must be cast in a form the model can "understand". Solutions that I have considered include immersing the model in a stylized historical trace (as I did with the POLI model), placing the model as a player in an otherwise-human simulation game (as I may yet do with AMed), or placing multiple computational agents together in purely artificial interaction (as in the autonomous agent literature). For the purposes of this paper, and in the spirit of illustration and demonstration (rather than test or even plausibility check), I have decided to try here a different approach. I have created a wholly fictional historical trace of a simple mediation system, making no effort at even surface resemblance to any particular real-world dispute. AMed will act as the mediator for this hypothetical situation, acting on the basis of the general theoretical beliefs described above and some situation-specific beliefs I will make up. Other actors will act as I make them act purely to exercise the model.

The mediation system and AMed's knowledge. In my hypothetical mediation situation, two disputants — Cartha and Reno — are negotiating over ownership of a piece of territory —

Flatland — which Cartha currently occupies. AMed here represents a private individual offering informal mediation. There are no identity ties between AMed and either disputant, but AMed is accepted by both Cartha and Reno as a legitimate negotiator. In short, AMed has only legitimacy, expertise, and informational resources available (see Bercovitch, 1992).

In AMed's knowledge of the two actors, they are roughly equal in power and other resources and are "medium-sized" powers. However (just to spice things up a little), Reno has secretly developed a weapon that will give it a decisive military advantage over Cartha, and so Reno does not have the incentive to negotiation or compromise that AMed (and Cartha) will assume it has. Cartha and Reno have a long history of conflict, so AMed expects them to be fairly hostile.

This situation is subjectively represented in AMed's knowledge base as a frame for each disputant (CARTHA and RENO) with slots defining their expected resources and motives. Each of these frames is listed as an instance frame and is linked to a superclass DISPUTANT frame and a superclass NEGOTIATOR frame. Both of these are quite simple. Disputants are expected to be hostile towards each other with little willingness to compromise if there is any other way to achieve their goals. Negotiators are expected to be willing to compromise since they are assumed to value a negotiated settlement (so these expectations are in tension). The motives in the CARTHA and RENO frames are actually inferred from the class links to DISPUTANT and NEGOTIATOR. In the DISPUTANT and NEGOTIATOR frames are also examples of "diagnostic" behaviors. That is, certain types of behaviors (messages) will tend to match the

disputant type while other behaviors will tend to match the negotiator type.[54] So AMed's expectations about the basic motives of the disputants may vary as the stream of messages is processed.

AMed has the general public goal of *getting a solution acceptable to both sides* and the private goal of *increasing its own reputation as a mediator* (this will set up incentives to appear active).[55] Structurally, these goals appear in the SELF frame along with a listing of the resource types available to AMed.

AMed instantiates the theoretical knowledge described in the previous section in three strategy frames: Communication-facilitation (COMM), Formulation (FORM), and Manipulation (MANIP). These are the most "developed" frames in AMed's knowledge at this point. They contain procedural knowledge (rules) designed to operate on several kinds of perceived messages, perceived resources, and general goals and capable of generating response messages (for example, if AMed interprets an input message as a misconstrual of the mediation system[56], a rule within the COMM frame suggests "clarify the situation" as a strategy. Additional rules give this strategy operational specificity (i.e., casts it in the form of a set of messages, e.g., SAY "<current interpretation of system>").[57]

---

[54]To make the demonstration work despite the simplicity of AMed's knowledge, I admit to having "canned" these diagnostic behaviors to match the six messages described below.

[55]Several theorists note the self-aggrandizement motive for mediators (e.g., Princen, 1992).

[56]I allow AMed's motives to be designated public or private. Only public motives will be expressed.

[57]All messages that AMed can emit are indicated with the keywords SAY or DO, corresponding to verbal and physical messages. The bracketed phrase is a variable which, in this

For this run, I have restricted the message vocabulary to the KEDS Modified WEIS Codes (see Appendix).

A demonstration run. AMed's <interpretation of the system> at startup attributes mixed motives to both disputants: they are seen equally as negotiators and disputants. AMed's own public motive is to reach settlement.[58] The two disputants are seen as having moderate and relatively equal power resources. AMed has neutral affect toward both (operationalized as affective tags = 0). AMed perceives itself as having only legitimacy, expertise, and informational resources. There are no proposals on the table at startup. I traced the process through six messages, listed with AMed's responses in Table 1. Below I detail two of these responses, to the first and last messages in the stream.

— Table 1 about here —

M1: <Reno> <DEMAND "Cartha withdraw from Flatland"><Cartha>

Perceptual layer: AMed populates the stimulus object slot in WM with the objects, <Reno>, <DEMAND "Cartha withdraw from Flatland">, and <Cartha>, which of course are at this point meaningless symbols. <Reno> and <Cartha> have matching frames in LTM, and DEMAND matches a slot in the DISPUTANT frame. These three frames are fed activation, which takes them above threshold, and all three are brought into WM. Once they are in WM (occupying three of the five available slots), the executive: (1) "fills in" the <Reno> and <Cartha> parts of the message stimulus object with the attributes listed in these frames; and (2) updates the mediation system stimulus object with a revised interpretation of Reno's motives and a decrement of the affective tag for RENO, as specified in the DISPUTANT frame (a higher level of hostility toward Cartha is now

---

case, will return AMed's own current interpretation of the mediation system, including the disputant's motives, resources, and any proposals currently on the table.

[58]I allow AMed's motives, stored in the SELF frame, to be either public or private. Only public motives would be directly expressed in a message.

attributed to Reno). In this demonstration run, I stop processing after only one round.[59]

Reasoning layer: Several implications are drawn from this message and the current representation of the mediation system. First, the features of the subjective mediation system constrain the range of responses (e.g., I have already mentioned that the more coercive mediation strategies are not available because AMed lacks the necessary resources). Second, several strategies in the COMM frame match the interpreted message (more particularly, the code word DEMAND appears as a condition in several rules in the COMM frame, so that the simple pattern matcher will compute a partial match to the frame).[60] COMM receives a jolt of activation, making it temporarily the most active memory object; as a result, COMM is transferred to a WM slot. There, the various rules that regulate actual strategies and responses in this category were available to the forward chaining inference engine in the executive.[61] Ultimately, AMed chose the strategies, "avoid taking sides" and "allow the interests of all parties to be discussed." Through more forward chaining, these strategies then triggered two (fairly derivative) messages as AMed's ultimate response to M1: (1) SAY "It would not be appropriate for me to take sides."; and (2) SAY "Why don't we discuss your interests."

M6: <Reno><HALT NEGOTIATION><Cartha>

Perceptual layer: In response to the five previous messages (listed in Table 1), AMed has by now arrived at a substantially revised interpretation of the mediation system: (1) Reno is now seen as highly hostile toward Cartha; (2) Cartha is seen as more hostile toward Reno than at the outset; (3) in part because there is no mechanism for changing anything in LTM that is not in a frame header, both Reno and Cartha are still seen as committed to the negotiation (that is, there is still a link between both actor frames and the NEGOTIATOR frame; indeed, that link has grown stronger because it has repeatedly carried activation each time the two actors were invoked, which happened with each message); (4) on the other hand, in the interpretation of the mediation system in WM, both

---

[59]Note that knowledge base for this demonstration run does not have the richness implied above, which would support multiple rounds of spreading activation and interpretation. Also, I have glossed over the fact that the entry of DISPUTANT (and even CARTHA and RENO) into WM was only highly likely, not certain. It is possible in the model for the more weakly activated frame NEGOTIATOR, which received activation spread from the CARTHA and RENO frames, to have been selected rather than DISPUTANT, which received activation from CARTHA, RENO, *and* the partial match from the message code, DEMAND.

[60]The current pattern matcher simply counts keyword matches, giving a higher score to frames with more matches.

[61]In essence, the forward chainer fires all rules in all frames in WM whose conditional (the IF part) matches information in WM.

actors are seen as more motivated by hostility than desire for a negotiated settlement, though this is far more extreme for Reno; and (5) the actors' resources are still seen as moderate and equal. M6 is quickly interpreted by drawing the relevant frames into WM and filling in the message stimulus object with information from the appropriate frames. The overall trend in the interpretation of the mediation system was reinforced by the message.

Reasoning layer: HALT NEGOTIATIONS, in conjunction with AMed's (mis)perception that Reno has nothing to gain by halting negotiations (i.e., Reno is not expected to be likely to gain its goals through any unilateral action because the actors are relatively equal in power), triggers the COMM frame again. From that frame (in particular, a set of rules designed to detect behavior by the disputants that fails to match the motives or resources attributed to them), AMed draws the conclusion that it needs to try to clarify the real interests of the actors. It assumes that Reno has misunderstood the realities of the situation and attempts to explain these realities (using the strategies, "clarify situation" and "supply missing information"). In addition, following another line of reasoning in COMM designed to restart stopped negotiations, AMed tries once again to "make contact with the parties" and "arrange for interactions between the parties". These two lines of reasoning culminate in: (1) SAY "<current interpretation of system>"; and (2) DO "Invite both parties to conference." Recall that SAY "<current interpretation of system"> means that AMed will explain to the disputants its own current interpretation of the mediation system, including the motives and resources it believes the actors all have.

This demonstration run was admittedly highly artificial and driven by a rather simplified version of the model. Nevertheless it raises my confidence that it would be useful to instantiate AMed with a more detailed version of the "mediation theories" articulated in the international mediation literature and place it within a more meaningful environment. The "behavior" of the model seems quite reasonable and, at minimum shows that the basic processes described in the theoretical and model sections of the paper will at least work (in the sense of generating behavior from input messages). On the down side, a great deal of work remains. In particular, I must improve the representation of basic processes within the model (learning and pattern matching, e.g.) and identify, code, and represent a richer body of mediation knowledge. Progress along these lines might eventually lead to actual experiments with an artificial mediator.

References

Anderson, John R. 1983. *The Architecture of Cognition*. Cambridge, MA: Harvard.

Bennett, Peter G. 1995. Modeling decisions in international relations: Game theory and beyond. *Mershon International Studies Review* 39: 19-52.

Bercovitch, Jacob. 1992. The structure and diversity of mediation in international relations. In *Mediation in International Relations,* ed. Jacob Bercovitch & Jeffrey Z. Rubin. New York: St. Martin's Press.

Bercovitch, Jacob, & Jeffrey Z. Rubin, eds., 1992. *Mediation in International Relations*. New York: St. Martin's Press.

Bond, A.H., & L. Gasser, ed. 1988. *Readings in Distributed Artificial Intelligence*. San Mateo, CA: Morgan Kaufman.

Boynton, G. Robert, & Milton Lodge. 1994. Voter's image of candidates. In *Presidential Campaigns and American Self-Images,* ed. Arthur Miller & B. Gronbeck. Boulder, CO: Westview.

Collins, Allan and Edward E. Smith, eds. 1988. *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann.

Conover, Pamela, and Stanley Feldman. 1984. How people organize the political world: A schematic model. *American Journal of Political Science* 28: 95-126.

Eysenck, Michael W. and Mark T. Keane. 1990. *Cognitive Psychology: A Student's Handbook*. London: Lawrence Erlbaum.

Feigenbaum, Edward A., and Julian Feldman, eds. 1995. *Computers and Thought*. MIT Press.

Fisher, Ronald J. 1997. *Interactive Conflict Resolution*. Syracuse, NY: Syracuse University Press.

French, J.R., & B.H. Raven. 1959. The bases of social power. In *Studies in Social Power,* ed. D. Cartwright. Ann Arbor, MI: University of Michigan Press.

Jones, Edward E. 1998. Major developments in five decades of social psychology. In *The Handbook of Social Psychology, Vol. 1*, 4th edition, ed. Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey. Boston: McGraw-Hill, pp. 3-57.

Kraus, Sarit. 1996. Beliefs, time, and incomplete information in multiple encounter negotiations among autonomous agents. *Baltzer Journals* (July 1, 1996).

Kraus, Sarit. 1997. Negotiation and cooperation in multi-agent environments. *Artificial Intelligence* 94(1-2): 79-98.

Kraus, Sarit, & Daniel Lehman. 1995. Designing and building a negotiating automated agent. *Computational Intelligence* 11(1): 132-71.

Kraus, Sarit, & Jonathan Wilkenfeld. 1993. A strategic negotiations model with applications to an international crisis. *IEEE Transactions on Systems, Man, and Cybernetics* 23(1): 313-23.

Kraus, Sarit, Jonathan Wilkenfeld, & Gilad Zlotkin. 1995. Multiagent negotiation under time constraints. *Artificial Intelligence* 75(2): 297-345.

Kreifelts, T., & F. Von Martial. 1990. A negotiation framework for autonomous agents. In *Proceedings of the Second European Workshop on Modeling Autonomous Agents in a Multi-Agent World*, pp. 169-82.

Kunda, Ziva. 1990. The case for motivated reasoning. *Psychological Bulletin* 108: 480-498.

Larson, Deborah Welch. 1994. The role of belief systems and schemas in foreign policy decision-making. *Political Psychology* 15: 17-33.

Levine, John M., & Richard L. Moreland. 1998. Small groups. In *The Handbook of Social Psychology, Vol. 2*, 4th edition, ed. Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey. Boston: McGraw-Hill, pp. 415-69.

Lodge, Milton, and Ruth Hamill. 1986. A partisan schema for political information processing. *American Political Science Review* 80: 505-540.

Lodge, Milton, Kathleen McGraw, & Patrick Stroh. 1993. An impression-driven model of candidate evaluation. *American Political Science Review* 87:399-419.

Lodge, Milton, Marco Steenbergen, & Shawn Brau. 1995. The responsive voter: Campaign information and the dynamics of candidate evaluation. *American Political Science Review* 89: 309-326.

Lodge, Milton, and Charles S. Taber. 1999. "Three Steps Toward a Theory of Motivated Political Reasoning." In Arthur Lupia, Mathew D. McCubbins, and Samuel L. Popkin, eds., *Elements of Reason: Understanding and Expanding the Limits of Political Rationality*. London: Cambridge University Press.

Miller, Arthur, Martin P. Wattenberg, and Oksana Malanchuk. 1986. Schematic assessments of presidential candidates. *American Political Science Review* 80: 521-540.

Minsky, Marvin, ed. 1968. *Semantic Information Processing*. Cambridge, MA: MIT.

Princen, Thomas. 1992. *Intermediaries in International Conflict*. Princeton, NJ: Princeton University Press.

Pruitt, Dean G. 1998. Social conflict. In *The Handbook of Social Psychology, Vol. 2*, 4th edition, ed. Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey. Boston: McGraw-Hill, pp. 470-503.

Raven, B.H. 1990. Political applications and the psychology of interpersonal influence and social power. *Political Psychology* 11: 493-520.

Riesbeck, Christopher K. and Roger C. Schank. 1989. *Inside Case-Based Reasoning*. Hillsdale, NJ: Lawrence Erlbaum.

Schank, Roger C. and Robert P. Abelson. 1977. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum.

Schank, Roger C. and Robert P. Abelson. 1995. *Knowledge and Memory: The Real Story*.

Spellman, Barbara A. and Keith J. Holyoak. 1992. If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles. *Journal of Personality and Social Psychology* 6: 913-933.

Sylvan, Donald A., & James F. Voss, eds., *Problem Representation in Political Decision Making*. London: Cambridge University Press.

Taber, Charles S. 1992. POLI: An expert system model of U.S. foreign policy belief systems, *American Political Science Review* 86(4): 888-904.

Taber, Charles S. 1999. The interpretation of foreign policy events: A cognitive process theory. In Donald A. Sylvan and James F. Voss, eds., *Problem Representation in Political Decision Making*. London: Cambridge University Press.

Taber, Charles, Milton Lodge, and Jill Glather. 1999. "The Motivated Construction of Political Judgements." In James H. Kuklinski, ed., *Thinking About Political Psychology*. London: Cambridge University Press.

Taber, Charles S., and Richard Timpone. 1994. The Policy Arguer: The architecture of an expert system. *Social Science Computer Review* 12(1): 1-25.

Taber, Charles S., and Richard J. Timpone. 1996a. Beyond simplicity: Computational modeling in International Relations. *Mershon International Studies Review* 40: 41-79.

Taber, Charles S., and Richard J. Timpone. 1996b. *Computational Modeling* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-113). Newbury Park, CA: Sage.

Touval, S., & I.W. Zartman, eds. 1985. *International Mediation in Theory and Practice*. Boulder, CO: Westview.

Zartman, I. William, ed. 1978. *The Negotiation Process: Theories and Applications*. Beverly Hills, CA: Sage.

Zartman, I. William, & Maureen R. Berman. 1982. *The Practical Negotiator*. New Haven, CT: Yale University Press.

Table 1: A Trace of the Demonstration Run

M1: <Reno> <DEMAND "Cartha withdraw from Flatland"><Cartha>

SAY "It would not be appropriate for me to take sides."
SAY "Why don't we discuss your interests."

M2: <Cartha><REJECT "Cartha withdraw from Flatland"><Reno>

SAY "It would not be appropriate for me to take sides."
-based on same reasoning used for M1
SAY "Why don't we step back for a minute and consider our options."
-based on reasoning found in the FORM frame that seeks to "reduce tensions"
-this came in because, as is possible, *both* the COMM and FORM frames were brought into WM during the reasoning phase
-when that happens, inferential chains can make use of rules from all frames in WM

M3: <Reno><ACCUSE "Cartha not negotiating in good faith"><Cartha>

DO "Call for proposals."
-from the FORM frame, the strategies of "control pace and formality of meetings" and "structure agenda"
-partly driven by the private goal of self-aggrandizement
-not really appropriate, this response highlights the preliminary nature of the model

M4: <Reno><THREATEN "military action"><Cartha>

SAY "We all have an interest in maintaining negotiations and avoiding conflict."
-from the FORM frame, the strategy of "highlight common interests"
-I worry at this point that the model has become trapped with the FORM frame, because the activation level of this frame continues to be increased the more it is used, making it more likely to be used in the future
-on the one hand, this represents a real cognitive tendency to become fixated; on the other hand, one would assume that an expert negotiator would not fall into this trap

M5: <Cartha><PROPOSE "Cartha cede some territory"><Reno>

SAY "Excellent proposal; let's talk about that!"
-from COMM frame, the strategies of "offer positive evaluations" and "encourage meaningful communication"

M6: <Reno><HALT NEGOTIATION><Cartha>

SAY "<current interpretation of system>"
DO "Invite both parties to conference."