Psychological Issues Raised by an AI Model of Reconstructive Memory

Janet L. Kolodner
Department of Computer Science
Georgia Institute of Technology
Atlanta, Georgia 30332

Lawrence W. Barsalou
Department of Psychology
Emory University
Atlanta, Georgia 30332

1. Introduction

This paper presents some psychological implications of an AI model of reconstructive memory. Psychologists have characterized human memory as reconstructive for years (e.g., [1], [6]). AI simulation of reconstuction goes further since computer implementation requires explicit specification of processes and representations. The particular AI model we consider here is Kolodner's [5] E-MOP based model, implemented in a computer program CYRUS. The model has three interrelated components: a retrieval process, an underlying memory organization, and processes for developing memory organization with the encoding of new events. The retrieval process was designed to imitate reconstructive retrieval strategies observed in people. The memory organization both supports and causes reconstructive retrieval. Processes for developing memory organization build new knowledge structures (i.e., learn) as new events are encoded. These new knowledge structures enable subsequent reconstructive retrieval of the new events.

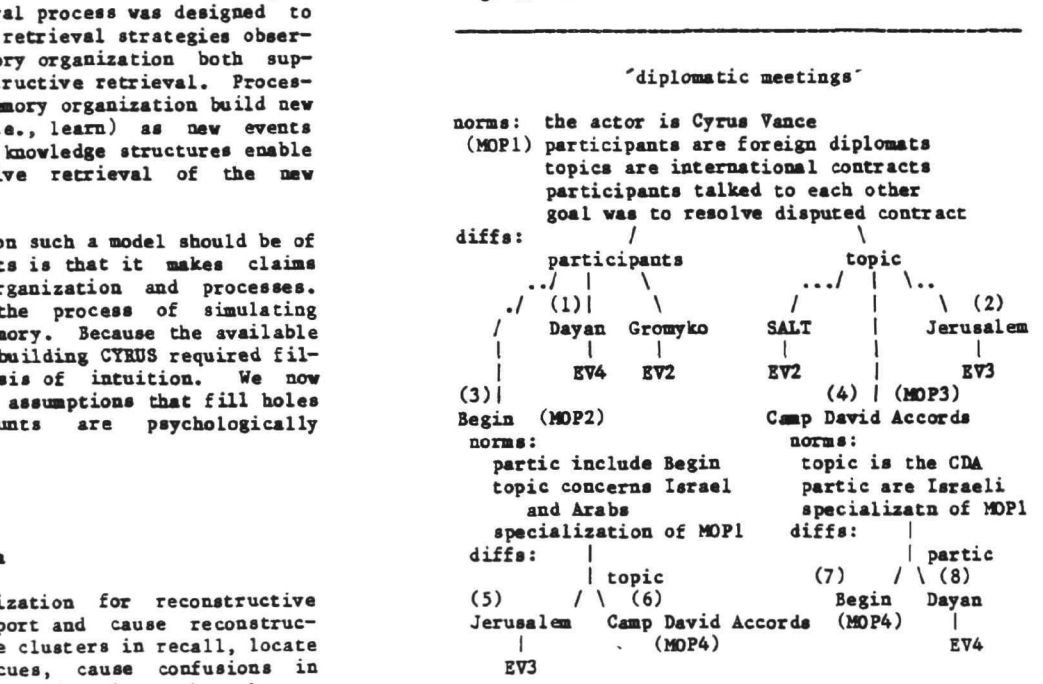
One important reason such a model should be of interest to psychologists is that it makes claims about human memory organization and processes. These claims stem from the process of simulating human reconstructive memory. Because the available model was incomplete, building CYRUS required filling it in in on the basis of intuition. We now ask whether the added assumptions that fill holes in psychological accounts are psychologically valid.

2. The E-MOP model

2.1 Memory Organization

A memory organization for reconstructive retrieval must both support and cause reconstruction. It must generate clusters in recall, locate and develop retrieval cues, cause confusions in recall and recognition, and emulate other characteristics of human remembering. Kolodner's memory organization uses conceptual categories called Episodic Memory Organization Packets, or E-MOPs (similar to Schank's MOPs [7]) that organize episodes in memory. A central assumption is that there is one E-MOP for each type of activity a person may be involved in, where type is defined as events that achieve a similar goal. Diplomats are involved in "diplomatic meetings", "diplomatic trips", "negotiations", and "state dinners". Each individual event is stored in the E-MOP(s) it fits into.

E-MOPs incorporate both episodic and generic memory (commonly, but incorrectly called semantic memory). The generic component consists of generalizations describing most of its members (i.e., some members may exhibit violations of these "norms"). Most "diplomatic meetings" discuss an international contract, for example, but a particular meeting might be called to plan an international event. An E-MOP's second component is its organization of member episodes. Episodes are organized based on how they differ from the E-MOP's norms. An episode is indexed and retrieved from an E-MOP by its relevant differences. When more than one episode has the same difference, a new sub-MOP is formed based on their similarities and differences. The figure below illustrates this organization:



The norms are features characteristic of diplomatic meetings. The episodes are indexed according to their similarities and differences in topic and participants. Meetings with the same topic or participants form E-MOPs whose norms are composed of their similarities. Below these norms, a set of similar instances are subsequently differentiated by their differences (i.e., mapped into indices). CYRUS organizes E-MOPs in three ways: (1) hierarchically, as just described; (2) by causal, temporal, and containment relationships between normative features; (3) by these same relations between indices.

2.2 Maintaining memory organization over time

Encoding new episodes requires both retaining the old organization to some extent and accomodat-

The work of the first author was partially supported by NSF under grant No. IST-8116892.

ing it to the new input. When a new episode is indexed identically to an old one, a new E-MOP must be formed to subsume them. That is, E-MOP formation is triggered by "reminding" [7], which occurs when the new episode retrieves the other similar one. The new E-MOP's norms are the similarities between the two items, and its indices are their differences. Because generalizations about a kind of event based on only two items may be inaccurate, subsequent episodes encoded with this E-MOP are used to refine these norms. If a feature not a norm for the first two episodes turns out to be normative for most others, what was initially an index can become a norm. Similarly, if a false generalization were made, norms for the first two instances can be relegated to indices.

2.3 Retrieval

Retrieval cues are abstracted from requests to remember an event. Such requests can be partial or complete specifications of the event to be retrieved. A request specifies an E-MOP to be searched and which indices within the E-MOP are to be traversed to find the event. An important assumption is that an E-MOP index cannot be traversed unless it is specified. In this way, retrieval is directed by the information in the request and further information that can be derived from it.

Since this process can fail in several ways, reconstructive strategies are proposed to deal with various types of failure. First, if the information in a request does not specify an E-MOP to be searched, then one or a small set of E-MOPs must be chosen. This process sees if any of the features stated in the request have E-MOPs associated with them (i.e., schema triggering).

A second type of failure stems from E-MOPs being untraversable unless their indices have been specified. A retrieval cue may specify features that don't correspond to E-MOP indices. Or, a retrieval cue may be so general that it doesn't specify enough features to direct traversal processes to a unique item. In that case, plausible features corresponding to E-MOP indices must be inferred from the given retrieval cues. A "meeting with Menachim Begin" might plausibly have taken place in Jerusalem. The strategies which make these inferences capitalize on an E-MOP's norms and knowledge about plausible relationships between different event features. Once such information has been specified, the corresponding indices are traversed. Interestingly, both types of strategies mentioned so far can lead to retrieval confusions and false starts.

A third type of strategy derives from the relationships between events in memory. Individual events refer to other events they are related to. If an event related to the requested event can be better specified, the related event can be used to further specify the requsted one. To recall a particular museum visit, for example, one might attempt to recall the trip it was part of.

3. Psychological Issues

This model stems from observations of how people remember, and what they forget. Although the processes and organization used to construct a complete model of reconstruction seem to work, are they really psychologically valid? One aspect of the model that has received empirical investigation to a large extent is reconstructive retrieval strategies [8]. People appear to elaborate upon requests to remember in many of the ways CYRUS does. Nevertheless, many issues remain untouched or at least require further attention. How does the organization of a set of events constrain the manner in which people elaborate on retrieval cues? That is, to what extent are such strategies content-dependent? Do people use the elaboration strategies used by CYRUS? Do they use others? What strategies are used most often, in what order, and for what reason? Similar to the content-dependence issue is the context-dependence issue. To what extent is elaboration affected by immediately previous searches for other events? for the same event? Given retrieval failure while searching an organized set of events, how do people select new parts of the organization for search? How does a retrieval access change organization? How sensitive is retrieval to incorrectly specified cues? Is the model too dependent on correctness?

Perhaps the most central issues the model raises are: How are events organized in memory? And how does this organization change over time? CYRUS assumes that events are the fundamental organizing units in memory. Is this true of human memory? If not, then what are the fundamental units? There may be several types of organizing principles. Others to be considered are: location (e.g., a local restaurant or bar); time (e.g., Christmas, summer); participants (e.g., Nixon, a spouse, a close friend)? If there are several ways events are organized, what determines which will apply to a given set of events? The content of the events? The goal the organization will serve? Perhaps several organizations simultaneously exist over a set of events.

A related issue concerns knowing what feature(s) should be used to discriminate two episodes sorted to the same E-MOP. There may be numerous features that distinguish two events, but only those that will be useful in the later evolution of generic knowledge should be chosen for indexing. How can such features be chosen? Another related issue is how many indices are grown each time reminding occurs. Another central issue concerns E-MOP construction. Is a new E-MOP constructed every time someone is reminded of an old event by a current one? To what extent is generic structure automatically acquired from and imposed on events? Or is conscious attention necessary to abstract normative information from previous events, organize it into E-MOPs, and apply it to new events? Human data may be informative on these points.

In the proposed memory organization, MOPs and their sub-MOPs form hierarchies in which common properties are stored once at the highest possible point in the hierarchy. This economy of storage parallels what psychologists call "cognitive economy" [2]. To date, it appears that the organization of semantic memory (i.e., lexical meaning) violates cognitive economy [3]. But to what extent is this violation true of other types of generic knowledge? Does human organization of events reflect cognitive economy? Or do people have much looser, less integrated and non-inclusive organizations for events?

An important aspect of E-MOPs is that they combine "episodic" and "generic" memories. This implies that episodic and generic memory are not separate entities but are intimately connected. If this is so, what exactly is the connection? When does episodic information (e.g., E-MOP indices) become generic (e.g., E-MOP norms or frame information)? When and how does generic information become confused with episodic information to

generate confusions? In E-MOPs, both happen as generalizations are refined and corrected.

There are a number or topics not covered in the original model which are nonetheless important to a theory of human memory organization and retrieval. One such issue concerns the roles of automatic versus conscious processes that encode information into memory. Temporal, spatial, and frequency information appear to be automatically acquired — even without knowing they are doing it, people encode these fundamental aspects of events [4]. In contrast, the acquisition of content information often seems to depend more on the use of conscious attention. When such information doesn't receive attention, the information is not acquired. How do these two types of processes interact to store events? Conscious attention may be responsible for the construction, organization, and reorganization of generic structure, since it usually contains content information. Automatic processes may be responsible for the strengthening of generic knowledge and the integration of spatial and temporal information into it. Finding algorithms for these latter phenomena and interfacing them with content-oriented processes appears to be an interesting problem.

A related issue is the role of similarity among events. This factor can facilitate people's memory performance on some occasions and interfere with it on others. Observing such phenomena in people's memory for events may further constrain the way in which we view generic knowledge of events and its use during retrieval. In E-MOPs, when a property doesn't correlate with other events, an index is set up differentiating the event with the deviant property from other events in the E-MOP. Correlation and differentiation play the role of keeping events suitably accessible. An event which conforms to the norms of an E-MOP will not be easily accessible because it won't have many indices differentiating it. On the other hand, events which have differentiating features will be accessible if those features are specified in or derived from a retrieval cue. What is the role of similarity and differentiation in people's memories? What is the actual effect they have on memory's organization?

Analogy is another area not covered in the original model. CYRUS does not address the migration of generic information from old E-MOPs to new ones. Generic knowledge associated with a particular E-MOP might be useful, however, in creating a new related E-MOP or in understanding something in a similar E-MOP. To what extent does "generic" structure generalize from one set of events to another? Must a completely new structure be built for each new set or does transfer occur? What procedures transfer the structure of an old E-MOP to a new one? How can knowledge in one E-MOP (e.g., for Vance) be used to understand something about a related referent (e.g., Haig)?

4. Future Directions

We are currently designing experiments that we hope will help answer the questions above. The experiments, no doubt, will raise additional questions. As a joint Artificial Intelligence and Psychology project, we will address these questions in the same way we have found it profitable to consider their ancestors --by building computer programs and by collecting human data.

5. References

[1] Bartlett, R. (1932). Remembering: A Study in Experimental and Social Psychology. Cambridge University Press, London.

[2] Collins, A.M., & Quillian, M.R. (1969). Retrieval time from semantic memory. JVLVB (8), 240-247.

[3] Conrad, C. (1972). Cognitive Economy in semantic memory. JEP (92), 149-154.

[4] Hasher, L. & Zacks, R.T. (1979). Automatic and effortful processes in memory. JEP:G (108), 356-388.

[5] Kolodner, J. L. (1980). Retrieval and organizational strategies in conceptual memory: a computer model. Research Report #187. Dept. of Comp. Sci., Yale, New Haven, CT.

[6] Norman, D.A. & Bobrow, D.G. (1979). Descriptions: An intermediate stage in memory retrieval. Cognitive Psychology (11), 107-123.

[7] Schank, R. C. (1980). Language and memory. Cognitive Science (4), 243-284.

[8] Williams, M. W. and Hollan, J. D. (1981). The Process of Retrieval From Very Long-Term Memory. Cognitive Science (5), 87-119.