A HYBRID MODEL FOR CONTROLLING RETRIEVAL OF EPISODES

Colleen M. Seifert

UCSD and NPRDC

A central problem for models of memory is the retrieval of episodes. Most models assume retrieval is an automatic process where an input is matched to the contents of memory, and an episode is activated based on similarity. If there is a high degree of similarity between an input and a particular case, and both cases share little with other cases in memory, then similarity alone may be enough to account for the spontaneous retrieval of the case. However, when there are many instances that overlap in similarity, or when the similarities are abstract in nature, it appears that predicting retrieval based on similarity measures alone is quite difficult. Schank (1982) argues that the process of reminding is *mediated* by an abstract knowledge structure used to understand the original event -- cases are activated as a consequence of activating organizing schemas that capture important similarities. Is similarity alone enough to cause the automatic activation of prior episodes? In the next sections, I will present evidence from empirical studies about the conditions under which such remindings occur; specifically, that reminding based on similarity is not an automatic process but involves a strategic or goal-based function. The same information, in the presence of different processing goals such as explanation or planning, will result in retrieval of different prior cases from memory. In order to account for this, I propose a hybrid model of retrieval that incorporates both the content-addressable character of distributed memory models with the controlling influence of goals in processing. In the model, goals act as a controlling mechanism to focus attention on features that will result in retrieval of relevant episodes.

Experimental Evidence for Activation of Episodes

The question of interest is whether memory organization (similarity) alone will predict the automatic activation of cases. Leaving aside the question of conscious awareness of the result of the retrieval, we can investigate whether the memory organization formed at encoding provides connections between cases such that activating one case in memory will result in the activation of other cases encoded with the same organizing structures. This activation may be the result of an automatic retrieval process, that is uncontrolled and non-optional, as in semantic priming, or it may occur only under some conditions, under strategic control.

The basic paradigm involved presenting pairs of stories that either share or do not share the organizing knowledge structure, and then testing whether the similarity-based memory organization that results causes different activation patterns. In the studies, time to answer a question about a previously read story is used to measure the accessibility of the story in memory. If the two stories are connected by the organizing structure in memory, then answering a question from one story should activate the other story in memory, resulting in faster responses to a question from that story.

Experiments with abstract thematic structures

The materials chosen for the experiments (Seifert, McKoon, Abelson, and Ratcliff, 1985), included similarities based on the types of structures evident in protocols of natural remindings; namely, the goal and plan interaction that occur as goals are pursued. For example, the adage "closing the barn door before the

horse is gone" can be characterized as a planning failure where X knows a plan to prevent goal failure, but delays execution to avoid the cost until the goal is failing; then, the plan is executed, but fails because it is not a recovery plan but a prevention plan (Dyer, 1983). If two stories based on the same theme are connected in memory, responding to an item from one should speed the time to respond to an item from the other, compared to where the preceding item refers to a story that does not share the same theme, and therefore should have no connection in memory between the two stories. The results showed that the shared knowledge structure did not affect the ease of access of the episodes in memory. However, in a second experiment we instructed the subject to think about the theme as they read, and to rate the similarity of the story pair after the test list. This time, there was a significant effect of thematic similarity on response times, in that responses were faster for test items when the story pair shared the same thematic organizing structure. The only difference between the two experiments was the instruction to attend to similarity. We conclude, then, that accessing the same schema does not automatically connect two episodes in memory; instead, some strategic purpose is necessary to promote activation of prior episodes. The strategic basis for the ability to utilize connections between episodes has been replicated in other experiments (Seifert, McKoon, Abelson, and Ratcliff, 1985).

Experiments with content-based structures

One question was whether the strategic result was specific to abstract structures, or would hold true of more content based structures. To test this, we examined the effect of memory organization packets (MOPs) (Schank, 1982) on connections between cases (for a complete description, see McKoon, Ratcliff, and Seifert, 1988). Pairs of stories were written to instantiate the same MOP (such as "going to the beach") but without overlapping lexical items in the description of MOP actions. In the experimental procedure, subjects read a long list of stories, and then, after reading all the stories, connections from one story to another were measured by priming in old/new recognition judgments of phrases from the stories. Since the test items included MOP information, any activation effects may be due to connections in semantic memory, and not reflect any activation from one case to another case. In a second experiment, the test items used as primes contained only information that refers to the story-specific representation. Results showed priming from one story to another story of the same MOP when test phrases were related to the MOP; however, when the test phrases were story-specific but not related to the MOP, there was no evidence of priming. One way to describe the results of these experiments is to say that subjects cannot discriminate whether two MOP-related phrases were from the same or different stories; for example, "spreading out her towel in a dry place" from one beach story and "found an empty space for her blanket" from another beach story are equally good primes for "slowly strolled into the cool ocean". The results as a whole demonstrate that case to case activation is not an automatic phenomenon resulting from connections in memory.

A paradigm for intentional reminding

Since a strategic process appears to be involved in the activation of cases, a next step was to look at reminding within a strategy-based task. Up to this point, in order to keep the subject unaware of the intent of the experiment, we examined activation rather than conscious reminding. Now that the activation appears to be dependent upon a strategic process, we can examine reminding in a paradigm more similar to natural remindings. In these experiments (Seifert, McKoon, Abelson, and Ratcliff, 1985), subjects were asked to study a set of stories, then to read a new set of stories followed by test items from the study set. The test stories had either the same theme (as described in the earlier experiments) or a

different theme than the story that the target item was from. The results showed a significant facilitation effect for same theme pairs. A test story appeared to activate a previous story based on its thematic similarity, resulting in response facilitation. Subjects reported that as they read the test stories, studied stories occasionally "came to mind"; if they did, they were a good predictor of the target item, and so the reminding was useful information to the task. It seems the strategic aspect of this task was that the purpose of the reminding was built into it -- usefulness in predicting the test item provided a functional purpose for the reminding. Also, subjects were conscious of the remindings and their utility, and this may have encouraged attention to themes and the resulting remindings. This methodology was the first to investigate reminding in the laboratory.

Incorporating Goals in Reminding

From the experiments presented above, I conclude that encoding a story does not always activate a thematically similar story *unless* there is also present a functional or strategic purpose for the reminding. These results are reminiscent of analogical transfer studies (Gick and Holyoak, 1983), where transfer of a story solution to a new problem was infrequent unless either instructions or multiple example stories were provided. The similarities in the experiments reported here are more obvious than those in the transfer literature, and yet subjects were not reminded unless encouraged by the strategic nature of the task. Intention appears to play a bigger role in retrieval than may have been assumed, because without it the retrieval does not occur. This strategic aspect of retrieval points out the importance of cognitive processing goals; apparently, such goals may act as a constraint on retrieval in that activation occurs that would not under other circumstances. In this sense, cognitive goals appear to form a context within which retrieval operates, and the nature of this context determines what remindings occur. The results suggest that, at the least, *intentional reminding* (Schank, 1982) plays a much bigger role than previously indicated. I define intentional very broadly, as not always consciously intended, but as a bias in processing.

A result that supports this conclusion is that the low transfer rates from a story scenario to a new problem (Gick and Holyoak, 1983) were increased when the context was made congruent in the two cases (K. Holyoak and L. Novik, personal communication, February 1988). That is, presenting the first case *as a problem* rather than as a "story understanding task" improved the spontaneous application of the prior solution to the new problem. Since processing goals appear to focus attention within the retrieval process, the types of goals and their connections to features must be examined. Situations that appear to foster remindings in humans are functions like explanation, planning, argumentation, decision making, and conversation. The kinds of processing goals that may be involved can be described at a general level, but certainly we can discover more about what subtasks of cognition may be involved (Chandrasekharan, 1987). For example, the understander seeks, in his processing of new input in conversation, to be reminded of a memory that relates to what he heard and provides evidence for the point of view he wishes to defend. Remindings can serve to verify your analysis of episode, illustrate why your reasoning is valid, justify or support a claim, give specific solution information, and perhaps provide an analogy.

A Distributed Model of Retrieval

I have argued from experimental results that strategic purpose is important because automatic processes like retrieval are mediated by processing goals. How might goals be incorporated into the reminding process? Retrieval can be modelled as a feature space with encoded episodes, as in a completion network in a parallel distributed processing model of memory (McClelland and Rumelhart, 1985). Memory access is determined by the similarity between input features and stored

patterns, where the process finds the activation pattern that bests fits the connection constraints. For example, given an input such as "a Professor forgot his computer password", a similarity match may find a particular episode pattern, such as "Prof. Jones forgot his password last week", to be a close enough match compared to other patterns in memory. If the result of the matching process is a set of features corresponding to a schema, the retrieved pattern might be "Absent minded Professors forget things." Another possibility is that nothing will match well enough for a stable pattern to be retrieved. These results can be characterized in terms of maximizing goodness of fit: a good fit may be an episode reminding, an adequate fit may a be a generalization, and a low fit may mean no close matches could be retrieved.

Rumelhart (1988) has proposed that the matching process can operate by relaxing the constraints on the match, resulting in analogy in cases where no distinct pattern could be retrieved by strict similarity to the input pattern. For example, if too many episodes share the feature set, or if some of the features are inconsistent with the patterns already encoded into the memory, then no episode pattern could be isolated by similarity alone. However, by softening the constraints on the match, some of the input features could be utilized as "don't care" constraints, allowing analogical matches that focus on some features and throw out features that may be preventing a match from occurring. The notion is to release features by allowing the network to "turn them off" progressively until a better match is found. Rumelhart's proposal is to feed additional input strength to the features differentially, and allow the system to find the overall best fit by overriding features with low values that conflict. A feature is thrown out only if doing so leads to better fit than the contribution from that feature adds to the match. This *relaxed match* process provides a mechanism for overriding features that are preventing a match (such as context information) while giving increased importance to features that may be helpful in analogical transfer to other episodes.

Methods of feature selection

The softening of constraints could be accomplished in several ways. The particular implementation Rumelhart proposes is to deliver a constant amount of activation to each of the features; the lower the constant, the more likely that feature's contribution to fit will be less useful than throwing it out, making it easier for the system to override it. The selection of *which* features to assign what values to is central to the solution of the relaxed match. One possibility is to select the values assigned to each feature at random, resulting in a blind focus to the match process. This is a feasible method, but we would expect then to generate a lot of episode remindings with nonsensical similarities and ignore important ones. For example, from the Professor example, we might retrive a reminding like "Joe owns a computer." The random selection problem is more impressive with more input features in the match; clearly, it would be easy to recall episodes that share features but without a coherent similarity -- a correlational match but not a causal one. It may be that human memory operates this way, but it most often appears that remindings result from coherent similarities.

Another method of assigning values to the input features is to order all the features along an abstraction hierarchy (Rumelhart, 1988). By weakening constraints on concrete features first, and then progressing to abstract features until reaching a match, greater weight is given to abstract features and more abstract analogies will result from the relaxed match process. This approach certainly seems plausible given the results from example remindings; however, there are several problems with an intrinsic ordering on feature importance. First, "abstract" is hard to define, as even thematic patterns like "closing the barn door after the horse has gone" can have pretty contentful features (i.e., "too late").

Though generality can be defined in terms of how many instances are captured by a category, abstractness is much more difficult to define implementationally. For example, is "intelligent" more abstract than "independent"? A second problem is that the relative importance of features may change dynamically or as a function of interaction of features. In choosing a car, color, size and speed may be given a particular ordering of importance; in selecting a mode of transportation, these same features may be given a different ordering. Finally, as pointed out by data on remindings, it is not always the most abstract features that are responsible for a reminding; often, content features play a role in the similarity that is not accounted for by a strict abstractness metric. Of course, abstractness may play a role in the relaxed match process despite these problems; certainly, there seems to be an implicit notion that people have about what types of features may be important in analogy.

Controlling retrieval with goals

However, an alternative proposal is to use the context of the current goals being pursued as a method of weighting features for a match. Remindings, I claim, have more to do with *what* the person is attending to than to any a priori notions about abstractness. It is these processing goals that provide more weight to particular features and allow others to disappear from a match. I propose that goals play a role in retrieval by affecting what features are attended to most in a similarity match process. Particular features of the case will be attended to based upon the cognitive goals operating at the time.

Here's an example of how goals may act to select features to attend to:

> **Input: Teletrack has a "no smoking" policy which means, in effect, that every section is a smoking section. The rule is disregarded often enough, by people seated in various sections, that no section can be guaranteed to be smoke-free. You argue that the management should allow smoking in some sections, if only to be sure that some sections can actually be non-smoking.**

Suppose that you are engaged in an argument about your comment. Your goal is to buttress your claim that, basically, control is better than abolishment when disobedience is common. Supporting your claim involves retrieving another instance with this pattern; based on the abstract characterization of the claim, an analogous situation is recalled. You may be reminded, on this basis, of the legalization of heroin in England. That case supports the claim by identifying a circumstance where control of heroin use had fewer bad effects on non-participants than the outright banning of the substance.

Suppose instead that your goal is to plan a way to get your assertion implemented. In order to get your solution effected, you have to figure out a way to get the authority involved, Teletrack, to recognize that their control problem is worse now than it would be under the smoking section plan. Planning for implementing your solution selects problem features that are relevant to the management of the policy. You may be reminded, on this basis, of the pet policy in student housing. Pets were banned, but then many students got pets and broke the rules more and more brazenly, causing quite a bit of disturbance. In an attempt to manage the effects of the pets, the management made several buildings "no pets" and let the pet owners congregate in the others. This suggest a plan for the smoking policy: at Teletrack, smoke as much as possible, and encourage others to, in a way that exacerbates the problem for the management. They will be more likely then to see the benefit of control rather than prohibition.

Even if the plan does not seem a like good one, the point here is that the processing that goes on prior to a reminding may be much more involved than simply encoding input facts into a single representation. Instead, the features that

play a larger role in retrieval may be ones that are tied to particular processing goals active at the moment. From this, we could expect that retrieval of the new episodes later would depend on a congruency of cognitive purpose.For models where only one processing goal is ever present, the goals may be implicitly encoded in the process model and memory representation, compromising the ability to make claims about how cases are utilized in general.

The "goodness" of an analogy depends on what you are interested in; therefore the relevance of features changes dynamically as a result of the goals of a system. The goals may be modelled as connections to the input features that "gate" the activation of the features, in terms of multiplicative influences on the strength of the input features (as in Hinton's multiplicative links). The goal influence can turn down irrelevant and turn up relevant features in the match, resulting in the features relevant to the goal being the focus of the relaxed match. The reason for the separation of the goal links to the features from the input line architecture is to allow the same goals to function autonomously to the retrieval process, since they will also affect other subprocesses besides retrieval. Tasks like creating an explanation or learning can also be affected by the current processing goals. While the implementation of an abstractness metric through input line activation may be possible, that information has to be determined over many different processing contexts -- the abstractness of features may be learned over many different tasks. The gating connections from goals will provide activation based on information about what is relevant in the *current* processing context, which changes depending on the goals of the system. Thus, the two types of information about feature importance can be incorporated into the same network model.

Extensions to Learning

Cognitive goals may affect other processes besides retrieval, operating to provide a context for saliency of features. The tendency to treat all features as equal, without regard to how particular features may be attended to in certain cognitive contexts, is a major problem in models of learning. For example, distributed models of learning as a class consider all features present as potentially equally important to the rule being learned. Consequently, it takes many trials to determine which features are actually predictive rather than occurring occasionally. This approach hurts learning rates in two ways; first, it may hurt mathematically in the learning algorithms to include so many features that take many trials to be ruled out; and second, it ignores information available in utilizing a priori notions of what features are likely to be related to the rule.

In a learning system, treating all features equally will mean a lot of effort spent on features that have no importance to the rule learned. With the addition of a mechanism to select the features most likely to be relevant to the rule, one can save time by focusing on features that will pay off more quickly. For example, the current state of the network can provide a focus on particular input features based on what features have been given higher weights in previous learning. Consequently, by utilizing prior knowledge, one can shut off features that aren't a priori relevant to learning. Of course, as I have been arguing in this paper, this is not enough when the same network serves other tasks. Which features are relevant in a domain can't be set in an a priori way that will be true across all learning contexts. One needs a mechanism to affect the attention at the nodes *depending upon* the processing context. The goals can affect which features are attended to by interacting in a multiplicative fashion with activation on the input nodes. This method can incorporate changes in relevance due to the particular goals being pursued.

A particular method to control learning might be to focus on input values which have small outgoing weights, in a sense betting that the information to be

learned will have to do with the features already shown to be relevant in previous learning. Of course, when the same network is used for other tasks, then a gating method will be required as proposed for the retrieval model, to allow the current system goals to directly impact which features are attended to. This problem of selecting features to attend to in learning has begun to receive some interest recently. M. Mozer (personal communication, March 1988) has begun studying the problem using a bootstrapping learning procedure that attenuates input values with small weights. The results are suggestive: with noisy inputs, the network decides to ignore the noise inputs *before* solving the mapping, and in the case of redundant inputs, the attentional mechanism selects one and shuts out the redundant feature.

It seems that people often have notions of what features *may* be important to learning in a particular domain. For example, in learning a new video game, the actions effective in killing enemies tend to be predictable from common-sense notions of physical causality (e.g., you have to be in a linear position with the target for a shot to hit it). To the extent that the actual game mechanics violate these notions, they may be increasingly hard for people to learn. In fact, the less face validity to the rule, the more often people may fail to perceive its presence among the possible factors. This is a drawback to the process of using default expectations for the relevancy of features; however, it may be more than compensated for by the ability to quickly detect the operation of more obvious factors. By attending to the seemingly related features, the learning process may be "smartened up" and sped up, at the cost of discovering nonintuitive or novel connections in the data. Within human systems at least, the gain from attention in speed of learning may more than make up for missing counterintuitive or unusual patterns. Utilizing cognitive goals to affect the attention paid to features based upon their inferable connections seems a promising approach for models of learning as well as retrieval. I think the combination of network strengths (content addressing) and a controlling mechanism (attention based on goals or previous learning) is a promising method to "smarten up" the behavior of networks in retrieval and speed up the learning process.

References

Chandrasekaran, B. (1987). Towards a functional architecture for intelligence based on generic information processing tasks. *Proceedings of the Tenth IJCAI*, Milan, Italy.

Dyer, M. G. (1983). *In-depth understanding: A computer model of integrated processing for narrative comprehension.* Cambridge, MA: MIT Press.

Gick, M., & Holyoak, K. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1-38.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114*, 159-188.

McKoon, G., Ratcliff, R., & Seifert, C. M. (1988). Making the connection: Generalized knowledge structures in story understanding. Unpublished Manuscript.

Rumelhart, D. E. (1988). Towards a microstructural account of human reasoning. Unpublished manuscript.

Schank, R. C. (1982). Dynamic memory: A theory of reminding and learning in computers and people. New York:Cambridge University Press.

Seifert, C. M., McKoon, G., Abelson, R. P., & Ratcliff, R. (1985). Memory connections between thematically similar episodes. *Journal of Experimental Psychology: Human Learning and Memory, 12* (2), 220-231.