Automatic and strategic aspects of knowledge retrieval

William H. Walker & Walter Kintsch

University of Colorado

Send proofs to:
W. Kintsch
Department of Psychology
University of Colorado
Boulder, Colorado 80309

This report is based on the dissertation of W.H. Walker, directed by W. Kintsch. We thank the other members of the committee, especially A. Ericsson for his help with the protocol analyses, as well as R.M. Shiffrin for his thoughtful comments. H.H. Walker is now with IBM, Rochester, Minnesota. The preparation of the report was aided by Grant MH 15872 from NIMH to W. Kintsch.

Revised 7-3-1984

Technical Report No. 129

Institute of Cognitive Science
University of Colorado
Boulder, CO 80309

ABSTRACT

An experiment was performed in which subjects retrieved members of natural categories. Subjects used a variety of retrieval cues in this task, making knowledge retrieval a two-tiered process: First, contexts were generated in which category members are likely to be found, and second, these were used as retrieval cues to produce the category members themselves. These retrieval cues were primarily of an episodic nature, rather than abstract-semantic. Similarly, in a script generation task subjects tended to generate scripts from personal experiences, rather than through the retrieval of an abstract schema or script. A loosely interassociated memory network is suggested by these results. A model of memory retrieval developed to account for the data of list learning experiments by Raaijmakers & Shiffrin was shown to be able to account for the automatic component of knowledge retrieval, but requires a more complex control structure before it can successfully simulate the retrieval strategies used by the subjects in these experiments.

Knowledge retrieval is an important component of many cognitive tasks. Models that describe adequately how people retrieve real-world knowledge are not yet available, however. One difficulty has been accounting for the complex control processes that appear to be involved in retrieval, and another is separating problems of knowledge use from those of retrieval itself.

We propose to disentangle this complex by taking advantage of what we know about retrieval processes from the study of list-learning paradigms. We assume that what this literature says about the automatic component of retrieval is essentially correct. That is, given a particular memory structure and a specific retrieval cue, current memory theory adequately describes the use of this cue to retrieve information from memory. This is only a partial theory of retrieval, but its advantage is that by taking it to its limits we can begin to see more clearly the outlines of what we don't understand. Our present goals, then, are modest ones: we would like to see how much a purely automatic retrieval process can do, and what empirical phenomena are beyond its scope. By acquiring this basic information we can start thinking about formulating a theoretical account of those aspects of knowledge retrieval that are now neglected.

Specifically, the strategy we propose is the following. We take a simple task that involves knowledge retrieval, and then try to describe the resulting data by means of a retrieval model that we know accounts adequately for the retrieval of a list of words memorized in the laboratory. How well are the data described by such a model, and what aspects of the data are missed by our model? At this point we only aim at a qualitative description of the major features of the data rather than a strict, quantitative fit.

The task we begin with is naming members of natural categories (e.g., Name all cars that you can think of). This free-emission, category retrieval task is an old one in psychology, and its main features are reasonably well understood. Traditionally, the interest has mainly been in the time course of retrieval, and hence this will be the major phenomenon our model is to simulate. Retrieval of members of a natural category is rapid initially, but gradually slows down (Bousfield & Sedgwick, 1944), though it may continue at a low rate indefinitely (Williams & Santos-Williams, 1980). Group curves are smooth and well described by negative exponential functions (Gruenewald & Lockhead, 1980), while individual curves are heavily scalloped, reflecting the retrieval of closely interassociated groups of items or clusters (Graesser & Mandler, 1978). Caccamise (1981) has extended this paradigm beyond category member retrieval by instructing subjects to retrieve everything they know about certain topics (e.g., nuclear energy). Her subjects retrieved many more ideas (propositions in the retrieval protocols) for familiar topics than for unfamiliar topics and produced more tightly organized protocols when the topics were familiar, while the average number of propositions retrieved per chunk was unaffected by topic familiarity or other experimental variations. Thus, there exists a stable data base in the free-emission paradigm to make a meaningful simulation possible.

The model chosen for the simulations is that of Raaijmakers and Shiffrin (1981). This model has several properties that make it attractive for our purposes: it accounts well for a large number of phenomena in the list-learning literature, and it represents somewhat of a consensus model in that it embodies features that are widely perceived as important, such as encoding specificity, redintegration, and associative networks.

1. AN EXTENSION OF THE RAAIJMAKERS AND SHIFFRIN MODEL

Only a very brief outline can be presented here of the crucial features of the Raaijmakers and Shiffrin (1980, 1981) model.

Memory is represented as a network of interassociated concept nodes, which are potentially complex structures containing information of various types which we need not specify here. The retrieval process is guided by a memory probe (retrieval cue) which resides in short-term memory and operates on this associative structure. Retrieval is a dynamic process, because the contents of the retrieval probe are continuously modified, depending on what happened during retrieval, and the memory network itself changes similarly during a retrieval session.

The memory probe contains a limited number of cues, including the situational context, task constraints, and subject generated information thought to be relevant to the retrieval task. Some of the cues in the probe are nonchanging, but others change during retrieval. Thus, as new items are sampled from memory, they may replace some of the old cues in the probe.

The retrieval structure consists of the strengths of association between all potential cues and the nodes in memory, and is thought of as a matrix in which rows represent cues and columns concept nodes. The cells of the matrix contain the strength of the direct associations between the cue-node pairs.

Retrieval consists of a sampling component and a recovery component. During sampling a node associated to the current retrieval cues is selected for retrieval; if sufficient information about this node is available, the node can be recovered and retrieval is completed. The sampling rule is an extension of Luce's (1959) ratio rule. The probability that a node will be sampled is equal to the product of the strength of the connections of that node with all the cues in the current probe, divided by the sum across all nodes of such products. Thus, no single cue has a preponderant weight in deciding the sampling of a node. To have a good chance at being sampled, a node must be associated at least moderately with all or most cues in the probe. A target strongly associated with most cues in a probe but unassociated with one of the cues can not be retrieved without a change in the composition of the probe: a suitable control process must be formulated to drop the odd cue.

The probability of recovery is determined by an exponential function that rapidly approaches 1.0 as the weighted sum of the absolute strengths of association between the sampled node and the cues in the probe becomes large. Therefore, sampling in this model is always successful in the sense that something will always be retrieved, and retrieval failures occur because of a failure to recover some sampled item. Whenever an item is recovered, the associative strength of the connections from the cues to the recovered item are increased in such a way that the successful retrieval path is reinforced. In addition, if there is room in the probe the recovered item is added to it; if not, one of the non-permanent cues in the probe is selected at random and replaced by the recovered item. Thus, retrieval is modelled as a dynamic process.

For present purposes, the model has four parameters. First, there is the probe size, r, which we set to r=4 for all conditions. Of these four cues, two were designated as permanent and two as changeable: whenever a new item was successfully retrieved, it replaced one of the changeable items in the probe set. Then there are two stopping criteria, KMAX, and LMAX. At the start of the retrieval process a variable K is initialized to K=0. Whenever a retrieval failure occurs or whenever an old, already retrieved item is retrieved again, K is incremented by 1. When K reaches KMAX, the retrieval process is terminated. A variable L is also initialized to zero, and is incremented by 1 whenever K is incremented, but reinitialized to zero whenever it reaches the value LMAX, at which point the current memory probe is purged of its nonpermanent cues and retrieval is started again. In accordance with Raaijmakers and Shiffrin, KMAX was set to 30 (except in those cases where retrieval continued for a fixed time period, as will be shown below), and LMAX was set to 3. The final parameter of the model is the amount the associative strengths between the cues in the memory set and the corresponding nodes in the network are incremented when retrieval is successful. This value was set to 0.2 for the first retrieval of an item, and $.2^j$ for the j-th retrieval of the same item on the basis of some exploratory work and the previous results of Raaijmakers and Shiffrin.

This model was implemented in FORTRAN on the VAX 11/780 of the Computer Laboratory for Instruction in Psychological Research. All simulations were run with the predetermined parameter values described above, because our intent was not to tune the model to optimally fit data, but rather to determine how well it would do in this new situation without any ad hoc improvements.

2. RETRIEVAL FROM NATURAL CATEGORIES

Experiment 1

The purpose of our first experiment was twofold: first, to provide some data that could be compared with the simulations, and, second, to collect concurrent verbal protocols of the task. The verbal protocols were used to help generate hypotheses about subjects' retrieval strategies. Thus, subjects named members of three different categories, one well known and presumably large and well structured (automobiles), one intermediate in size and structure (soups), and one small and ill-structured (laundry detergents). Half of the subjects produced concurrent verbal protocols, while the other half, which did not produce protocols, served as a control for possibly interfering effects of the protocol procedure. In addition, subjects performed a chunking and rating task to check on the structures produced in the free-emission task.

Method

Subjects. Twelve introductory psychology students (eight male and four female) satisfied a course requirement by participating in this study.

Procedure. Each subject participated in two private sessions. Subjects were asked to name as many members as they could of categories of objects encountered in everyday life. They were given 12 minutes for each category. If pauses of more than 40 seconds occurred, they were encouraged to keep trying. All responses were tape recorded.

As a warm-up task, all subjects named as many cameras as they could in three minutes. Then the three experimental categories (automobiles, soups and laundry detergents) were presented, in a different order for each subject.

Half of the subjects were instructed to "think out loud" as they named category members. It was emphasized that they were to express whatever thoughts crossed their minds and not to worry about producing proper English sentences. The retrieval strategies used by these subjects were identified from the protocols they produced in this way (see below). These subjects were also given two brief warm-up tasks to familiarize them with the thinking-out-loud procedure. For the remaining subjects the experimenter tried to identify the strategies used during retrieval retrospectively. Following retrieval he asked subjects to describe how they had remembered category members and to confirm or disconfirm his hypotheses about their retrieval strategies.

In the second session, subjects were given slips of paper with the category members they had produced in the previous session and were asked to sort them into natural categories as they saw fit. These clusters were then named, and the cluster names sorted again. Furthermore, for each subject a matrix was prepared with the category members the subject had produced in the rows and the retrieval strategies which had been identified in the first session in the columns. Subjects were asked to identify and rank up to three retrieval strategies related to each category member. Subjects were told not to guess and to respond only when they were reasonably confident of their choice.

All tapes were transcribed with marks at the end of every five seconds so that cumulative retrieval curves could be generated.

Results

Figure 1 shows the cumulative retrieval of Automobile names as a function of time for the verbal protocol group and the control group, while Figure 2 shows the same data for a single subject. The typical smooth, negatively accelerated function is obtained here for the group curve, in contrast to the scalloped individual function. The retrieval curves for Soups and Laundry Detergents are similar, except that fewer category members are retrieved.

Insert Figures 1 & 2, Table 1

While the control subjects in Figure 1 produced somewhat more items than the verbal protocol subjects, the difference between these two groups was minor and did not reach statistical significance. As Table 1 shows, this was also the case for the other two experimental categories. An analysis of variance of the data shown in Table 1 with the factors Group (between subjects) and Category (within subjects) confirms this impression, with the main effect for Group yielding an $F(1,10)=1.45$, $p>.05$, and $F(2,20)<1$ for the interaction. There was, however, a highly significant main effect for Category, $F(2,20)=54.29$, $p<.001$: the larger and more familiar the category, the more members were named.

There was, of course, a large difference between the verbal protocol and control subjects in the number of old (i.e. previously reported) items produced, $F(1,10)=7.82$, $p<.025$: control group subjects tended to edit out already produced but reretrieved items, while reretrievals were reported by verbal protocol subjects.

Subjects' responses on the verbal protocols were assigned to five different scoring categories. First, all protocols were broken into phrases, which were defined as basic idea units, usually consisting of a single clause or single words. For example, consider the following fragment from Subject 1's automobile protocol where the phrase segmentation has been indicated by slashes:

> "Just keep on picture where cars are (R), / parking lot (R), / ...a sales lot (R), / ...on the road (R), / Subaru Brat (C), / Celica (C), / ...picturing cars in front of my dorm (R), / ...ones I come out and see all the time (R), / ...Midget (C), / ...I know like no names (M), / ...Camaro (C), / my sister got her first Camaro (D)."

Four of the five scoring categories are indicated above by the paranthesized capital letters. Category members (C) were names of category members (e.g., Monte Carlo) or subcategories (e.g., Chevrolet). Names of models and manufacturers, reference to particular cars and other marginal category members were also accepted. Descriptors (D) were descriptive statements which sometimes preceded but most often followed a category member and in some way elaborated on it (e.g., Dodge trucks; I wrecked one of those once.) Retrieval-related statements (R) were statements directly related to retrieval, which were neither category members nor descriptors (the R-statements from one of the protocols are shown in Table 2, to illustrate the variety of statements in this category). The remaining three categories were Metastatements (M - e.g., What time is it? or I am getting pretty fed up with this.), General Task Related Statements (G - e.g., My girlfriend has a Subaru. Her dad works at Rocky Flats and loves it. He's pretty wierd.), and Unrelated Statements (U - e.g., I was thinking of breakfast.). Less than .1% of all responses fell into the latter two categories. 20% of all statements were metastatements, which were equally divided among the three categorization tasks and occurred at a steady rate throughout the experimental sessions.

Insert Table 2 about here

The number of non-category member retrieval cues (R) subjects mentioned in the concurrent verbal protocol and in the retrospective interview differed significantly. While an average of 29.4 cues were identified for verbal protocol subjects, an average of only 5.2 cues were identified retrospectively for control subjects, $F(1,10)=36.72$, $p<.001$. Presumably this does not mean that the control subjects used fewer such cues, but only that they could remember fewer retrieval cues on their retrospective reports. It was also the case that more retrieval cues were reported for the large categories than for the small ones. The average number of retrieval cues was 23.2 for Automobiles, 15.4 for Soups, and 13.3 for Detergents, $F(2,20)=13.38$, $p<.001$.

In the second session subjects performed two tasks. First, they sorted previously produced category members into clusters of related items. Second, to the extent they could, they assigned their category members to the retrieval strategies that had been identified for them. In general, subjects performed this second assignment well: only 7.3% of all items were not assigned to a retrieval strategy.

Thus, the data could be clustered in two ways: once on the basis of the sorting task and once on the basis of clustering during retrieval. These two structures did not corresond well. First of all, in the verbal protocol condition the number of unique retrieval statements (R) was more than double the number of clusters sorted (12.3 retrieval strategies versus 5.8 sorted clusters, $F(1,10)=13.38$, $p<.005$). Even more striking, the clusters derived from the sorting task overlapped only 19% with the clusters defined by the retrieval strategies. Apparently, the principles of organization in the clustering task were quite different from those operative in the retrieval task.

Not only the number of clusters and their content were different in the sorting and retrieval conditions, even more important are the differences in the nature of the clusters themselves. This becomes clear if one classifies all cues and clusters as semantic or episodic. Semantic clusters included such things as <u>small cars</u>, <u>foreign cars</u>, <u>General Motors cars</u>, <u>luxury cars</u>, and so on, while episodic clusters and cues included things like <u>"my" cars</u>, <u>friends' cars</u>, <u>cars I have wrecked</u>, <u>cars I have seen on TV</u>, etc. Fewer than one percent of the clusters and cues were not sortable into either the semantic or episodic category. Only 13% of the sorted clusters were episodic, while 77% of the retrieval strategies were episodic. When sorting the items they named for the categories used, subjects tended to rely on context-free, semantic dimensions. But when they were asked to retrieve members of those categories, their tendency was to rely on episodic retrieval cues.

Discussion

People appear to have available a large and varied set of retrieval strategies that they use when naming category members. This suggests that the organization of nodes in memory must be characterized by a corresponding variety of associative connections. That is, nodes in memory must be connected associatively to a large number of other retrieval routes.

For example, the car name Camaro may be associated with a variety of semantic and episodic nodes, perhaps with Chevrolet, sports car, gas-guzzlers, fast car, Jim Rockford of television fame, the woman at work who just bought one, Donna C. who owns one, General Motors, the colors bronze, black and green, and so on. Which connections subserve retrieval depends on the retrieval cue currently active.

In the free-emission category retrieval task only the category name is given as a retrieval cue. Other retrieval cues must, to a very large extent, be retrieved themselves. After naming their first three or four detergents (or soups, or cars) most subjects began to deliberately search for situations in which detergents (or soups, or cars) are encountered (e.g., in laundromats, grocery stores, the laundry room in the dorm, etc.). The situations recovered were then used as cues to retrieve additional detergent names. If one plots the retrieval of retrieval cues as it is done for Automobiles in Figure 3, functions are obtained which appear to be more linear than the retrieval functions for the category members themselves (Figure 1). In applications of the SAM model to predict switches from one category to the next in categorized free recall, Shiffrin (personal communication) has assumed that subjects cycle through a series of retrieval cues using a constant failure criterion to govern switching. This predicts relatively linear generation and use of retrieval cues while at the same time predicting overall recall to rise in the usual negatively accelerated fashion.

Insert Figure 3 about here

----

2. SIMULATION 1

Rather than model the details of specific semantic fields (like cars), we focused on building the formal properties of natural categories into our simulations. In every case we worked with a retrieval cue by memory node matrix of dimension 100 x 100. The "content" or "meaning" of the cues and memory nodes remained unspecified. The associative strengths between cues and memory nodes (the cells of the matrix) were assigned at random under varying constraints. Thus, large categories were simulated by matrices in which every cue (row) tended to be associated with many nodes (columns). For small categories retrieval strengths were assigned so that cues were associated with fewer nodes on the average.

To simulate retrieval from large, intermediate and small categories, we built three types of matrices. We constructed an intermediate structure by filling a matrix with values sampled randomly from a uniform distribution on the interval 0 to 1, and we created a highly interassociated matrix by taking the fourth root of all entries in an intermediate structure and multiplying by 0.78. Low levels of interassociation were simulated by raising the entries in an intermediate matrix to the third power. While these procedures are quite arbitrary, they generated structures of the desired kind.

Thus our application of the Raaijmakers & Shiffrin model differs in two important respects from the practice of the original authors: we do not restrict the search set to the items in the appropriate category, but rather want to see how well the model performs without these restrictions, and we work only with a single interassociation matrix, while Raaijmakers & Shiffrin distinguish associations among the items of a list from item-context associations. With these modifications, the Raaijmakers & Shiffrin model with the parameter set discussed above was used to retrieve items from large, intermediate and small categories. To simulate naming automobiles, the retrieval process was started on the high strength matrix with two randomly selected items placed permanently in the probe set, leaving two items in the set to vary. The simulation was run for a fixed amount of time with the stopping criterion KMAX set high enough to prevent self-termination. The results of such a run are shown in Figure 4.

Insert Figures 4-7 about here

Group retrieval curves were generated by averaging six individual runs and are shown in Figures 5-7. These simulations duplicate several major features of the experimental data (shown in the insets in Figures 5-7): retrieval curves are smooth and negatively accelerated; the level of associative strengths, which serves here as our analogue of category size, increases the steepness as well as the asymptote of the retrieval curves: there are relatively more re-retrievals of previously retrieved items ("old items") in the low associative strength condition than in the high strength condition.

On the other hand, there are aspects of these simulations which do not coincide with the experimental data. First of all, many more old items are retrieved in the simulations than by control subjects. This probably reflects the tendency of subjects in the free-emission task to edit their responses, which was also observed in the comparison of the control data with the verbal report data, and which has been reported by other investigators as well (e.g., Caccamise, 1981). It would be quite trivial to add an editing process to the Raaijmakers and Shiffrin model to simulate this processing component.

A more interesting deviation from the experimental data is observed, however, when the retrieval functions for individual subjects are compared with the single simulation runs: the former are clearly more scalloped than the latter, as can be seen by comparing Figure 2 with Figure 4. This is a difference of fundamental importance. In the Raaijmakers and Shiffrin model the retrieval probe changes dynamically as retrieval progresses, but only retrieved category members themselves function as supplemental cues. As our experimental data have shown, this is not the case with real subjects: an important component of retrieval is the retrieval of retrieval cues themselves, so that retrieval becomes a two- or multi-step process. First, a plausible (strategic) retrieval cue is selected, then that cue is used to produce category members, causing the pronounced scallop in the retrieval function. At present the model is able to simulate the cued retrieval of items or even of strategic cues, but it cannot model the second component, the coordination of the retrieval of cues followed by the retrieval of category members.

3. RETRIEVAL FROM MULTICATEGORY STRUCTURES

Simulation 2

While Simulation 1 showed that the Raaijmakers and Shiffrin model provides a reasonable first approximation for certain aspects of knowledge retrieval processes, one could object that the simulation was too unrealistic. While it is not unreasonable to simulate natural categories like Automobiles with a 100 x 100 matrix, in human memory that matrix would be embedded in a much larger structure, and an observation of primary importance is that when people are trying to name automobiles, they do not stray much beyond that semantic field. Some additional simulations demonstrated that the model behaves similarly in three analogous situations. Of course, the various structures we have investigated here do not exhaust the range of possibilities, but the way these structures are operationalized here reflects important features of the organization of memory and knowledge.

Single clusters. Either 20, 40, or 80 nodes in the basic 100 node matrix were chosen to form a single cluster. This was done by raising the strength values of their interconnections by adding to it a randomly determined amount from a uniform distribution on the interval 0 to 1, and then normalizing the resulting strength values to the 0-1 interval. Thus items within a cluster were more strongly interassociated on the average than other items in the matrix.

Simulations were performed with the parameter values already mentioned. Their results were in good qualitative agreement with the predictions: smooth, negatively accelerated group curves were generated, with both their slopes and asymptotes increasing with cluster size, as expected. Most importantly, hardly any items were retrieved from outside the clusters when the memory probe consisted of items from within the cluster (0.4, 0.0, and 0.0% for clusters of size 20, 40, and 80, respectively).

Overlapping clusters. Real memory structures do not consist of a single cluster embedded into some larger matrix, but of many, typically overlapping clusters. Such structures were approximated by forming two overlapping clusters in our basic 100 x 100 item matrix. The first cluster consisted of items 1 - 30, while the second cluster consisted of items 11 - 40. Within each cluster, associative strength values were incremented, so that the items in the cluster overlap were actually incremented twice.

Simulating retrieval functions from this structure with standard parameter values yielded encouraging results. If the probe set consisted of items belonging only to one or the other cluster, retrieval was restricted to items from that cluster: in all cases, less than 1% of the retrieved items came from outside the original cluster. If the probe set consisted of items that belonged to both clusters, retrieval occurred primarily from the overlap area, but a substantial number of items were also retrieved that belonged only to a single cluster (20%). Almost no items not belonging to either cluster were retrieved (0.3%).

Hierarchical structures. Hierarchical structures were simulated by forming one cluster of items 1-22, and then forming four subclusters of five items each among the first 20 items. As with other simulations, items within the cluster were more associated with one another than items outside the cluster. The same pattern of relative strengths held for items within subclusters. Items 21 and 22 were arbitrarily designated as the permanent cues in the retrieval probe. These items were moderately associated with the other 20 items in the cluster. Once again, retrieval stayed almost completely within the overall cluster, and then proceeded from subcluster to subcluster, with the observed likelihood of staying within a subcluster being more than three times as high as would be expected by chance. Thus, the model behaves much like people in a free-recall task with a categorized word list.

We conclude from these exercises that the Raaijmakers & Shiffrin model can account for some of the basic phenomena of knowledge retrieval in naming tasks rather well, if it is given reasonable memory structure to operate upon. The much larger size of these structures than the list structures for which it has been developed appear to be no obstacle, and the model generates retrieval functions that are quite adequate in light of known experimental results. This is the case even when the memory structure is internally organized in fairly complex ways, imitating some features of human memory organization. At the same time, we have noted in the discussion of Simulation 1 some serious limitations of the model: it can represent not more than one component of the knowledge retrieval system, albeit a basic one.

4. RETRIEVAL FROM SCRIPT-LIKE STRUCTURES

Retrieval from loosely structured natural categories exemplifies only one type of knowledge retrieval, and perhaps not a prototypical one. We have therefore also investigated retrieval of information from stereotyped temporally and causally ordered situations or scripts. The relevance of scripts to psychological processing is well established today (e.g., Bower, Black, & Turner, 1979; Grasser, 1981), though the exact nature of these structures and the way they are used in various cognitive processes is still poorly understood. An experiment is reported here that further explores this matter as well as some simulations with the unmodified Raaijmakers & Shiffrin model.

Experiment 2

Method

Subjects. Three male and three female students from the same pool served as subjects in this study in individual sessions.

Procedure. After appropriate warm-up tasks, subjects participated in three script-naming tasks. They were asked to tell what typically happens in certain common situations, pretending to inform someone from another culture who has no idea of what is involved. The situations were _Going to a restaurant for a meal_, _Going to a grocery store to buy groceries_, and _Going to a doctor's office for a check-up_. The order in which the subjects performed these tasks was varied. Afterwards, each subject participated in a retrospective interview, designed to elicit information about the strategies used in the experimental tasks.

All responses were tape recorded and transcribed with five-second time marks. From these protocols, cumulative retrieval functions were constructed as in Experiment 1, with clauses (text propositions, in the sense of van Dijk & Kintsch, 1983) as units.

Results

The data produced in this experiment were in excellent agreement with the results reported by Bower, Black, & Turner (1979). Any event mentioned by at least 50% of the subjects in either Bower et al. or the present study was generated by more than 25% of the subjects in the other study.

The retrieval functions obtained in the three experimental tasks are shown in Figure 8. It is obvious that these functions are quite different from those obtained for category member retrieval: instead of being negatively accelerated, these functions increase at an essentially constant rate. The linear nature of these functions is not an artifact of the method of averaging different subjects together that are stopping at different times since the retrieval functions for individual subjects show the same pattern of increase, though they are more scalloped than the group functions.

Insert Figure 8 and Table 3 about here

Subjects' retrospective reports produced some rather interesting results. Table 3 provides a classification of the retrieval strategies mentioned by subjects in their retrospective reports. As we know from Experiment 1, retrospective reports represent probably an underestimate of the total number of strategies actually employed in the course of retrieval, but we have no reason to suppose that the proportions of different strategy types are thereby biased. There are two noteable results in Table 3. First, for most retrieval episodes more than a single strategy was reported. Thus, a multiplicity of experiences appears to be associated with retrieval of scripted events. In fact, on only 11% of all retrieval episodes subjects reported using a single retrieval strategy. The second major result apparent in Table 3 is that only 10% of the retrieval strategies subjects reported were normative, abstract, context-free retrieval cues - i.e., the classical notion of schema or script - while the great majority of strategies that were reported were clearly of an episodic nature. By far the most frequent strategy was thinking of a single place in which the script could occur and generating the script sequence in that way. Often, this involved visualizing themselves going through the appropriate actions.

Visits to a grocery store, with which our subjects had a great deal more experience than with doctors' offices or restaurants, were described in a somewhat different way: all subjects reported using some concrete place as a cue for generating an appropriate description, rather than a particular visit to a store, frequently in connection with an imagery strategy.

Discussion

The crucial difference between the category naming task in Experiment 1 and the script description task here is that scripts are tightly organized temporal-causal sequences. Subjects found the task of describing scripts a very easy one; they started at the beginning and rapidly generated an appropriate sequence (the average time was less than 2 minutes for the script descriptions, while subjects were given 12 minutes for the naming task).

Subjects were not retrieving a set of sequenced facts which, together constituted a description of a visit to a grocery store, a doctor's office, or a restaurant. The processes that led to the descriptions were far more complex than that. Subjects did rely on the retrieval of information, but the information retrieved consisted largely of memories of places in which the script could be enacted and particular instantiations of the scripts, with relatively little context-free information.

Yet subjects were not describing the locales about which they said they were thinking, nor were they detailing those script instantiations which they said they had in mind. These disparities between the subjects' descriptions and what they later said they were doing foregrounds the fact that in addition to retrieval, other processes were involved in the production of script descriptions. Apparently a large amount of editing is involved in this task. Concurrent verbal reports might throw some light on this question.

It is quite likely that a generalization process was the bridge between subjects' recollections of instantiations of a script and their descriptions. Subjects apparently thought of a particular visit to a restaurant or doctor and derived from that particular visit a description of a typical visit. A particular experience thus became decontextualized. To the extent that script descriptions are based on such decontextualization, one cannot say that scripts are held in memory separate from context. Schank (1982) and van Dijk & Kintsch (1983) have made similar points.

Simulation 3

Although it is obvious from the discussion above that the basic Raaijmakers & Shiffrin model could not possibly account fully for the processes of information retrieval in script descriptions, it is nevertheless worthwhile to explore how far the model actually does reach.

For this purpose, a chain of nine items was constructed by incrementing the associations between items n and n+1 and associating them with a superordinate script node. When run, the simulation would sample one of the nine items in the sequence and then retrieve two or three of the following items in the sequence in order. Thus, a run might retrieve the items in this order: 5, 6, 7,/ 2, 3,/ 1,/ 8, 9,/ 4 (the slashes mark the points at which all nonpermanent cues were purged from the probe).

Obviously, the model could not handle the script description task without modifications. First of all, the program did not know where to find the first item in the sequence, and the chain-like memory structure was not sufficient to guide the retrieval process. Clearly, what the model needs is a more extensive control structure, which is simply not part of its original formulation. Instead of having two permanent and two dynamically but randomly changing cues in the probe set, a control process is needed that updates the probe in such a way that the (or a) correct script sequence is selected.

5. GENERAL DISCUSSION

There were two main goals of this study. First of all, we wanted to explore how well a model taken over wholesale and almost without modifications from the list-learning literature could account for knowledge retrieval in some simple experimental situations. Secondly, through experimental analysis, we wanted to get a clearer idea what the characteristic features of retrieval processes in these situations actually are. We comment briefly on both of these in turn.

The answer to our first question is an unambiguous "yes-and-no". Yes, because we found no reasons in our analyses to doubt that the basic, automatic component of knowledge retrieval is identical with that identified in the memory research employing various list-learning paradigms. Given a retrieval structure characterized as an associative network and a particular retrieval cue, certain automatic, dynamic processes occur that lead to the retrieval or non-retrieval of items in that structure. The Raaijmakers & Shiffrin model appears to provide an adequate characterization of these processes, both in the list learning and in the knowledge retrieval case. This does not mean that the model is perfect, or that alternative characterizations are not possible. But here we have at least one sturdy and useful tool for the further exploration of the processes involved in knowledge retrieval.

The second part of our answer about the status of the Raaijmakers & Shiffrin model must be negative, however, for it clearly does not provide a full account of the processes involved in either the category naming task or the script description task. This is, of course, no news, because it would have been quite unreasonable to ever suppose something else. However, we are somewhat ahead in this game, because even though we always knew that the model could not possibly be complete, we now have some hints about the nature of its incompleteness. Which brings us to our second point.

Knowledge retrieval, even in such simple tasks as naming category members and describing scripts, is not or at least not primarily determined by fixed, pre-existing organizational patterns in the memory structure. It is much more of a constructive process, where appropriate patterns are being created as a particular task demands out of a loosely structured, complexly interrelated memory system. Van Dijk & Kintsch (1983) have argued this point, and the present results provide some experimental support for their claims.

When naming category members, very few items are directly retrieved. We have no lists of laundry detergents or automobiles in memory, or only very fragmentary ones. Instead, the subject retrieves memory nodes associated with the retrieval topic, and then uses them for the further retrieval process. The category is thus generated, but is not in any sense a memory unit that can simply be retrieved. Furthermore, these intermediary memory nodes that are involved in retrieval are typically not abstract, context-free and general, but can be very personal and ideosyncratic. While the category or script that is thus generated is semantic, general, decontextualized, the processes involved in its generation rely heavily on episodic memory.

Recently, Strube (1984) has reported some interesting research which strongly supports the findings reported in the present study. Employing a classical free-association paradigm, he observed 1) that only the very first items are retrieved directly in response to the stimulus word, 2) that further items were mediated through various "situational contexts" which correspond directly to the retrieval statements identified in the present study (Table 2), and 3) that these retrieval cues had a concrete, experiential, episodic character. (In this regard Strube cites a study by Ziehen (1898) where this very same point was emphasized). Finally, Strube provides a simulation of his data, using once again the Raaijmakers & Shiffrin model as a basis. The way he applies this model is, however, rather different from what was done here. Strube attempts to simulate a fragment of an ideal subject's long-term memory, namely the domain of animal names, and then studies retrieval processes in that domain. This allows a more detailed comparison between his subjects' protocols and the simulation than was possible here, but of course does not permit him to investigate different types of memory structures as was done here. It is noteworthy, however, that the memory structure he arrives at is a loosely inter-associated network, much like the one we have used.

A rather different approach to the problem of memory organization and knowledge retrieval was used in another series of recent studies by Kolodner (1983a, 1983b, 1984). However, her conclusions agree very well with the present results. Kolodner provides a detailed simulation of a particular person's long-term memory, and her concerns are much broader than what we have tried to do in the present paper. Her retrieval model, at least at the level of an information flow chart, appears to be very similar to the one proposed here, emphasizing reconstructive, context-dependent processes and the importance of retrieval cues which are not themselves to be retrieved items. Indeed, we would claim that the present results provided direct empirical support for some of the postulates of her theory (e.g., Memory Principle 6 in Kolodner, 1983b).

If we turn from category member generation tasks to script generation tasks, we find similarly that scripts are not found in memory and reproduced, but are put together from relevant information retrieved from memory. The problem for the theorist then becomes to describe exactly how scripts, or categories, are generated, and how their precise constitution is fine-tuned to task demands, as it apparently is. While this is by no means a trivial task, our results permit us to form some hypotheses about how to proceed. The naming task, for instance, might be conceptualized as a two-step procedure, in which the original memory probe is used to retrieve some of its highest associates. This will yield a few category members, but also some promising material that can be used for further retrieval attempts. It is not clear how complex the control processes for such a two-step procedure need to be, but it is conceivable that a really accurate simulation of the naming task could be constructed in this way. The control processes for the script description task are undoubtedly more complex. But at least we know now what we want to do, and what additional data need to be collected to make further progress possible. The trick is to generate certain structures out of a memory in which these structures are not built in beforehand by the theorist. We certainly need to know more about the control processes involved - but it appears that we don't have to worry about the basic, automatic retrieval mechanism itself: memory theory provides us with a suitable model.

If our conjectures about knowledge retrieval in the human organism are right, workers in artificial intelligence might ask themselves some questions about the design principles of knowledge retrieval systems. Typically, such systems are semantically oriented; will they work with the required flexibility, or might it be worthwhile imitating human processes? If so, how does one create rich, episodic artificial knowledge bases that would support human-like, adaptable retrieval processes?

REFERENCES

Bousfield, W.A. & Sedgwick, C.H.W. The analysis of sequences of restricted associative responses. Journal of General Psychology, 1944, 30, 149-165.

Bower, G.H., Black, J.B., & Turner, T.J. Scripts in memory for text. Cognitive Psychology, 1979, 11, 177-220.

Caccamise, D.J. Cognitive processes in writing: Idea generation and integration. Doctoral dissertation, University of Colorado, 1981.

Graesser, A.C. Prose comprehension beyond the word. New York: Springer, 1981.

Graesser, A.C. & Mandler, G. Limited processing capacity constrains the storage of unrelated sets of words and retrieval from natural categories. Journal of Experimental Psychology: Human Learning and Memory, 1978, 4, 86-100.

Gruenewald, P.J. & Lockhead, G.R. The free recall of category examples. Journal of Experimental Psychology: Human Learning and Memory, 1980, 6, 225-240.

Kolodner, J.L. Maintaining organization in a dynamic long-term memory. Cognitive Science, 1983, 7, 243-280 (a).

Kolodner, J.L. Reconstructive memory: A computer model. Cognitive Science, 1983, 7, 281-328 (b).

Kolodner, J.L. Retrieval and organizational strategies in conceptual memory: A computer model. Hillsdale, N.J.: Erlbaum, 1984.

Raaijmakers, J.G.W. & Shiffrin, R.M. A theory of probabilistic search of associative memory. In G.W. Bower (Ed.) The psychology of learning and motivation, Vol. 14. New York: Academic Press, 1980.

Raaijmakers, J.G.W. & Shiffrin, R.M. Search of associative memory. Psychological Review, 1981, 88, 93-134.

Schank, R. Dynamic memory. Cambridge: Cambridge University ress, 1982.

Strube, G. Assoziation: Der Prozess der Erinnerns und die Struktur des Gedaechtnisses. Berlin: Springer Verlag, 1984.

van Dijk, T.A. & Kintsch, W. Strategies of discourse comprehension. New York:

Academic Press, 1983.

Williams, M.D. & Santos-Williams, S. Method for exploring retrieval processes using verbal protocols. In R.S. Nickerson (Ed.) Attention and performance VIII. Hillsdale, N.J.: Erlbaum, 1980.

Ziehen, Th. Die Ideenassoziation des Kindes. Berlin: Reuther & Reichard, 1898.

Table 1

Mean Number of Category Members

named by Group and Category

	CATEGORY		
GROUP	AUTOS	SOUPS	DETERGENTS
CONTROL	61.5	29.5	14.0
VERBAL PROTOCOL	54.5	23.7	10.5

Table 2

Retrieval related statements from the protocol of S2, Soups. Verbatim quotes are indicated by quotation marks. The number of category members retrieved are shown after each statement (repetitions are in brackets).

"The first soup is the one I hate most" - 1

good soups - 6

Campbell soups - 8(3)

soups "I eat most often" - 2(3)

"Restaurants I have eaten in" - 0

"I think of Chinese restaurants" - 3

"In Italian restaurants it is usually..." - (1)

"I have eaten soup in Greek restaurants" - 0

"I have never eaten in Mexican, a real Mexican restaurant" - 0

"Campbell soups comes out 75 soups" - 1(1)

"I am trying to think of different vegetables that might be in soups" - 3(3)

"I just thought of that because I thought of a French restaurant, La Crepe, and they have great..." - 1

Table 3

Strategies of various types subjects reported using on the three script generation tasks. (Note that subjects were allowed to report more than one strategy for each retrieval task).

Reported Strategy	Number of Script Retrievals for which strategy was reported.	Percentage of Script Retrievals for which strategy was reported.
Remembered normative, abstract, context-free fact	4	22%
Thought of a single place and its lay-out	15	83%
Pictured or visualized self in a place	10	56%
Remembered a particular visit	6	33%
Thought of a number of different places	3	17%
Remembered a number of different visits	3	17%

Figure Captions

Figure 1. Mean cumulative retrieval of automobile names as a function of time for the Verbal Protocol and Control groups. Retrieval of both new and old items are shown.

Figure 2. Cumulative retrieval of automobile names as a function of time (Verbal Protocol Subject 1). Retrieval of both new and old items are shown.

Figure 3. Mean cumulative production of new non-category member retrieval cues for automobiles as a function of time (Verbal Protocol subjects only).

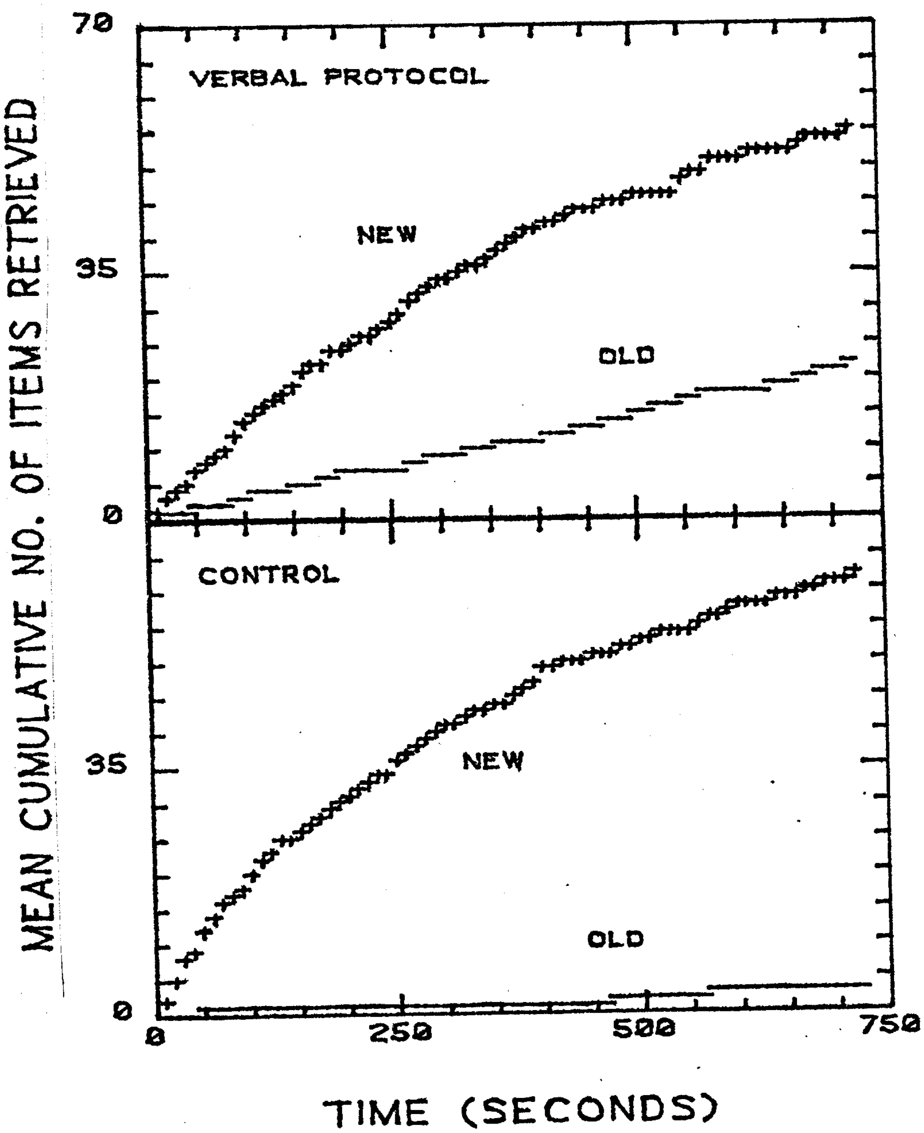
Figure 4. Cumulative retrieval of items as a function of time (arbitrary units) for one simulation of retrieval from a large natural category (standard parameter settings: KMAX = 100).

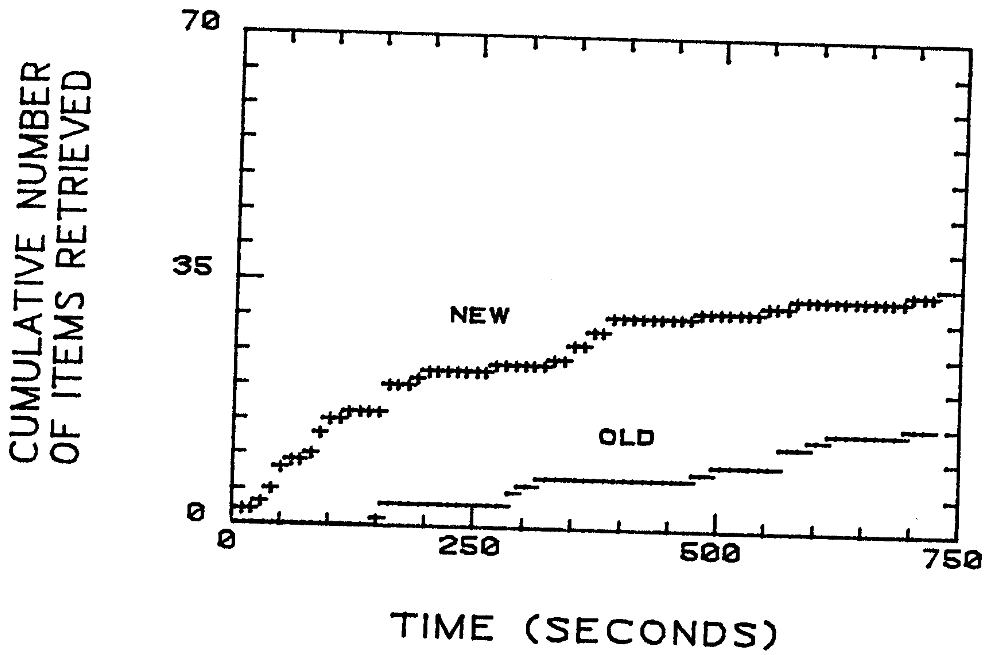
Figure 5. Mean cumulative retrieval of items as a function of time (arbitrary units) for six simulations of retrieval from a large natural category (standard parameter settings: KMAX = 100). Inset is top panel of Figure 1.

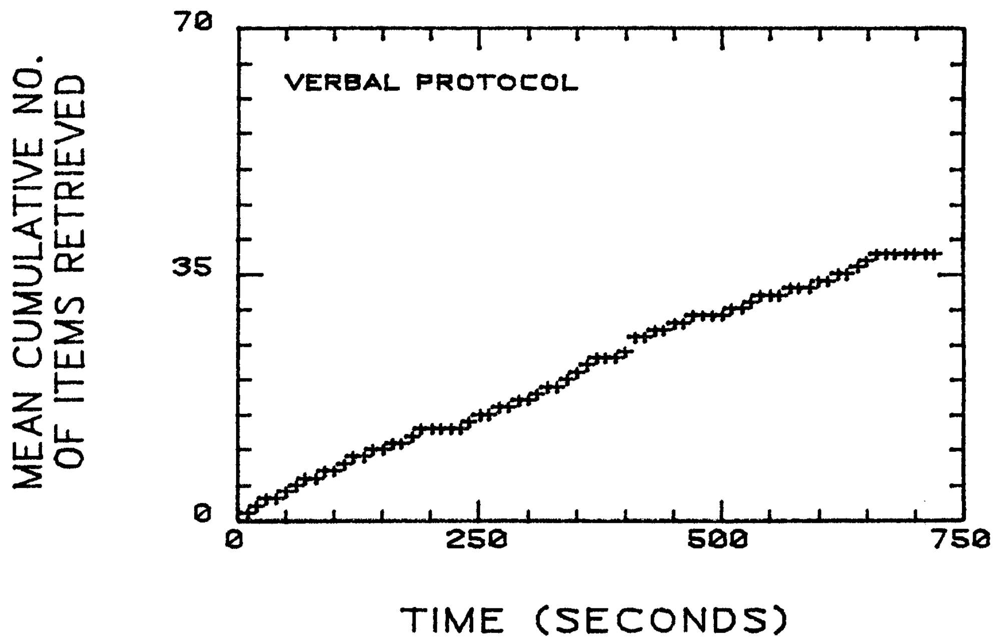
Figure 6. Mean cumulative retrieval of items as a function of time (arbitrary units) for six simulations of retrieval from a medium-sized natural category (standard parameter settings: KMAX =100). Inset shows Verbal Protocol subjects' cumulative retrieval of Soups (scale values match those of Figure 1).

Figure 7. Mean cumulative retrieval of items as a function of time (arbitrary units) for six simulations of retrieval from a small natural category (standard parameter settings: KMAX = 100). Inset shows Verbal Protocol subjects' cumulative retrieval of Laundry Detergents (scale values match those of Figure 1).

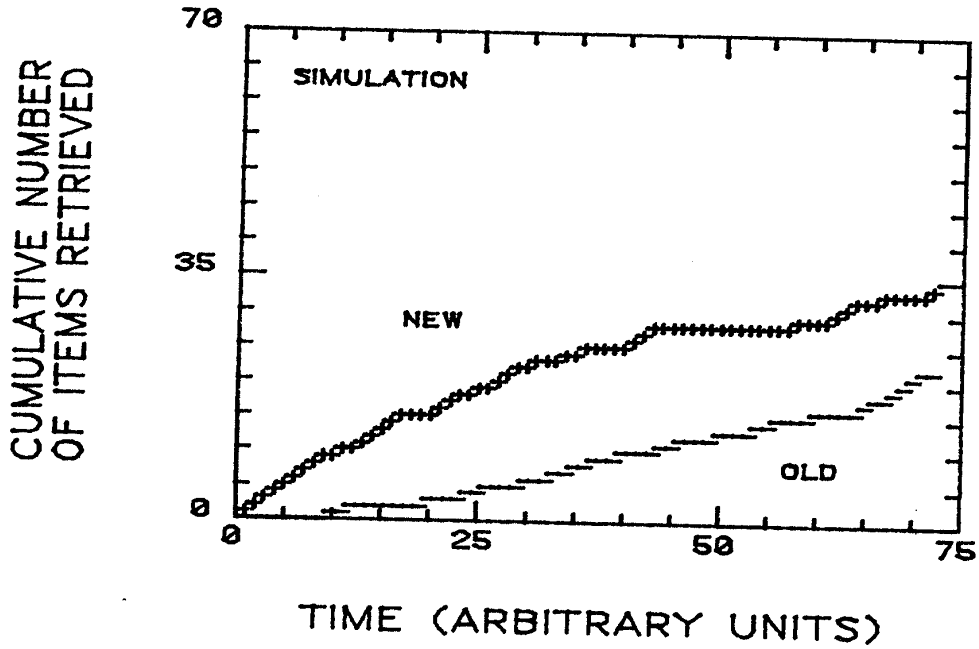
Figure 8. Mean cumulative number of clauses produced during script production as a function of time by script. The numbers next to the curves denote the number of subjects whose data are being averaged to produce the points plotted in each section of the curves (both cumulative retrieval curves of a subject who described a visit to a grocery store twice were included).

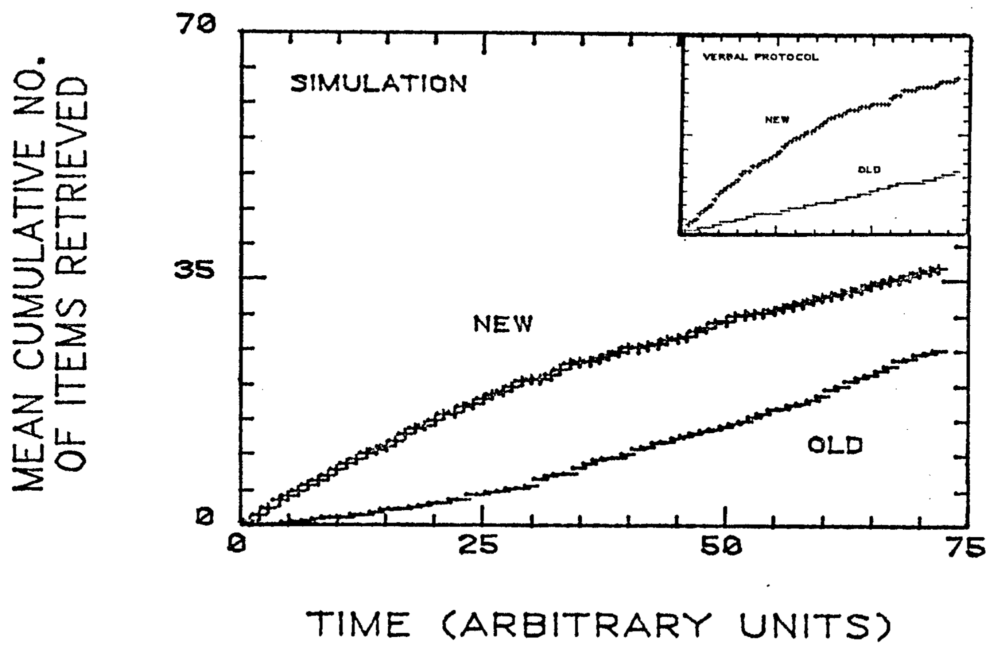

70
VERBAL PROTOCOL
NEW
35
OLD
0
CONTROL
NEW
35
OLD
0
0
250
500
750
MEAN CUMULATIVE NO. OF ITEMS RETRIEVED
TIME (SECONDS)


70
35
0
CUMULATIVE NUMBER
OF ITEMS RETRIEVED
NEW
OLD
0
250
500
750
TIME (SECONDS)

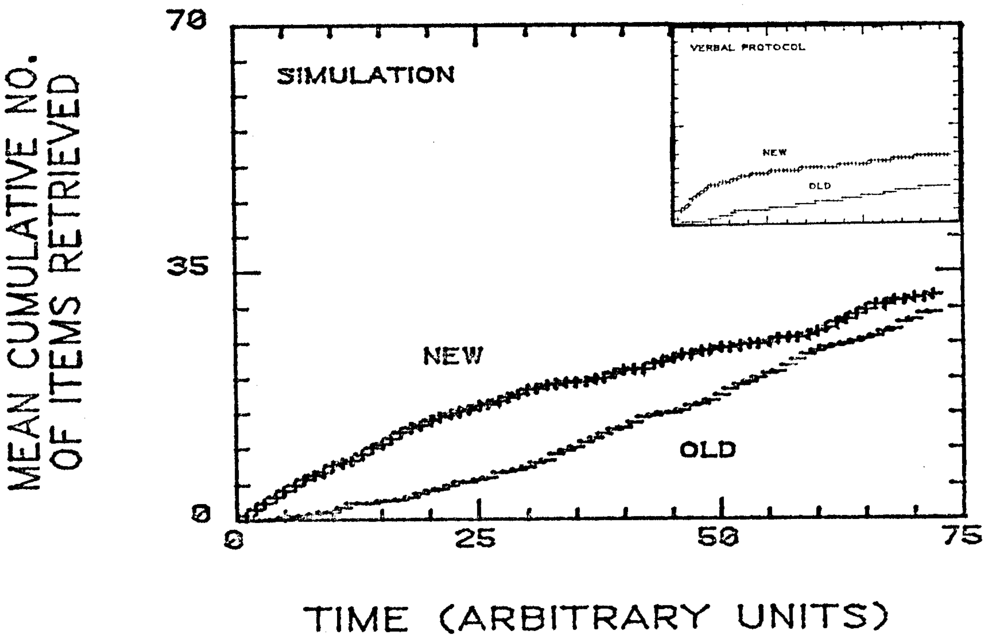


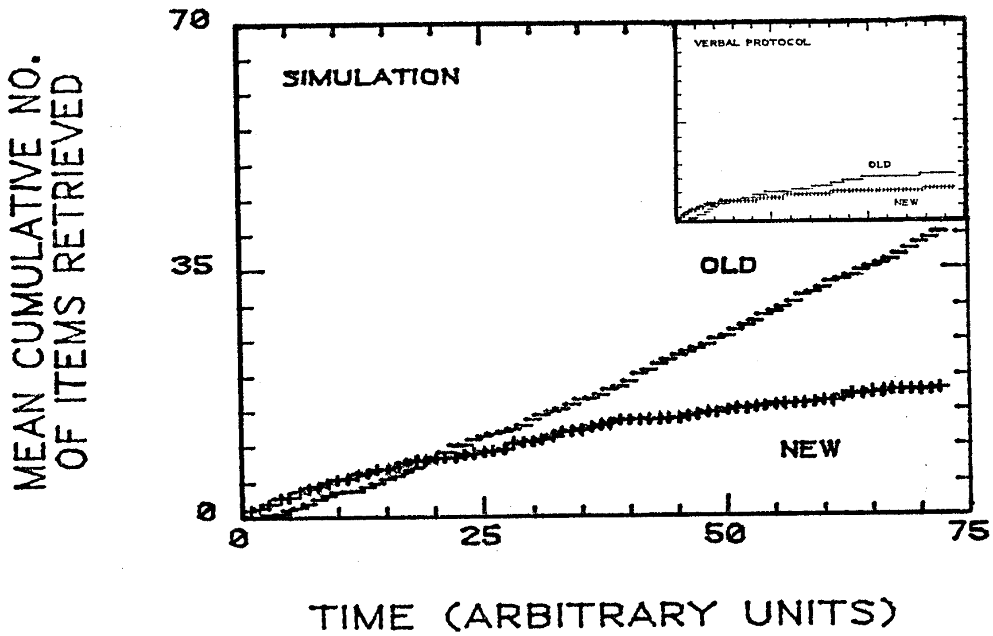
VERBAL PROTOCOL
MEAN CUMULATIVE NO. OF ITEMS RETRIEVED
70
35
0
0
250
500
750
TIME (SECONDS)


CUMULATIVE NUMBER OF ITEMS RETRIEVED
70
35
0
SIMULATION
NEW
OLD
0
25
50
75
TIME (ARBITRARY UNITS)

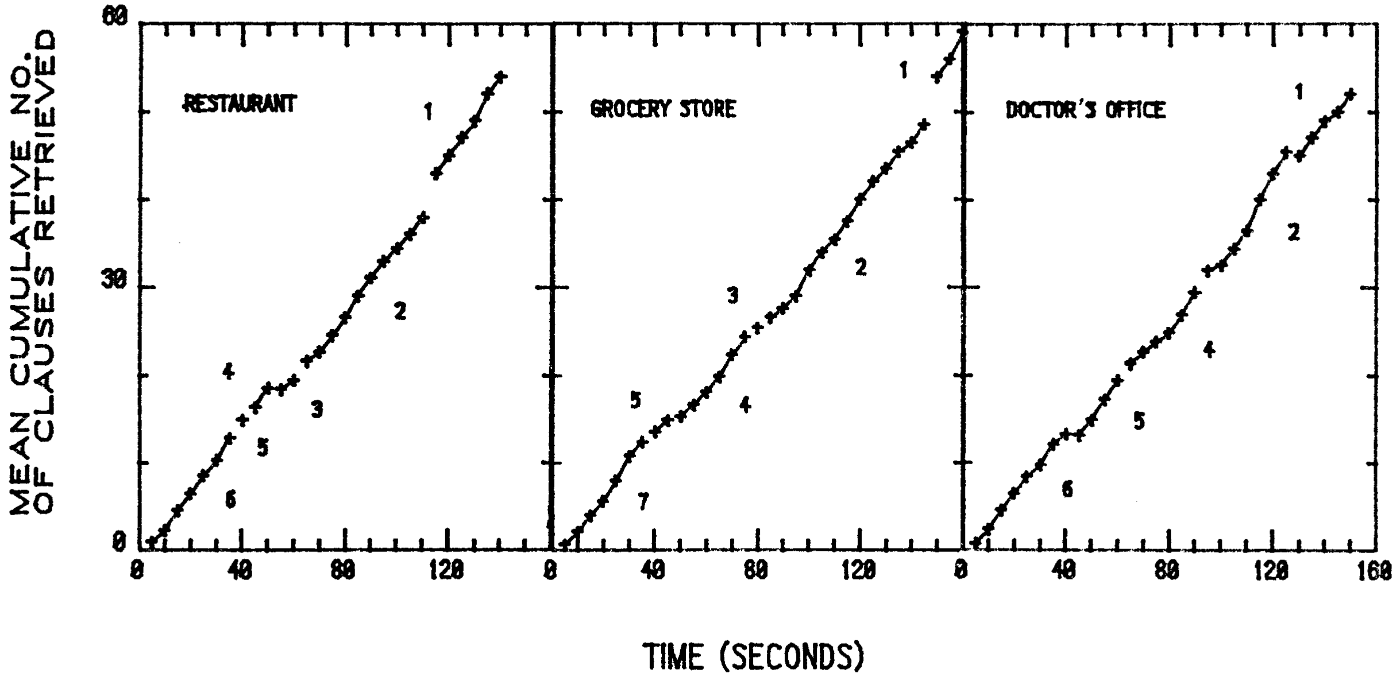


70
SIMULATION
VERBAL PROTOCOL
NEW
OLD
35
NEW
OLD
0
0
25
50
75
MEAN CUMULATIVE NO. OF ITEMS RETRIEVED
TIME (ARBITRARY UNITS)


SIMULATION
VERBAL PROTOCOL
NEW
OLD
NEW
OLD
70
35
0
0
25
50
75
MEAN CUMULATIVE NO. OF ITEMS RETRIEVED
TIME (ARBITRARY UNITS)



MEAN CUMULATIVE NO.
OF ITEMS RETRIEVED
TIME (ARBITRARY UNITS)
SIMULATION
OLD
NEW
VERBAL PROTOCOL
OLD
NEW
70
35
0
0
25
50
75


MEAN CUMULATIVE NO. OF CLAUSES RETRIEVED
0
30
60
RESTAURANT
GROCERY STORE
DOCTOR'S OFFICE
1
2
3
4
5
6
7
0
40
80
120
160
TIME (SECONDS)