) 38 p 40 FB see minor corrections! GB

The Formation of Verbal Schemas: Mediation and Interference Processes

Francis S. Bellezza

Ohio University

and

Gordon H. Bower

Stanford University

Running head: Verbal Schemas

Abstract

Four experiments ~~were performed to~~ studied the formation of memory schemas for sentences. Sentence frames composed of three concepts were created along with five instances for each sentence frame. During training, varying numbers of instances of each sentence frame were presented to subjects. In the critical test list which followed, ~~only~~ one instance of each frame was presented, and recall of only ~~the~~ this test list was required. The two transfer processes proposed by Thorndyke and Hayes-Roth (1979) were in evidence. Increasing the number of instances of a sentence frame in the training list had two separable effects: (a) It increased the probability that a schema was formed in memory representing the presented sentence frame. This schema acted as a mediating structure to facilitate recall of the schema-related sentence presented later in the critical test list. (b) ~~It~~ Prior schema-instances also increased the likelihood that details of sentences creating the schema would compete at recall with the ~~schema-related test sentence~~ instance of that schema presented during the critical test. This interference ~~process~~ among details diminished recall of the specific sentence presented in the test list. These two processes, one positive and one negative, combined to determine recall performance. ~~Other factors manipulated in~~ The experiments ~~were~~ also varied the spacing and variability of sentence-frame exemplars on the training list and the ~~temporal interval~~ delay between the presentation of the training and test lists. A mathematical model is presented which generally fits the pattern of the data collected in Experiments 3 and 4. ~~Other~~ The proposals of Thorndyke and Hayes-Roth regarding schema formation are also discussed.

The Formation of Verbal Schemas: Mediation and Interference Processes

In recent years investigators in the areas of cognitive psychology, artificial intelligence, and social psychology have come to believe that organized knowledge structures in memory play an important role in how people perceive, comprehend, and remember information (Bobrow & Norman, 1975; Graesser & Nakamura, 1982; Minsky, 1975; Rumelhart & Ortony, 1977; Schank & Abelson, 1977; Taylor & Crocker, 1981). we will refer to These knowledge structures ~~are referred to here~~ as memory schemas. A memory schema can be thought of as a prototype for a class of objects, persons, situations, events, sequences of events, actions, or sequences of actions. It seems that a new schema can be created in memory by the repeated occurrence of experiences that are in some way variations on a common structure ~~similar to each other~~. The invariant characteristics of these experiences are abstracted and stored in memory by mechanisms which are not yet well understood by memory investigators. Memory schemas can be extended to help ~~allow~~ a person ~~to~~ perceive, comprehend, and remember novel experiences which have not ~~never~~ been previously encountered. ~~This cognitive processing can occur because~~ schemas are available to structure an experienced event, even though some details of that event may be ~~are~~ novel.

Memory schemas play an important role in remembering information. For example, if an opaque ~~a~~ passage ~~is presented~~ describes ~~describing~~ the actions associated with ~~the~~ washing one's ~~of~~ clothes, subjects reading the passage may neither understand nor remember the

(Bransford + Franks, 19xx)

information presented, unless they are informed beforehand what the passage is about. Unless the appropriate knowledge structure is first activated, no referential situation can be called up to interpret the information presented (Bransford & Johnson, 1972). A schema provides an "ideational scaffolding" (Ausubel, 1963; Anderson, Spiro, & Anderson, 1978) to which new information can be associated. When a set of information is interpreted in terms of a particular schema, it "instantiates" (instances) that schema (Anderson, 1978). Furthermore, the person reading or hearing the information can use the activated schema to make inferences. Thus, when a passage is presented based on a restaurant script (Schank & Abelson, 1977), experimental subjects will often assume that the person described as eating in a restaurant looked at a menu, even though this was not mentioned in the passage. By making inferences from the activated schema, default values can be ~~generated~~ expected for objects, persons, and actions not explicitly described. When asked to indicate to whom the order for a meal was given, subjects may indicate that it was a waiter or waitress, even ~~if~~ though this information was not provided in the passage.

Another important characteristic of schema-based remembering is that multiple events, each of which instantiates the same schema, may nevertheless have their idiosyncratic information "tagged" so that the events are to some degree discriminable in memory (Bower, Black, & Turner, 1979; Graesser & Nakamura, 1982; Smith & Graesser, 1981; Thorndyke & Hayes-Roth, 1979). Early formulations of the role of schemas in remembering tended to minimize the degree to which this discrimination took place (see Alba & Hasher, 1983, for a review of this literature). Yet ~~the capacity of the memory system~~ people's ability to discriminate among stored instances of a schema is an important capability. Further discussion of discrimination among multiple instantiations is given below.

Research on memory schemas has increased dramatically in the last few years (Alba & Hasher, 1983), but there are relatively few studies exploring the formation of memory schemas. In these studies the development of schemas used in categorizing random visual patterns typically has been monitored (Anderson, Kline, & Beasley, 1979, 1980; Franks & Bransford, 1971; Posner & Keele, 1968, 1970; Reed, 1972). In this research subjects were trained to classify a number of visual stimuli into a smaller number of categories. The ~~degree of~~ development of a schema in memory was determined by how well subjects classified stimuli never presented before.

Research using categorization judgments of verbal material in the form of sentences has also been studied (Elio & Anderson, 1981) but less frequently than visual material. Our purpose here is to further investigate the formation of verbal schemas. In the experiments to be presented here, verbal material in the form of sentences was used to investigate the development of memory schemas. Moreover, the degree of schema formation was assessed not by categorization of instances but by the recall of presented instances of the schema (Bower, 1974; Hayes-Roth & Thorndyke, 1979; Thorndyke & Hayes-Roth, 1979). Thorndyke and Hayes-Roth

(1979) reported an important set of experiments of this type. They presented subjects with a series of passages. Some of these passages were related to one another, such as passages about different constellations of stars.

Thorndyke and Hayes-Roth found that manipulating the degree of similarity of the content of passages affected two processes involved in recall. First, recall was enhanced through experience, because as more instances of similar material were presented, a memory schema was formed for this material. This schema acted as a memory structure into which new instances of the schema could be assimilated. Later, retrieval of the schema from memory enabled the information that recently instantiated it to be recalled. The schema acted as a mediator for information that fit it. Second, the process of schema formation also interfered with the recall of specific instances of the schema. As additional instances of the same schema were presented, the subjects found it increasingly difficult to discriminate between the information presented most recently and the information presented earlier in training when the schema was being formed.

Hence, presenting passages of the same kind had two contrary effects on recall, one positive and one negative. The similarity of related information allowed the subjects to abstract from it common concepts and form memory schemas. This provided a schema in memory to mediate later learning. But at the same time this training information interfered with the learning of new specific instances of the schema. The problem of explaining both the positive and negative effects of stimulus similarity on learning is not a new problem in psychology (Osgood, 1949). To get around this theoretical impasse, Thorndyke and Hayes-Roth assumed that schema development preceded in time the interfering factor of accumulated details. From this assumption they expected that recall of the most recent schema instance would first increase as a function of the number of previous instances in training and then decrease. This decrease was expected to occur because the accumulating collection of details in memory associated with the schema interfered with further learning mediated by the schema. This was the result they obtained.

The goals of the present experiments were to attempt to replicate the results obtained by Thorndyke and Hayes-Roth (1979) and to provide further experimental support for the schema mechanisms they proposed. The materials used were different from those of Thorndyke and Hayes-Roth and were created so that experimental manipulations involving repetition and similarity could be easily performed. Sentence frames were made up by relating three randomly sampled categories of nouns (see Appendix A). To illustrate, one sentence frame was "The <u>public official</u> engaged in a financial transaction involving a <u>commercial business</u> on <u>some planet</u>." Thus, specific instances of the sentence frame above were "The mayor bought a store on Mars", "The senator sold a restaurant on Venus", and "The judge purchased a bank on Mercury". Five sentence exemplars of this kind were created for each sentence frame.

In the experiments, subjects were presented with two lists of sentences; the training list and the test list. Subjects were forced to process the training sentences in such a manner to ensure their comprehension. But when the test list was presented, subjects tried to memorize these sentences in preparation for a later series of recall tests. Only sentences from the test list were to be recalled. ~~It was~~ We expected that as the number of sentence exemplars from a sentence frame increased in the training list, the probability of forming a memory schema corresponding to that sentence frame would also increase. Because the memory schema represents an abstraction from the sentence exemplars, it was assumed to be comprised of associations among the ~~concepts~~ categories used to create the ~~sentence~~ common-action frame itself. The schema could then become a mediating structure available in memory for facilitating later schema-related learning. Increasing the number of sentence-frame exemplars in the training list should increase the probability that the corresponding test-list sentence is processed in terms of the memory schema.

As proposed by Thorndyke and Hayes-Roth (1979), formation of the memory schema should have a positive, mediating effect on the free recall of the schematic sentence presented on the test list. However, the presentation of multiple sentence exemplars in the training list should, in addition, interfere with later learning. Retrieval of the schema at the time of the recall test may produce many confusing details from the sentences presented earlier. This confusion ~~will result in poorer~~ can produce inter-list confusions, thus lowering recall of the details of the test sentence (a form of proactive interference). In summary, repetitions of a sentence frame on the training list should result in a positive transfer at the level of concepts, but in a negative transfer of specific ~~instances~~ details.

In their research Thorndyke and Hayes-Roth (1979) manipulated two experimental factors. One was the amount of training material (instances) provided for each schema; ~~and~~ the second was the time interval between the training material and the test material. ~~In the~~ Our experiments ~~presented here~~ manipulated these two ~~manipulations were used.~~ variables as well as ~~In addition,~~ two others ~~factors were also tested~~. One was the spacing ~~of~~ between schema instances in the training list. Spacing could be varied easily, because each schema instance was represented by a single sentence. ~~It was~~ We hypothesized that increased spacing of schema instances would make for stronger learning of the schema as well as better discrimination among ~~memory representations of these~~ instances ~~sentences easier and improve~~, thus improving recall of the specific instance ~~by reducing interference~~.

A second, additional factor was the variability of the exemplars. ~~Sometimes~~ In some conditions, only one sentence from a sentence frame was repeatedly presented in the training list; ~~whereas~~ in ~~another~~ other conditions, up to four different examples of the sentence frame were used. Our hypothesis was that presentation of only one sentence would ~~inhibit~~ retard the formation of a generalized memory schema. ~~It seemed that,~~ for abstraction and schema formation to occur, a variety of instances of the same conceptual relations must occur. With low variability in the training list it was expected that test-list recall would be less than with high variability. With low variability the schematic structures needed to support recall would ~~not be~~ be less available.

Experiment 1

__Method__

__Materials.__ One hundred and eleven categories of five words each were selected. ~~Some~~ Most of these categories were taken from Battig and Montague (1969), such as articles of furniture, fruits, and types of music. Also, some of the Battig and Montague categories were divided into subcategories, such as precious versus nonprecious metals, individuals versus team sports, and song birds versus predatory birds. Other categories not in ~~their~~ those norms were used, such as royal personages, brands of automobiles, and types of soil. Ninety ~~of these~~ categories in total were used to form the 30 sentences in the main list, and 21 categories were used to create filler sentences. The categories are listed in Appendix A.

The ninety categories were randomly grouped into thirty sets of three categories each. Because each category contained five instances, five similar sentences could be made by choosing one instance from each category and then adding common-action verbs and other necessary words. The common-action verbs used in the five sentences were different but similar in meaning. These five sentences defined a ~~sentence~~ common-action frame ~~and its corresponding sentence schema.~~

Two sets of five sentences created ~~for~~ from six ~~particular~~ categories are shown in Table 1.

Insert Table 1 about here.

The procedure used for creating the sentences in the main list was also used to create the seven sets of filler sentences. Each of these sets also contained five sentences.

Lists. A training list was made up ~~by sampling 18 sets~~ of ~~sentences, that is, 18 sentence frames.~~ From six of the frames no ~~sentence~~ was included in the training list; from another six frames three ~~sentences~~, and from final six frames four ~~sentences~~. ~~But~~ each of these four sentences was presented twice in the training list, and each sentence was presented before any were repeated. ~~Hence,~~ the number of repetitions of each sentence frame ~~in the training list~~ was either 0, 3, or 8. Thorndyke and Hayes-Roth (1979) also used a range of 0 to 8 repetitions.

The second factor manipulated was the spacing of exemplars in the training list. Each sentence examplar was separated from the other exemplar sentences from the same frame by 0 or 3 other sentences. For those six sentence frames from which no sentence was presented, two of the frames were arbitrarily designated as representing a spacing of zero, three, or eight intervening sentences. This was done ~~for~~ the statistical design, although the "spacing" of ~~no~~ presentations is undefined. The training list contained a total of 90 sentences. Sixty-six of these represented sentence frames upon which the subjects were being trained, and 24 were filler sentences used to fill ~~any~~ gaps remaining in the list caused by the spacing manipulation.

The test list was composed of 28 sentences. The first five sentences and the last five sentences were filler ~~sentences~~, providing primacy and recency buffers. The middle 18 sentences had not been presented in the training list, but 12 of them were new instances of the 12 sentences frames presented in the training list. In addition, six sentences were included that represented those six frames never shown in the training list. These six sentence frames represented the zero-repetition condition.

A computer program was used to create the lists and print them for each subject. The program was written to select randomly common-action ~~sentence~~ frames for each condition and to select randomly individual sentences from each of the frames chosen. Also, the arrangement of the spacing and repetition conditions was different in each form of the list.

<u>Procedure</u>. Subjects were tested individually using a memory drum which presented the sentences one at a time. Twelve different training and test forms were created and two subjects were tested on each form. The sentences from the training list were presented for 10 seconds each, and the subject had to rate each sentence on a 5-point comprehens~~ion~~ibility scale. A rating of 5 indicated that the sentence was very easy to comprehend, and a rating of 1 indicated that the sentence was very difficult. Immediately after the training list was presented, the subject was informed that for the next list, the test list, comprehens~~ion~~ibility ratings were ~~would~~ not required ~~have to be made~~. Rather, each sentence had to be carefully studied to prepare for a later free recall test. Each sentence in the test list was shown for 10 seconds. Following the presentation of the test list, subjects wrote down in any order as many of the test sentences as they could recall ~~remember~~. A second, fill-in test followed free recall in which each sentence from the test list was presented with only the third noun missing. The

subject had to write in the missing noun. ~~This was the fill-in test.~~

Results

Scoring procedures. In scoring free recall all the filler sentences were ignored. Also, each test sentence was scored with regard to only the three nouns. The verbs and other types of words were ignored. We also ignored the order of the nouns so that active-to-passive shifts were considered as correct recalls. One potential problem was that subjects might recall entire sentences from the training list rather than from the test list. If this occurred, ~~the~~ a simple measure of recall ~~would~~ does not represent how the training list influenced the learning and recall of the test list, but rather would indicate that the subjects could not distinguish between the two lists. To determine the degree to which the training and test lists were ~~confounded~~ confused in recall, each sentence recalled in the free-recall test was placed into one of four categories. A sentence could be a fragment containing only one or two nouns instead of three, a sentence could contain the three correct nouns from the test list, a sentence could contain three incorrect nouns but from the correct categories, and, finally, a sentence could contain a mixture of correct nouns and incorrect nouns from the correct categories. Table 2 shows the proportion of sentences presented on the critical test list that were recalled in each of these categories.

Insert Table 2 about here.

Almost no sentences were recalled that did not fit into one of these four categories. That is, almost no sentences were recalled that represented combinations of two or more sentence frames. The number of sentence fragments recalled was quite small, as was the recall of sentences containing all category intrusions. For most of the sentences the nouns recalled were either all correct or were a mixture of correct nouns and intrusions from the correct categories. This was also true of the cued-recall procedure used in Experiments 2, 3, and 4, as shown in the lower part of Table 2. Hence, it can be concluded that subjects were recalling sentences primarily from the test list, although errors were occurring in these sentences because of intrusions of similar nouns from the training list. Almost all category intrusions were nouns presented in the training list.

In order to use all the data from Table 2, including data from the sentence fragments and from sentences containing a mixture of correct and incorrect nouns, the analyses reported here are based on the overall proportion of nouns correctly recalled from each experimental condition. Because the proportion of ~~correctly~~ perfectly recalled sentences was low, especially in the free-recall condition, the proportion of nouns recalled provided more useable data than did complete sentence recall. Table 3 shows the frequency of correct nouns, category intrusions, and other

---

Insert Table 3 about here.

---

nouns recalled in the free-recall and cued-recall conditions of the four experiments. A preponderance of recalled nouns were correct nouns, but with a significant proportion representing category intrusions. Few other types of nouns, such as extracategory intrusions, were recalled.

<u>Measures of recall</u>. In Experiment 1, the two measures taken ~~of recall~~ were ~~used~~: free recall followed by a fill-in test in which the subject had to fill in the noun missing at the end of each test sentence. As argued by Thorndyke and Hayes-Roth (1979), successful free recall of a test sentence is dependent on <u>both</u> retrieval of the schema representing the sentence frame <u>and</u> successful discrimination of ~~among~~ the most recent ~~various~~ nouns associated with the schema. The fill-in test, however, seems ~~to us~~ much less dependent on the previous formation of a schema. The large number of recall cues presented for each test sentence in the fill-in task meant that discrimination of the most recent instances of words ~~among potential words stored~~ in memory was more important than retrieving an intact schema.

A second aspect of the scoring procedure was that we scored not only whether the correct nouns were recalled but also whether the correct noun category was recalled regardless of the correctness of the specific word. For example, if the test sentence was "The senator sold a restaurant on Venus," but ~~and~~ the subject recalled Mars for Venus, then this would be scored as recall of the correct category for the third noun but not the correct word. These ~~so-called~~ intracategory intrusions were expected to occur increasingly as the number of sentences representing a sentence frame increased in the training list. A large number of correct category recalls indicates that a schema for the sentence frame has been created. A large difference between correct recall and category recall ~~means~~ that the subject was ~~unable to discriminate~~ in memory the nouns stored for the training sentences ~~from the nouns~~ for the test sentence.

__Sentence ratings__. Table 4 shows the mean comprehensi~~on~~ ratings obtained at each repetition of a sentence frame in the training list.

Insert Table 4 about here.

The effect of repetitions was significant, $F(7, 151) = 6.04$, $MSe = .110$, $p < .001$. However, this increase occurred only after the fourth sentence was presented, and the training sentences from each frame began to be repeated. A trend analysis on Repetitions 1 to 4, on which the sentence exemplars were different, showed no significant linear trend.

__Correct recall__. The proportion of nouns correctly recalled and the proportion of nouns recalled from the correct category are shown in Table 5 for each repetition condition.

Insert Table 5 about here.

The effects of training repetitions on the proportion of nouns recalled in free recall was marginally significant, $F(2, 46) = 3.11$, $MSe = .033$, $p = .05$, but the linear component of a trend analysis was [more] significant, $F(1, 46) = 6.22$, $MSe = .033$, $p < .025$.

The proportion of [most-recent instance] nouns correctly recalled in the fill-in test was also affected by the number of repetitions of the sentence frame in the training list, $F(2, 46) = 7.97$, $MSe = .081$, $p < .005$. As can be seen in Table 5, the proportion of ~~correct~~ [recent-instance] nouns recalled *decreased* as the number of ~~repetitions~~ [prior instances of the frame] *increased*.

Category recall. When the proportion of words recalled from the correct category was used as a measure of free recall, the effect of repetition was significant, $F(2, 46) = 7.87$, $MSe = .063$, $p < .005$. When recall of nouns from the correct category was used as a measure for the fill-in task, recall performance *increased* as ~~the function of~~ the number of repetitions *increased*, $F(2, 46) = 10.31$, $MSe = .042$, $p < .001$. In the four analyses of variance performed in Experiment 1, spacing was never a significant factor nor was the Spacing X Repetition interaction. [Therefore, we will present none of the results divided according to the spacing of the repetitions.]

Discussion

The results of Experiment 1 provided some evidence for the mediation and interference processes proposed by Thorndyke & Hayes-Roth. The notion that recall was facilitated by mediating schematic structures was supported by a number of results. As the number of repetitions of exemplar sentences increased in the training list, the proportion of ~~correct~~ [recent-instance] nouns free recalled from the test list also increased. The formation of a schema for a sentence frame during training provided a mediating structure by which the appropriate sentence from the test list could be learned. Each newly formed schema could be retrieved and used to recall its corresponding test sentence during the free-recall task.

Recall of category intrusions also provided evidence of schema formation. Both in the free-recall task and in the fill-in task, ~~recall~~ intrusions of earlier ~~other~~ instances from the category ~~intrusions~~ increased with repetitions. This meant that subjects knew what category of word to recall even if they could not call the ~~correct~~ most recent noun. Recall of category intrusions indicates that schema abstraction had taken place in the sense that sets of categories had become part of the schemas corresponding to sentence frames. This abstraction occurred even though no category labels were ever presented.

The results of Experiment 1 also provided evidence for the interference ~~process~~ in list-discrimination proposed by Thorndyke and Hayes-Roth (1979). Correct recent-instance responses in the fill-in test declined with repetitions rather than increased as in the free-recall task. This result is not surprising. Recall in the fill-in test was not strongly dependent on a well-formed schema being ~~formed~~ available in memory, because a strong retrieval cue was provided. On the other hand, the major task of the subject in the fill-in task was to discriminate between the noun presented on the (most recent) test list and those presented earlier on the training list. However, repetitions of the sentence frame in the training list strengthened many same-category words in memory so that this discrimination was ~~difficult~~ error-prone. ~~could not take place~~.

Some results of Experiment 1 offer no support for the idea that a memory schema was formed during training. Comprehensi~~on~~bility ratings for the different exemplars from each sentence frame did not increase as additional exemplars were presented. If schema formation were taking place, we would expect each additional sentence instantiating the schema to be more easily comprehended. Second, no U-shaped curve was found relating recall performance to the number of training repetitions. Thorndyke and Hayes-Roth proposed interference effects follow in time the effects of schema formation to produce a U-shape curve.

A surprising result from Experiment 1 was that the spacing of repetitions in the training list had no effect on either comprehension ratings or on any of the measures of recall performance. We expected an effect because spacing of repetitions should make each sentence-frame instance temporally distinct. Perhaps the spacing values of 0 and 3 intervening items were not different enough to affect performance.

<u>Experiment 2</u>

The U-shaped curve relating recall performance to number of repetitions (Thorndyke & Hayes-Roth, 1979) was not found in Experiment 1. Experiment 2 represented a second attempt to produce a U-shaped recall curve. Also, visual-imagery ratings were used rather than comprehension ratings to try to detect any increase in ease of processing additional exemplars from the same frame. The number of repetitions of sentence frames in the training list was again 0, 3, or 8. But because spacing of repetitions was found not to be a significant factor in Experiment 1, the variability of repetitions, rather than their spacing, was manipulated in Experiment 2. In Experiment 1 the levels of free recall were not high, so in Experiment 2 a cued-recall test was used rather than a free-recall test. During testing the first noun from each test sentence was presented to the subject. From this cue the subject had to recall the complete sentence. [these tests] ~~This test~~ of cued-recall [were] ~~was~~ followed by the fill-in test[s] used in Experiment 1.

Method

The procedure used in Experiment 2 was the same as in Experiment 1 with the exceptions listed below. The training list contained 94 sentences with three filler sentences at the beginning and end of the list. Filler sentences were not needed within the list, because the spacing between exemplars representing the same sentence frame was always zero. The test list was 30 sentences long with three filler sentences at the beginning and at the end of the list. [explain constant vs variable repetitions in here] Each of the six experimental conditions was replicated four times within the lists. The sentences in both lists were presented for 8 seconds each. ~~and~~ [D]uring the training list the subject gave a 5-point imagery rating for each sentence. A rating of 5 indicated that it was very easy to form a visual image of the sentence, and a rating of 1 indicated that it was very difficult to form a visual image. Following presentation of the test list, subjects [received] ~~were administered~~ the cued-recall test followed by the fill-in test.

Results

Sentence ratings. The effect of training repetitions [on imagery ratings] was significant, $F(7, 161) = 34.13$, $MSe = .105$, $p < .001$, as was the Repetition X Variability interaction, $F(7, 161) = 5.00$, $MSe = .092$, $p < .001$. As can be seen in Table 4, when the exemplar from a sentence frame was always the same in the training list, the mean imagery ratings increased at a faster rate than when more than one ~~exemplar~~ instance was used. However, even in the variable-exemplar condition, the imagery ratings increased over the first four sentence ~~exemplars~~ instances, which were all different sentences. A trend analysis showed the linear component of the curve to be positive and significant, $F(1, 161) = 12.02$, $MSe = .092$, $p < .001$.

Correct recall. Repetition had no significant effect on the proportion of nouns correctly recalled in the cued-recall test, although the means tended to increase with training repetitions (see table 5). Analysis of the fill-in test showed that repetition had a significant effect, $F(2, 46) = 4.88$, $MSe = .047$, $p < .025$. As was found in Experiment 1 and shown in Table 5, the proportion of recent-instances ~~nouns~~ correctly recalled *decreased* as the number of repetitions *increased*. There was no significant effect of variability of instances on either measure of correct instance recall.

Category recall. Using the proportion of words recalled from the correct category in the cued-recall task, repetition had a positive and significant effect, $F(2, 46) = 7.25$, $MSe = .058$, $p < .005$.

Scoring the fill-in test as to whether the nouns recalled were from the correct category showed that repetition, $F(2, 46) = 17.49$, $MSe = .038$, $p < .001$, and variability, $F(1, 23) = 9.24$, $MSe = .064$, $p < .01$, were significant. Table 5 shows that category recall increased with the number of repetitions. In addition, when only one exemplar was used in the training list, recall of the correct category was .62. But when the exemplars used were varied, recall was .75.

Discussion

The results of Experiment 2 were similar to those of Experiment 1. Using a cued-recall procedure rather than a free-recall procedure did not produce the U-shape curve of Thorndyke and Hayes-Roth (1979). Experiment 2, like Experiment 1, however, did produce data that demonstrate the schema-based mediation and transfer processes proposed by Thorndyke and Hayes-Roth. The mean imagery ratings increased with repetition. This effect did not occur with the comprehension ratings of Experiment 1. This was true even though the first four repeated sentences from the same sentence frame in the variable condition had almost no specific words in common. This increase in ratings indicated that information from a currently developing schema could be utilized even as the training sentences were presented.

The interference process was in evidence also. Correct responses in the fill-in task decreased with repetitions in the training list, even though the proportion of correct category recalls increased. The larger number of training sentences produced a greater number of specific interfering nouns.

Variability also affected performance in Experiment 2, but not in the manner we expected, for variability did not improve recall accuracy. However, subjects did find it easier to form a visual image for those sentences repeatedly presented compared to varied sentence-frame exemplars. With the same sentence repeatedly presented for each frame, the subjects could concentrate on improving their one image rather than having to change the components of the image.

Also, recall of any word from the correct category was better under variable exemplars. So subjects were more cognizant of the category of word needed in the variable-exemplar condition than in the constant-exemplar condition. Paradoxically, the larger number of category words did not diminish recall of the correct word. It may be that in the variable-exemplar condition memory schemas were better formed than in the constant-exemplar condition. Correct recall may not have been better with variable exemplars because both schema formation and interference occurred more quickly and offset one another.

Experiments 3 and 4

In the two experiments presented so far, the time interval separating the training and test lists was not manipulated. Thorndyke and Hayes-Roth (1979) reported that recall performance on the test list increased when a 24-hour time interval was interposed between the training and testing phases of the experiment. Experiments 3 and 4 were similar to Experiments 1 and 2 except that in Experiment 4 the test list was presented 24 hours after the training list. It was expected that recall performance would be better in Experiment 4 than in Experiment 3. Furthermore, correct recall for the fill-in task was expected to increase with training repetitions rather than decrease. After 24 hours the nouns used in establishing the sentence schema should be temporally well-differentiated from those in the test list. ~~Because of this~~ Therefore, proactive interference should be reduced.

Method

The materials used in Experiments 3 and 4 were the same as those used in Experiments 1 and 2. Three different factors determined the composition of the training list. The number of repetitions of each sentence frame was either 0, 3, or 8. The spacing of exemplars from the same sentence frame was either 0 or 3. The third factor was variability of the repetitions. Either ~~they were~~ all the same sentence or were maximally different, as explained in the Method section of Experiment 2. The training list consisted of 70 sentences, 26 of which were filler sentences. The main list was 18 sentences long including three filler sentences at the beginning and end of the list. Each of the 12 experimental conditions was represented by one sentence frame. Twelve different forms of the matched training and test lists were made up, and two subjects were tested on each form. Subjects rated each sentence in the training list for imagery using a 5-point scale. In the test list subjects simply studied the sentences in anticipation of a later recall test. The sentences in both training and test lists were each presented for 10 seconds. Three different tests of recall were used. Immediately after the presentation of the test list each subject was administered the test of free recall, then cued recall, and then the fill-in test. The subjects were given as much time as needed to complete each test.

Experiment 4 differed from Experiment 3, ~~because~~ after rating the sentences on the training list, each subject was dismissed. The subject returned 24 hours later, was presented the test list for study, and then was tested on the three recall measures. Also, the set of list forms used in Experiment 4 were different from those of Experiment 3.

Results

<u>Sentence ratings</u>. In Experiment 3 both the variability factor, $F(1, 23) = 12.71$, $MSe = 3.17$, $p < .001$, and the Repetition X Variability interaction, $F(7, 161) = 2.57$, $MSe = .222$, $p < .025$, significantly influenced imagery ratings ~~were significant.~~ The mean ratings are shown in Table 6.

Insert Table 6 about here.

When the sentence exemplars for a sentence frame were the same during training, the mean rating obtained was 3.91. When the exemplars were maximally different, the mean rating was 3.26. As can be seen in Table 6, the significant Repetition X Variability interaction seems to be caused by the ratings for the constant exemplars increasing at a rate faster than that found with variable exemplars. A trend analysis on the first four repetitions in the variable-exemplar condition showed no significant linear trend.

For Experiment 4 the imagery ratings ~~results~~ in Table 6 formed a pattern similar to that of Experiment 3. However, in Experiment 4 only repetition was a significant factor, $F(7, 161) = 8.59$, $MSe = .174$, $p < .001$. Furthermore, a trend analysis on the first four repetitions of the variable-exemplar condition showed a significant linear trend, $F(1, 168) = 5.16$, $p < .025$.

Correct recall. In Experiments 3 and 4 three tests of recall were administered: first free recall, then cued-recall using the first noun in each test sentence, and then a test in which the subject had to fill in the last noun of an otherwise complete sentence from the test list. The proportion of words recalled in each repetition condition is shown in Table 7.

Insert Table 7 about here.

In Experiment 3 analysis of the proportion of correct nouns recalled in the free-recall test showed that training repetitions ~~was~~ had a statistically significant ~~factor~~ effect, $F(2, 46) = 9.75$, $MSe = .133$, $p < .001$. As can be seen from Table 7, recall of nouns from the test list increased with the number of repetitions of the sentence frame in the training list. In Experiment 4 the proportion of nouns correctly recalled in the free-recall test also increased with repetitions, $F(2, 46) = 12.12$, $MSe = .097$, $p < .001$.

In Experiment 3 analysis of the proportion of correct nouns recalled in the cued-recall test showed that repetition was significant, $F(2, 46) = 4.56$, $MSe = .145$, $p < .05$. As can be seen in Table 7, cued recall of the nouns first increased and then decreased. Tukey's HSD test (Kirk, 1968), however, showed no significant differences among the three means. The factor of variability was also significant, $F(1, 23) = 4.96$, $MSe = .137$, $p < .05$. When sentence frames were represented by the same sentence, the proportion of nouns recalled from the test list was .39, whereas when variable exemplars were used, the proportion was .29. The cued-recall measure used in Experiment 4 gave results somewhat different from those of Experiment 3. The proportion of correct nouns recalled increased with repetitions, $F(2, 46) = 19.59$, $MSe = .161$, $p < .001$. Also, variability was not a significant factor.

Analysis of the proportion of correct nouns recalled in the fill-in test of Experiment 3 showed variability to be significant, $F(1, 23) = 6.97$, $MSe = .220$, $p < .05$. When constant exemplars were used in the training list, the proportion of nouns correctly recalled was .56, but when exemplars were variable, the proportion was .42. There was also a significant Repetition X Spacing interaction, $F(2, 46) = 4.25$, $MSe = .202$, $p < .05$. Post tests on the interaction means (Cicchetti, 1972), however, revealed no significant differences, and the differences among means displayed no meaningful patterns.

Category recall. Analysis of the propportion of nouns free-recalled from the correct category in Experiment 3 showed a significant effect of repetition, $F(2, 46) = 13.62$, $MSe = .181$, $p < .001$, with recall increasing with number of repetitions. This relation was also found in Experiment 4, $F(2, 46) = 17.32$, $MSe = .138$, $p < .001$.

When cued-recall in Experiment 3 was scored as to whether each noun came from the correct category, only repetition was significant, $F(2, 46) = 10.58$, $MSe = .194$, $p < .001$. In Experiment 4 repetition was also significant, $F(2, 46) = 12.12$, $MSe = .121$, $p < .001$.

Analysis of the proportion of correct categories recalled on the fill-in test of Experiment 3 indicated that repetition was a significant factor, $F(2, 46) = 10.92$, $MSe = .162$, $p < .001$. Also significant was the Repetition X Spacing interaction, $F(2, 46) = 5.29$, $MSe = .073$, $p < .005$. Post tests on the interaction means (Cicchetti, 1972) showed only one significant difference. With zero spacing, category recall was .48, .92, and .81 for 0, 3, and 8 training repetitions, respectively. With a spacing of three intervening sentences, category recall performance was .65, .75, and .88. This interaction may be spurious, because with zero training repetitions no sentence-frame exemplars were presented in the training list, and these obviously could not be spaced. Yet, with zero repetitions, zero spacing resulted in a recall value of .48 and a spacing of three resulted in a value of .65. The Repetition X Spacing interaction would seem to be spuriously significant ~~by chance.~~

In Experiment 4 the fill-in test resulted in three significant main effects. Repetition was significant, $F(2, 46) = 22.33$, $MSe = .122$, $p < .001$. As can be seen from Table 7, recall increased from 0 to 3 spacings but not from 3 to 8. Also significant was spacing, $F(1, 23) = 11.50$, $MSe = .121$, $p < .005$. When the spacing of exemplars was 0, the proportion of categories correctly recalled from the test list was .81. When the spacing was 3, this value was .57. Finally, variability was significant, $F(1, 23) = 5.66$, $MSe = .157$, $p < .05$. If the exemplars presented were the same, then recall was .69. If they were variable, then recall was .80.

Discussion

There were a number of similarities in the recall patterns found in Experiments 3 and 4, even though in Experiment 3 presentation of the test list and its recall took place immediately after presentation of the training list, and in Experiment 4 presentation and ~~testing~~ recall of the test list was delayed for 24 hours. The proportion of nouns correctly recalled in the free-recall test was .22 in Experiment 3 and .23 in Experiment 4. The proportion of nouns correctly recalled in the cued-recall tests were .34 and .36, respectively. In the fill-in test the proportions were .49 and .52, respectively.

Furthermore, many of the results of Experiments 3 and 4 parallel those of Experiments 1 and 2. In general, correct recall and category recall increased with repetitions. The purpose of Experiments 3 and 4 was to determine if any of the recall tests show an advantage of Experiment 4 over Experiment 3 resulting from the 24-hour interval between the training and test lists. One difference appeared. In Experiment 3 correct recall in the fill-in task was .50, .52, and .45 with 0, 3, and 8 repetitions, respectively, of the sentence frame in the training list. These results correspond to the general decrease with repetitions found in Experiments 1 and 2. In Experiment 4 with its 24-hour delay, however, these proportions were .44, .59, and .54, indicating a general increase with repetitions.

To test for the presence of this interaction statistically, the proportion of nouns correctly recalled on the fill-in test from both experiments was reanalyzed adding delay as a factor.

The Repetition X Delay interaction was not significant, $F(1, 92) = 2.25$, but a trend analysis showed that the Linear Component X Delay interaction was significant, $F(1, 92) = 3.87$, $MSe = .218$, $p < .05$, using a one-tailed test. The hypothesis tested was that recall decreased with repetitions in the 0-hour delay condition but increased with repetitions in the 24-hour delay condition.

It is not clear why repetitions may have interacted with delay of learning in the fill-in test but not in the tests of free and cued recall. It may be that the fill-in test maximizes the recall of category intrusions because the sentence with only one noun missing provides such a strong set of recall cues.

Although the levels of performance were similar in the two experiments, the degree of variability of the exemplars in the training list did have differential effects. In both experiments constant exemplars resulted in higher imagery ratings than did variable exemplars. In Experiment 3 in which the test list immediately followed the training list, cued recall of correct nouns and the fill-in test of correct nouns showed better performance in the constant-exemplar condition than in the variable-exemplar condition. It appears that the variable-exemplar condition generated more interference. In Experiment 4, with a 24-hour delay between the training and test lists, variability had no effect on correct recall. This ~~was the result~~ would arise if the delay ~~of~~ produced greater discriminability between the training list and the test list.

Since variability had so little positive effect on recall performance, it may be that only one exemplar is needed in the training list to create a memory schema. Subjects may be able to generalize from each presented noun to its more general category with only one presentation of a sentence-frame exemplar. For example, presenting the sentence "The rabbi cast an elephant from steel" one or more times may form a schema that could be described as "The clergyman created an animal out of some metal." Whatever the explanation, it was surprising to find that variability of exemplars had seemingly so little effect on recall performance.

Most of the results of Experiments 3 and 4 correspond to those of the first two experiments. Imagery ratings tended to increase with training repetitions. Correct recall tended to increase with repetitions in the free-recall and cued-recall tests. As discussed above, in the fill-in test of Experiment 3 correct recall decreased with repetitions as in the earlier experiments, but ~~a~~ an increase with repetitions took place in Experiment 4. Not surprisingly, recall of correct category words increased with training repetitions.

Spacing of training exemplars had little effect in these experiments. The few times ~~effects of~~ spacing effects appeared, it ~~appeared~~ seemed to ~~occur~~ arise by chance. It appears that what is needed in future experiments are much larger spacing intervals that those used here.

A Model for Sentence Schemata

Although the data of Experiments 1 to 4 provide evidence for the opposite-acting schema-based mediation and interference processes proposed by Thorndyke and Hayes-Roth (1979), no inverted U-shaped curves were found of the kind they reported. Also, recall of the test sentences seemed to be minimally affected by a 24-hour delay between the training and test lists. Considering the differences in the types of materials used here and used by Thorndyke and Hayes-Roth, the results may not be surprising. Perhaps some different range of factor parameters would replicate the recall patterns reported by Thorndyke and Hayes-Roth.

Another reason why schema mediation and interference processes may be difficult to ~~study~~ disentangle is that their effects may occur concurrently, rather than with interference being delayed until after schema-formation as Thorndyke and Hayes-Roth supposed. To better understand the data of Experiments 3 and 4 in terms of mediating and interfering processes, a model was developed to try to fit the data. A graphic representation of this model is shown in Figure 1.

Insert Figure 1 about here.

Only number of repetitions was included as a factor. Spacing and variability manipulations were ignored.

The assumptions of the model are as follows: On each presentation of a sentence, either in the training list or in the test list, there is a probability f that a schema will be formed in memory for that sentence frame. Also, once the schema has been formed there is a probability r that on any presentation the memory schema will become accessible for recall during the free-recall task. Hence, the memory schema can be in one of three states: State 0 in which the schema has not been formed; State F in which the schema has been formed but is not retrievable in a free-recall task; and State R in which the schema has been formed and is retrievable. The transition probabilities for the three states are given below:

$$
\begin{array}{cc} & \text{N+1} \\ & \begin{array}{ccc} \text{State 0} & \text{State F} & \text{State R} \end{array} \\ \text{N} \quad \begin{array}{c} \text{State 0} \\ \text{State F} \\ \text{State R} \end{array} & \begin{bmatrix} 1-f-fr & f & fr \\ 0 & 1-r & r \\ 0 & 0 & 1 \end{bmatrix} \end{array}
$$

The matrix of transition probabilities from Trial 1 to Trial k is given below (see Levine & Burke, 1972, chap 6):

$$
P^{k} = \begin{bmatrix} (1-f-r)^{k} & A & B \\ 0 & (1-r)^{k} & 1-(1-r)^{k} \\ 0 & 0 & 1 \end{bmatrix}
$$

where $A = f[(1-f-fr)^{k} - (1-r)^{k}]/(1-f-fr)-(1-r)$

and $B = 1 - (1-f-fr)^{k} - A$

In order for free recall to occur by retrieving a sentence schema from memory, the schema must be in State R. If the schema is in State R and is retrieved, then it is also necessary to retrieve the correct response associated with each node (slot) of the schema. This is the noun recently presented in the test list. This noun is arbitrarily assigned a strength of 1. The other category nouns presented in the training list interfere with recall of the correct noun and are each assigned a strength c. Each copy in memory of all k-1 nouns presented in the training list is assigned a strength c. So the total stength of the interfering nouns is $(k-1)c$, whether they are the same noun.

~~or not~~. The set of ~~responses resulting in~~ extraneous associations or the editing out of intrusions leading to no overt response is assigned ~~the~~ a net strength of d. Consequently, the total strength of all ~~responses in~~ associations coming out of each of the three nodes of a sentence schema is $1 + (k-1)c + d$, where $k-1$ represents the number of category instances presented in the training list. Hence, the probability of a correct response given that the appropriate schema is available in memory is

$$1/[1+(k-1)c+d].$$

The probability of recalling a category intrusion is

$$(k-1)c/[1+(k-1)c+d].$$

And the probability of not responding to this slot is

$$d/[1+(k-1)c+d].$$

These equations are applied to each of the three nouns associated with each sentence schema. For example, the probability of recalling a test sentence in the free-recall test made up of two correct nouns and a category intrusion is

$$[1 - (1-f-fr)^{k} - A]\, 3[1/[1+(k-1)c+d]]^2 [(k-1)c/[1+(k-1)c+d]],$$

where the number of exemplars presented in the training list was $k-1$.

In the test of cued recall, the first noun from each test-list sentence was presented as a cue, and the subject had to write out the rest of the sentence. For the sentence to be recalled, ~~it is assumed that~~ we assume that the corresponding sentence schema ~~has~~ must have been formed in memory. The schema does not have to be accessible in the sense that it has to be ~~recallable~~ retrievable, as in the free-recall task. Because a cue is provided, the schema can be either in State F or in State R to produce recall. However, the noun presented as a cue has a probability u of retrieving the schema from memory. The subject uses this cue by first inferring to what category it belongs, and then retrieving the schema that includes that category of nouns. For example, the probability of recalling one correct noun and one category intrusion given the first noun of a test sentence as a cue is

$$u[1 - (1-f-fr)^k]\,2[1/[1+(k-1)c+d]][(k-1)c/[1+(k-1)c+d]],$$

where k-1 exemplars of the sentence frame were presented in the training list.

In the fill-in test it is assumed that if the schema is in State F or State R, then the schema is available to the subject when the almost complete sentence is presented. The probability of a correct response is

$$[1 - (1-f-fr)^k][1/[1+(k-1)c+d]],$$

where k-1 exemplars of the sentence frame were presented in the training list.

Fitting the model. To fit the model, data from 16 categories of responding were used from Experiments 3 and 4. These categories and the proportions of recall associated with each one are shown in Table 8.

Insert Table 8 about here.

The following categories of responding were used from the free recall test: all three nouns correct, two nouns correct with one category intrusion, two nouns correct with one omission, one noun correct with two intrusions, one noun correct with one intrusion and one omission, one noun correct with two omissions, three category intrusions, two intrusions with one omission, and one intrusion with two omissions. In cued recall there were only two nouns available for recall, so the outcomes used were two correct nouns, one correct noun with one category intrusion, one correct noun with an omission, two intrusions, and one intrusion with an omission. From the fill-in test two outcomes were used: recall of the correct noun and recall of a category intrusion. The proportion of nouns recalled in each of these 16 response categories was computed for 0, 3, and 8 repetitions of the sentence exemplars in the training list. This resulted in 48 data points in each experiment. These 48 proportions were then estimated from the model after the best-fit parameters were obtained. The observed and predicted proportions are shown in Table 8 for Experiments 3 and 4. Each proportion was based on 96 observations, 4 per subject. The model parameters used were estimated separately for the two experiments.

An expression for each of the 48 data points was derived from the model. Examples of these have already been discussed. A chi-squared minimization procedure (STEPIT, Chandler, 1965) was used to estimate the model parameters that best fit the data from each experiment. Since the chi-squared minimization procedure is based on frequencies, each of the observed proportions in Table 8 were multipled by 96 before a chi-squared statistic was computed. However, many of the observed frequencies were small and close to zero thus not meeting the assumption of the chi-sqaure test which assumes these frequencies to be normally distributed. To deal with this problem, only observed proportions greater than .10 in both experiments were used to compute the chi-squared statistic. With this procedure 12 of the 48 proportions were used in the test of goodness of fit, although all 48 proportions were estimated from the parameters derived from the 12 proportions. These estimated proportions are shown at the bottom of Table 8. The parameter values and the chi-squared values derived from the two experiments are also shown in Table 8. For both experiments the chi-squared statistic obtained was significant, indicating that the frequencies of occurrence generated by the model are significantly different from those observed. Rather than attempting to refine the model in order to obtain a better fit with the data, we suggest that this model does a reasonably good job of fitting the pattern of proportion of responses observed, even if all the values are not precisely matched. Of the 96 observed proportions presented in Table 10, only 10 of the predicted proportions were more than .10 distant from the observed proportions.

Figures 2 and 3 show the observed and expected proportions of correct

----

Insert Figures 2 and 3 about here.

----

responses and category intrusions from the fill-in tests of Experiments 3 and 4. As discussed previously, the patterns of ~~correct~~ instance vs. category recall tend to move in opposite directions as training repetitions increased. In Experiment 3 there is no improvement predicted with increasing repetitions of the sentence frame in the training list from 0 to 3, but in Experiment 4 there is a predicted increase. The primary reason for this difference is that the value of the parameter *c*, which represents the strength in memory of each category intrusion, differs in the two experiments. In Experiment 3 the value of *c* is .10 and in Experiment 4 the value is .05. The formula for the probability of a correct response in the fill-in test with k repetitions of the sentence frame is

$$[1 - (1-f-fr)^{k}]\,[1/[1+(k-1)c+d]].$$

The k repetitions are composed of k-1 repetitions of the sentence frame in the training list and 1 repetition in the test list. As k becomes large in value, the expression $[1 - (1-f-fr)^{k}]$ approaches the value 1. In Experiment 3 the predicted probability of correct recall on the fill-in test starts at the value .47 with k = 1 and ends at the value .50 with k = 9. However, in Experiment 4 the starting value is .43 and the final value is .67. So the increase in recall found in Experiment 4 is predicted by the model. The larger value of parameter *c* in Experiment 4 compared to Experiment 3 may result from the fact that in Experiment 4 the test list was presented and tested 24 hours after the training list. Hence, there was less interference from specific nouns presented in the training list during the learning and recalling of the test list.

It should be noted that the probability of a schema becoming available in memory after a presentation of an exemplar, parameter *f*, is somewhat smaller in Experiment 4 than it is in

Experiment 3. This is also true for the recall parameter *r*. These differences also demonstrate a small amount of forgetting of the schema formed by the training list after 24 hours.

General Discussion

The results of the four experiments reported here lend support for the schema-based mediation and interference processes proposed by Thorndyke and Hayes-Roth (1979). The most important result was that increasing the number of sentence exemplars in the training list had two effects on the recall of the same-frame sentence presented in the test list. One effect was to increase the probability of recall of the test sentence by making more likely the formation of a sentence schema based on the sentence frame. This sentence schema could then be instantiated by the test sentence and make recall of the test sentence more likely. The increased availability of the sentence schema was evidenced by the increased recall of nouns from the categories that were used to make up the sentence frame. This effect was observed in all four experiments in the tests of free and cued recall, even though the repetition effect was sometimes not statistically significant.

The second ~~and~~ result was a negative effect of increased training. ~~was also observed.~~ As an increasing number of examples of a sentence frame were presented in the training list, a sentence schema was more likely to be formed. In addition, however, subjects had an increasingly difficult time recalling the details of the test sentences. A schema formed from a large number of examples was necessarily associated with many different details from the training sentences, and these often intruded during recall of the test sentence. This phenomenon can be seen most clearly in the results of the fill-in tests of Experiments 1 and 2. The proportion of correct nouns recalled decreased as the number of repetitions of the sentence frame in the training list increased. However, recall of nouns belonging to the categories used in the sentence frame increased. This result showed that the details of the most recent instance of ~~sentence representing~~ a common-action ~~sentence~~ frame become more prone to interference ~~less available~~ as previous training is increased, even though the schema itself was more likely to have been formed and be available in memory.

Unlike the Thorndyke and Hayes-Roth (1979) experiments, the materials and procedure used in the present experiments allowed the manipulation of both the spacing of sentence exemplars in the training list as well as ~~and~~ their variability. However, spacing of sentence exemplars during ~~in the~~ training ~~list~~ had little effect on correct recall of sentences from the test list. Variability of exemplars may have ~~had a~~ facilitated ~~facilitating effect on~~ the formation of memory schemas to a small degree. The primary effect of the variable-exemplar condition was to increase the recall of category nouns (compared to the constant-exemplar condition). In Experiments 2 and 4, this increase was not accompanied by a decrease in correct recall; ~~but~~ in Experiment 3 a decrease in correct recall did occur. So the evidence linking variability of exemplars to rate of schema formation is mixed.

Spacing and variability could be important factors in future experiments. The sentences from the same sentence frame used in these experiments were so similar and so easily related to one another that it is possible that the spacing values used were too small to have a large effect on recall. If larger spacing values are used, some effects on recall may be observed.

A delay of 24 hours between training and presentation of the test sentences did not affect overall correct recall but did seem to change the relation between correct recall and the number of training repetitions. With immediate presentation of the test list, there was a tendency for correct recall on the fill-in task to increase with repetitions, whereas with a delayed presentation there was a tendency for correct recall to increase. In general, however, the effects of delay were small.

Little evidence was found of the U-shaped curve connecting correct recall with the number of training sentences. As mentioned above with regard to spacing and variability manipulations, the effects of delay and repetitions would become more apparent if the experimental parameters are given a greater range of values.

It was consistently observed that that mean imagery ratings increased as addition numbers of exemplars from the same sentence frame were presented on the training list. This occurred even when the exemplars were different. When an exemplar was presented it made contact in memory either with the schema being formed for the sentence frame or with previously presented exemplars from the same sentence frame. In either case the different instances of a sentence frame were not stored independently of each other. The possibility of independent storage has been proposed by Medin and Schaffer (1978). If the previous presentation of a ~~sentence~~ instance from a sentence frame affects the rating of a new sentence from that frame, then it must be assumed that the storage of the new sentence will also be affected. Therefore, independen~~t~~ce should not occur.

__The Model__

It may be premature to attempt to develop a mathematical model to fit the data collected using the experimental paradigm introduced here. In fact, the model presented produced performance means that were significantly different from the performance means observed in the experiments, although the pattern of results was duplicated fairly well. The model takes into account the effects of repetition in the training list, but not the effects of spacing, variability, or other factors that future research may show to be important. However, it may be valuable to compare more closely the model presented here with the learning process proposed by Thorndyke and Hayes-Roth (1979). They are similar in many respects. Two mechanisms are included. One is schema formation and the other is item interference.

An important difference, however, between the Thorndyke and Hayes-Roth proposal and the model presented here is in the use of context tags. In the model the most recent detail is most strongly associated with the category concept in the schema. It has a strength of 1 compared to a strength _c_ for each potential intrusion and a strength _d_ for no overt response. The model also implies that the probability of obtaining the correct noun from one category of a sentence schema is independent of the intrusions. Table 2 shows that the sentences recalled are made ~~up either of all correct nouns or of~~ a mixture of correct nouns and category intrusions. The proportion of recalled sentences made up entirely of intrusions is small. The question remains, however, whether recall of the correct noun is completely independent from one category to another within the same sentence schema.

To determine this, an additional analysis was done of the free and cued recall of Experiments 3 and 4. The total number of correct nouns, category intrusions, and omission errors were counted from the sentences recalled by all subjects. Because the proportion of correct responses was a function of the number of repetitions, totals were computed separately based on the number of repetitions of the sentence frame in the training list. From these base frequencies the expected frequency for each category of response was computed, where it was assumed that the correct recall from one category in a sentence schema is independent of the correct recall from another category. After this ~~was done~~, a chi-squared test was performed to determine if the observed and expected frequencies of the various response categories were different. Table 9 shows the results of this analysis for free recall, and Table 10 shows the results for cued recall. The analysis was done only in those cases where each of the response categories used had expected frequencies of 5 or greater.

Insert Table 9 and Table 10 about here.

Because of small frequencies, the free recall of Experiment 4 was not used. Nor was the cued recall of Experiment 4 used involving three repetitions in the training list. The number of parameters estimated was three, so the degrees of freedom associated with each chi-squared test was the number of response categories minus three. In each analysis these three parameters represented the overall proportions of correct responses, intrusions, and omissions. Combining the chi-squared values for the free-recall data in Table 9, the result is χ^2 (3) = 1.66, which is not significant. However, combining the chi-squared values for cued recall in Table 10, χ^2 (4) = 27.83, $p < .001$. It appears that in cued recall the nouns recalled for each sentence were not independently retrieved in cued recall. Inspection of Table 10 shows that typically more correct-correct (CC) and intrusion-intrusion (II) noun combinations were recalled in cued recall than was expected by chance, along with fewer correct-intrusion (CI) combinations. This tendency does not exist in the free-recall data.

These analyses indicate that under cued-recall tests a noun retrieved from the test list is likely to be accompanied by another correct noun from the same sentence. This means that there must be some additional mechanism in our proposed model for linking or associating nouns to one another as well as to the sentence schema. This mechanism could involve the context tags proposed by Thorndyke and Hayes-Roth (1979) and Anderson and Bower (1972a), or it could also involve a temporal search mechanism as proposed by Bellezza (1982). The nouns from a given probability of recalling the correct noun from another category. In contrast, Thorndyke and Hayes-Roth proposed that the details of a particular sentence instantiating a schema are associated not only to the schema itself but also to a context tag. In this manner various instances of the schema can be discriminated in memory (see Figure 2 in Thorndyke and Hayes-Roth, 1979). As the number of competing context tags increases, the learner looses the ability to discriminate among the various context tags, and as a result, recall performance decreases. It is not clear from the discussion of Thorndyke and Hayes-Roth whether context-tag confusion is the only mechanism causing recall deficits. If it is, then a common recall error should be to recall an intact sentence fitting a schema that is the wrong sentence. Context confusion indicates that if one noun in a sentence is correct, then the other nouns should tend to be correct. Similarly, if one noun is wrong, then the other nouns should also tend to be wrong. In the Thorndyke and Hayes-Roth formulation, it is not clear whether a sentence can be recalled that is a mixture of correct nouns and category intrusions. This may happen in the Thorndyke and Hayes-Roth process if context tags are forgotten for some of the nouns, and the subject guesses by choosing some other noun recently associated with the schema which also has no context tag. But this seems to be an exceptional situation in their account of how schema instances are discriminated in memory.

It is clear from our results that subjects do not recall sentences that are composed of either all correct nouns or all sentence could be indirectly connected by being linked to the same context tag. Under free recall, however, this dependence does not appear. So the model is correct in assuming independence among the nouns free recalled. Stated another way, in free recall there is no evidence that the correct recall of one correct noun from a sentence was linked to the recall of another (Anderson & Bower, 1972b; Foss & Harwood, 1975).

The model assumes indepdendence among the responses. Future development of this model may have to include more sophisticated context checking mechanisms. But before these mechanisms can be added and verified, more data involving free and cued recall will have to collected.

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93*, 203-231.

Anderson, J. R., & Bower, G. H. (1972a). Recognition and retrieval processes in free recall. *Psychological Review, 79*, 97-123.

Anderson, J. R., & Bower, G. H. (1972b). Configural properties in sentence memory. *Journal of Verbal Learning and Verbal Behavior, 11*, 594-605.

Anderson, J. R., Kline, P. J., Beasley, C. M. (1979). A general learning theory and its application to schema abstraction. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 13, pp. 277-318). New York: Academic Press.

Anderson, J. R., Kline, P. J., & Beasley, C. M. (1980). Complex learning processes. In R. E. Snow, P. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction. Vol. 2: Cognitive processes of learning and problem solving* (pp. 199-235). Hillsdale, N. J.: Erlbaum.

Anderson, R. C. (1978). Schema-directed processes in language comprehension. In A. Lesgold, J. Pelligreno, S. Fokkema, & R. Glaser (Eds.), *Cognitive psychology and instruction* (pp. 67-82). New York: Plenum.

Anderson, R. C., Spiro, R. J., & Anderson, M. C. (1978). Schemata as scaffolding for the representation of information in connected discourse. *American Educational Research Journal, 15*, 433-440.

Ausubel, D. P. (1963). The psychology of meaningful verbal learning. New York: Grune and Stratton.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut norms, Journal of Experimental Psychology Monograph, 80, Pt. 2, 1-46.

Bellezza, F. S. (1982). Updating memory using mnemonic devices. Cognitive Psychology, 14, 301-327.

Bobrow, D. G., & Norman, D. A. (1975). Some principles of memory schemata. In D. G. Bobrow & D. A. Norman (Eds.), Representation and understanding (pp. 131-149). New York: Academic Press.

Bower, G. H. (1974). Selective facilitation and interference in retention of prose. Journal of Educational Psychology, 66, 1-8.

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

Bransford, J. D., & Johnson, M. K. (1972). Contextual prequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 11, 717-726.

Chandler, J. P. (1965). Subroutine STEPIT. Program QCPE 66. Quantum Chemistry Program Exchange, Indiana University.

Cicchetti, D. V. (1972). Extension of multiple-range tests to interaction tables in the analysis of variance: A rapid approximate solution. Psychological Bulletin, 77, 405-408.

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 397-417.

Foss, D. J., & Harwood, D. A. (1975). Memory for sentences: Implications for human associative memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 1-16.

Franks, J. J., & Bransford, J. D. (1971). Abstraction of visual patterns. *Journal of Experimental Psychology*, *90*, 65-74.

Graesser, A. C., & Nakamura, G. V. (1982). The impact of a schema on comprehension and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 60-109). New York: Academic Press.

Hayes-Roth, B., & Thorndyke, P. W. (1979). Integration of knowledge from text. *Journal of Verbal Learning and Verbal Behavior*, *18*, 91-108.

Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences*. Belmont, Calif.: Wadsworth.

Levine, G., & Burke, C. J. (1972). *Mathematical model techniques for learning theories*. New York: Academic Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification data. *Psychological Review*, *85*, 207-238.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York: McGraw-Hill.

Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. *Journal of Psychological Review*, 56, 132-143.

Posner, M. J., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

Posner, M. J., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83*, 304-308.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3*, 382-407.

Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquistion of knowledge* (pp. 99-135). Hillsdale, N. J.: Erlbaum and Associates.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, N. J.: Erlbaum and Associates.

Smith, D. A., & Graesser, A. C. (1981). Memory for actions in scripted activities as a function of typicality, retention interval, and retrieval task. *Memory & Cognition, 9*, 550-559.

Taylor, S. E., & Crocker, J. (1981). Schematic bases of social information processing. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium* (Vol. 1, pp. 89-134). Hillsdale, N. J.: Erlbaum and Associates.

Thorndyke, P. W., & Hayes-Roth, B. (1979). The use of schemata in the acquisition and transfer of knowledge. *Cognitive Psychology, 11*, 82-106.

Footnotes

This research was supported in part by a grant to the first author from the Field-Wiltsie Foundation. The authors would like to thank Greg A. Camillo, Debbie A. Hartstein, Barbara A. Lombardo, Cynthia S. Marshall, and Kellie A Vaughan for their help in collecting and scoring the data. Thanks also goes to Ohio University Computing and Learning Services for making computer time and their facilities available. Requests for reprints should be sent to Francis S. Bellezza, Department of Psychology, Ohio University, Athens, Ohio 45701.

Table 1

Examples of the Sentence Frames and Sentences Used in the Experiments

Categories: optical instrument, U. S. state, and foreign city

The microscope was moved from Ohio to Berlin.

The telescope was transported from Wisconsin to London.

The binoculars were sent from Maryland to Paris.

The spyglass was shipped from Idaho to Tokyo.

The glasses were mailed from Oregon to Rome.

Categories: female first name, fossil fuel, and type of soil

Mary prospected for oil in the mud.

Sue hunted for gas in the dirt.

Ann looked for coal in the clay.

Jane searched for gasoline in the slime.

Carol sought kerosene in the dust.

Table 2

Proportions of Sentence Types Recalled in the Four Experiments Using Free Recall and Cued Recall

Experiment	Fragments	All correct	All intrusions	Mixture
		Free recall		
1	.03	.08	.02	.09
3	.06	.13	.01	.10
4	.04	.17	.00	.07
		Cued recall		
2	.10	.19	.03	.13
3	.11	.24	.07	.13
4	.06	.32	.03	.09

Table 3

Frequency of Recall of Correct Nouns, Category Intrusions, and Other Nouns in the Four Experiments

Experiment	Correct	Intrusions	Other	Total
		Free recall		
1	182	82	11	275
3	187	46	3	236
4	208	31	1	240
		Cued recall		
2	575	105	21	701
3	352	81	3	436
4	375	47	4	426

Table 4

Mean Ratings of Comprehensibility or Imagery as a Function of the Number of Repetitions of the Sentence Frame in the Training List.

Condition	Number of repetitions							
	1	2	3	4	5	6	7	8
				Experiment 1				
	3.55	3.65	3.77	3.64	3.91	3.95	3.98	3.92
				Experiment 2				
Constant	3.27	3.45	3.69	3.89	4.02	4.14	4.22	4.27
Variable	3.30	3.44	3.59	3.57	3.73	3.71	3.81	3.81

Note. In Experiment 1 ratings were of comprehensibility and in Experiment 2 ratings were of imagery; 5 is high and 1 is low.

Table 5

Proportion of Nouns Correctly Recalled and Categories Correctly Recalled in Experiments 1 and 2 as a Function of the Number of Exemplars of Sentence Frames Presented in the Training List

Condition	Number of repetitions		
	0	3	8
	Experiment 1		
Correct instance recall			
Free recall	.10	.13	.19
Fill-in	.67	.50	.44
Category recall			
Free recall	.11	.19	.31
Fill-in	.75	.81	.94
	Experiment 2		
Correct recall			
Cued recall	.25	.28	.31
Fill-in	.47	.45	.34
Category recall			
Cued recall	.28	.38	.47
Fill-in	.55	.74	.76

Table 6

Mean Ratings of Imagery as a Function of the Number of Repetitions of the Sentence Frame in the Training Lists of Experiments 3 and 4.

Condition	Number of repetitions							
	1	2	3	4	5	6	7	8
	Experiment 3							
Constant	3.54	3.77	3.97	3.88	3.98	3.98	4.06	4.10
Variable	3.27	3.22	3.23	3.31	3.33	3.35	3.02	3.35
	Experiment 4							
Constant	3.79	3.90	4.12	4.18	4.22	4.20	4.28	4.28
Variable	3.74	3.68	3.77	3.96	4.10	4.14	4.10	3.98

Table 7

Proportion of Nouns Correctly Recalled and Categories Correctly Recalled in Experiments 3 and 4 as a Function of the Number of Exemplars of Each Sentence Frame in the Training List

Condition	0	3	8
	Experiment 3		
Correct recall			
Free recall	.08	.27	.30
Cued recall	.27	.43	.32
Fill-in	.50	.52	.45
Category recall			
Free recall	.09	.32	.40
Cued recall	.31	.57	.56
Fill-in	.56	.83	.84
	Experiment 4		
Correct recall			
Free recall	.11	.22	.36
Cued recall	.25	.42	.47
Fill-in	.44	.59	.57
Correct category			
Free recall	.12	.24	.43
Cued recall	.27	.49	.63
Fill-in	.55	.81	.86

Table 8

Observed and Expected Proportions of Each Response Category Recalled in Experiments 3 and 4

Response category	Experiment 3 Observed	Experiment 3 Predicted	Experiment 4 Observed	Experiment 4 Predicted
		0 repetitions		
Free recall				
3C	.06	.06	.09	.04
2C, 1I	.00	.00	.02	.00
2C, 1O	.02	.03	.02	.01
1C, 2I	.00	.00	.00	.00
1C, 1I, 1O	.02	.00	.00	.00
1C, 2O	.00	.01	.01	.00
3I	.00	.00	.00	.00
2I, 1O	.00	.00	.00	.00
1I, 2O	.00	.00	.00	.00
Cued recall				
* 2C	.17	.24	.21	.24
1C, 1I	.06	.00	.03	.00
1C, 1O	.15	.10	.06	.05
2I	.01	.00	.00	.00
1I, 1O	.00	.00	.00	.00
Fill-in				
* C	.50	.47	.45	.43
I	.06	.00	.12	.00

Table 8 (cont.)

Response category	Experiment 3 Observed	Experiment 3 Predicted	Experiment 4 Observed	Experiment 4 Predicted
	3 repetitions			
Free recall				
* 3C	.17	.15	.17	.17
2C, 1I	.10	.13	.05	.08
2C, 1O	.04	.09	.02	.05
1C, 2I	.00	.04	.00	.01
1C, 1I, 1O	.01	.05	.03	.01
1C, 2O	.01	.02	.00	.00
3I	.01	.00	.00	.00
2I, 1O	.00	.01	.00	.00
1I, 2O	.00	.01	.00	.00
Cued recall				
* 2C	.32	.26	.37	.37
1C, 1I	.15	.16	.09	.11
1C, 1O	.05	.11	.04	.07
2I	.05	.02	.01	.01
1I, 1O	.02	.03	.01	.01
Fill-in				
* C	.52	.64	.60	.74
* I	.31	.19	.22	.11

Table 8 (cont.)

Response category	Experiment 3 Observed	Experiment 3 Predicted	Experiment 4 Observed	Experiment 4 Predicted
		8 repetitions		
Free recall				
* 3C	.15	.11	.28	.20
* 2C, 1I	.12	.25	.10	.24
2C, 1O	.05	.06	.01	.06
1C, 2I	.04	.20	.03	.10
1C, 1I, 1O	.01	.10	.01	.05
1C, 2O	.03	.01	.00	.00
3I	.02	.05	.00	.01
2I, 1O	.02	.04	.02	.01
1I, 2O	.00	.01	.00	.00
Cued recall				
* 2C	.22	.15	.38	.27
* 1C, 1I	.14	.24	.10	.22
1C, 1O	.06	.06	.04	.05
2I	.15	.10	.07	.06
1I, 1O	.05	.05	.07	.02
Fill-in				
* C	.45	.50	.57	.67
* I	.40	.40	.30	.27

Table 8 (cont.)

Response category	Experiment 3 Observed	Experiment 3 Predicted	Experiment 4 Observed	Experiment 4 Predicted
Parameters				
r	.21		.14	
f	.47		.41	
c	.10		.05	
d	.20		.09	
u	.51		.62	
Chi-square (7)	30.30		34.64	

Note. C = correct noun recalled; I = category intrusion recalled; O = word omitted. If an * precedes the response category label, then the observed proportions were used in the parameter estimation procedure.

Table 9

Observed and Expected Frequencies of Response Categories From the Free-Recall Tests of Experiments 3 and 4.

Response category	Observed	Expected
	Experiment 3, 3 repetitions, $\underline{X}^2$ (1) = .93	
CCC	15	16.15
CCI	10	8.67
CCO	4	4.35
Other	3	4.83
	Experiment 3, 3 repetitions, $\underline{X}^2$ (2) = 1.33	
CCC	15	12.31
CCI	11	13.05
CCO	5	6.06
CII	4	4.62
Other	8	6.96

Table 10

Observed and Expected Frequencies of Response Categories From the Cued-Recall Tests of Experiments 3 and 4.

Response category	Observed	Expected
Experiment 3, 3 repetitions, X^2 (1) = 1.60		
CC	31	29.16
CI	15	19.00
CO	6	5.64
Other	7	5.20
Experiment 3, 8 repetitions, X^2 (2) = 9.48		
CC	21	15.77
CI	13	23.79
CO	6	5.57
II	14	8.97
IO	5	4.80
Experiment 4, 3 repetitions, X^2 (1) = 16.75		
CC	35	28.70
CI	10	20.92
CO	4	6.72
Other	14	6.56

Appendix A

Categories of Words Used to Make up Presented Sentences

List sentences

optical instr.	U. S. state	foreign city
female first name	fossil fuel	type of soil
writing instr.	cutting instr.	Ivy League school
printed material	cloth	kind of precipitation
disease	word type	U. S. region
fruit	room in house	time interval
non-alcoholic bev.	color	season of year
small bird	precious stone	wooden boat
member of clergy	savanna animal	nonprecious metal
tool	team sport	vegetable
synonym for child	flower	body of water
racing sport	religious building	foreign country
brand of auto	insect	American coin
male relative	musical instrument	royal personage
steel ship	gaseous element	seasoning
weapon	dessert	string-like binder
predatory bird	personal ornament	unit of distance
timepiece	type of road	type of music
military rank	type of toy	container
type of crime	railroad car	violent weather
young woman	U. S. city	building stone
occupation	dwelling	geological formation
elected official	commercial business	planet
U. S. president	alcoholic drink	science

Verbal Schemas

fish	small mammal	dance
table utensil	furniture	organized entertainment
male first name	part of body	building part
farm animal	footware	clothing
auto part	precious metal	continent
snake	tree	vehicle

Filler sentences

reptile	digit	circular object
predatory cat	letter	type of school
type of stick	punctuation mark	vowel
soft drink brand	kitchen furniture	foreign money
indian tribe	cartoon character	holiday
type of spirit	air vehicle	type of nut
breed of dog	type of ape	type of berry

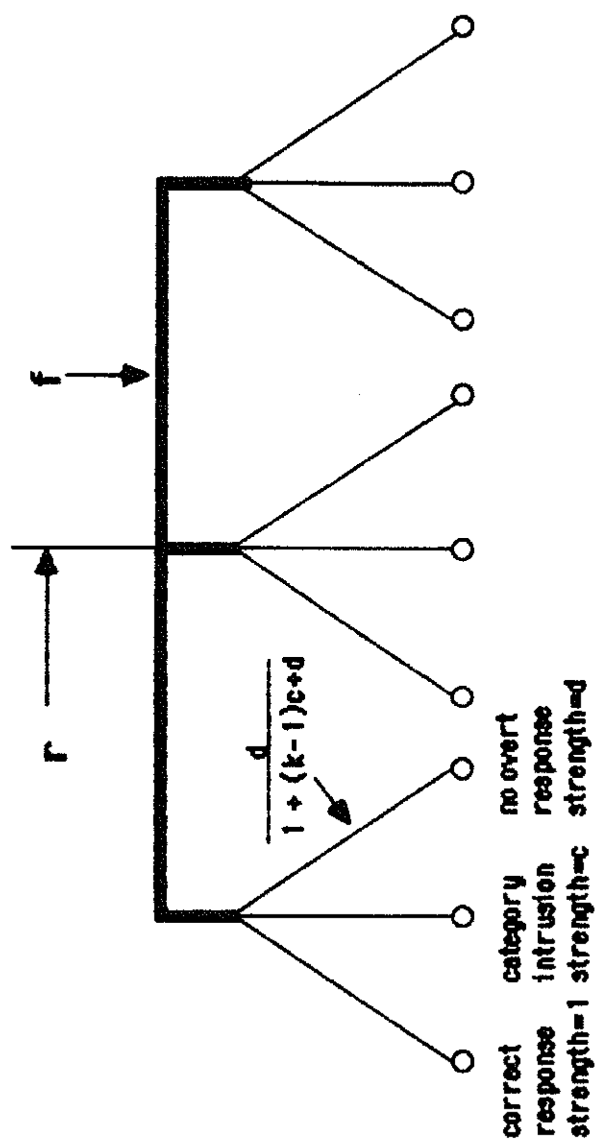
Figure Captions

Figure 1. Diagram of the model proposed for the sentence schemas.

Figure 2. The observed and predicted proportion of correct nouns and category intrusions recalled in the fill-in test of Experiment 3.

Figure 3. The observed and predicted proportion of correct nouns and category intrusions recalled in the fill-in test of Experiment 4.

Figure 1



correct response strength=1
category intrusion strength=c
no overt response strength=d
$\frac{d}{1+(k-1)c+d}$
r
f

Figure 2

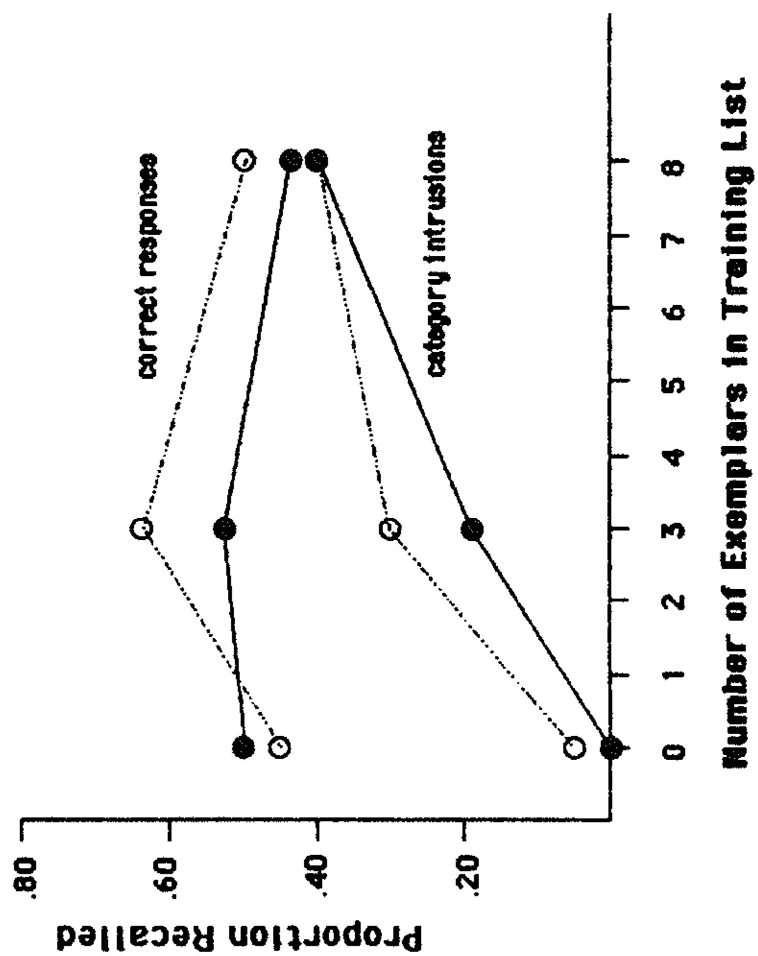
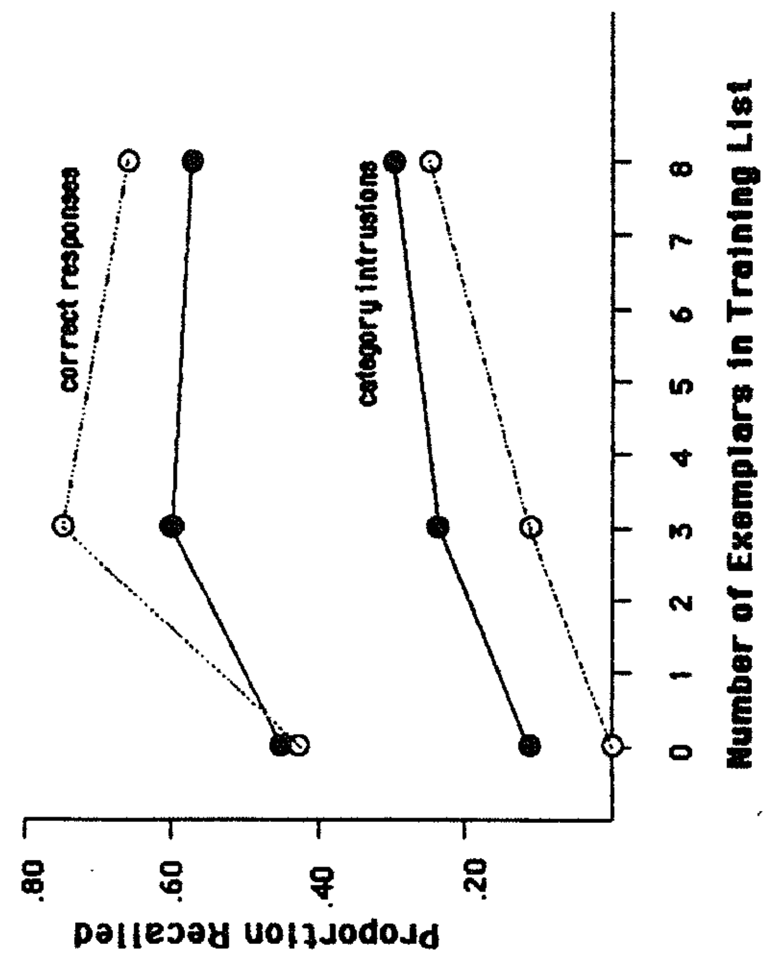

Proportion Recalled
.80
.60
.40
.20
correct responses
category intrusions
0
1
2
3
4
5
6
7
8
Number of Exemplars in Training List

Figure 3


correct responses
category intrusions
Proportion Recalled
.80
.60
.40
.20
0
1
2
3
4
5
6
7
8
Number of Exemplars in Training List