*Journal of Experimental Psychology: Human Learning and Memory*
1975, Vol. 104, No. 3, 249–260

# Item-Specific and Relation-Specific Interference in Sentence Memory

John R. Anderson
*University of Michigan*

In two experiments, sentences like *Nixon kicked the milkman* tested whether memory for such a sentence would be interfered with by learning other sentences about *Nixon*. Experiment 1 examined interfering effects of subsequent sentences on the retention of an original sentence. Experiment 2 looked at interfering effects of competing sentences on recognition reaction time. Both experiments found evidence for item-specific interference. The experiments also questioned whether this interference was relation specific. That is, does learning sentences in which *Nixon* is the object interfere with memory for sentences in which *Nixon* is the subject? As much interference was found with sentences where the relation was changed as with sentences where the relation was the same. Implications of these results for network models of memory are discussed.

Several computer simulation models of memory (Anderson & Bower, 1973; Quillian, 1969; Rumelhart, Lindsay, & Norman, 1972) represent human knowledge in terms of labeled networks of interconnected ideas. An important issue in such models is how these networks are searched to retrieve relevant information. Anderson and Bower (1973) proposed a search algorithm and tested some of its psychological implications. The experiments in this article are further tests of its predictions. The Anderson and Bower model has been dubbed HAM, an acronym for Human Associative Memory.

## *An Overview of HAM*

The fundamental assumption in HAM is that information is represented in memory in terms of networks of propositions about concepts. Propositions in HAM are not to be identified with sentences; they are more abstract entities that are not language specific. Figure 1 illustrates a propositional representation set up by HAM's parser to encode the sentence *Nixon kicked the student*. The words in HAM's memory are connected by W (word) links to nodes that represent the ideas of *Nixon*, *kicking*, and *student*. The general concepts of *kicking* and *student* are connected by $\epsilon$ links to the specific instances referred to by the sentence. *Nixon*, a proper name, already refers to a specific individual. *Nixon* is the subject of the underlying proposition, and the predicate is *kicked the student*. The top node on the tree represents the main proposition; an S branch points to the subject and a P branch to the predicate. The predicate branch is subdivided into an R branch pointing to the relation and an O branch pointing to the object.

*Accessing propositional information.* Suppose a subject is probed with *Nixon*, *kick*, or *student*, or some combination of these. How does he retrieve the original proposition of which these concepts are part? The HAM retrieval model has been specified and tested by Anderson (1974a), Anderson and Bower (1973, chaps. 9 & 12), and Thorndyke and Bower (1974). In these

Experiment 1 was supported by Grant MH-13500 from the National Institute of Mental Health (NIMH) to G. H. Bower. Experiment 2 and preparation of this manuscript were supported by Grants MH-24360 from NIMH and GB-40298 from the National Science Foundation to me. I would like to thank J. Psotka for assisting with Experiment 2.

Inquiries concerning this article should be sent to John Anderson at the Human Performance Center, 330 Packard Road, Ann Arbor, Michigan 48104.

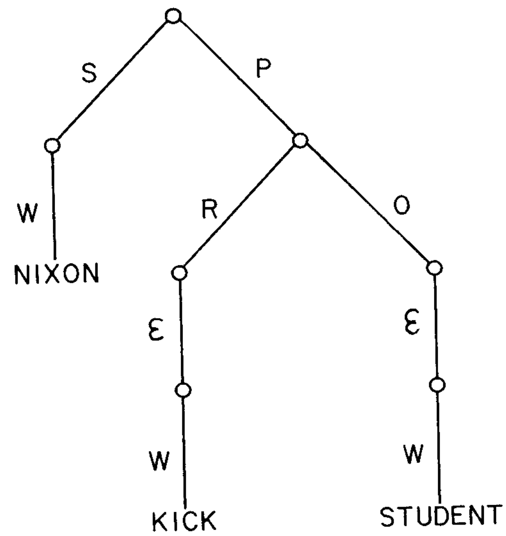


FIGURE 1. HAM's propositional representation for the sentence *Nixon kicked the student.* (S = subject, P = predicate, R = relation, O = object, $\epsilon$ = instance, and W = word.)

experiments, subjects were presented with a probe sentence and were asked to judge whether they had learned the sentence. It is assumed that, upon presentation of a probe sentence, parallel search processes begin from all the concepts in the probe. These processes are independent and each begins a serial search of memory. Search time is determined by whichever process first finds the target proposition or determines that the proposition is not stored. There is, in effect, a race between the serial processes, with reaction time being determined by the fastest.

This search algorithm predicts that the time needed to verify a proposition about a particular concept should increase as more propositions are acquired about that concept. This prediction has been consistently verified. For instance, Anderson (1974a) had subjects commit to memory sentences like *A hippie is in the park* and manipulated the number of sentences involving *hippie* or *park*. Subjects were asked to decide if test probes were among the sentences they had learned. They were approximately 100 msec slower for every additional sentence associated to a concept in a test probe.

Anderson and Bower also relate this search process to interference with retention of sentential materials. They propose that the search for a target proposition is subject to a cutoff time after which searching will cease. Since the number of propositions learned about a concept in a lifetime will be huge, this is a simple way to prevent long and fruitless searches of memory. As a consequence of this assumption, the more new facts a subject learns about a concept, the harder it should be to retrieve an old fact. This provides a mechanism for retroactive interference.

Retroactive interference has typically been displayed in a paired-associate paradigm, where the stimulus may be a digit and the response an adjective. It is interesting to see whether interference effects can be obtained with sentential materials. Some (e.g., Ausubel, 1963; Miller, 1951) have speculated that the concepts of interference would not generalize from nonsense syllables and single words to meaningfully learned linguistic material.

Recently, serious questions have arisen as to whether interference effects are really due to interference in the connection between a specific stimulus and its response (e.g., Postman, Stark, & Fraser, 1968). The argument is that interference results from loss in the availability of a whole class of responses. For instance, in the A–B, A–C paradigm, it is suggested that as a function of A–C learning, the B responses become less available as a class. This is clearly counter to HAM's analysis of the phenomenon. The most convincing evidence that all interference is not generalized response loss comes from a mixed-list design in which some pairs in the two lists bear an A–B, A–C relation and others an A–B, C–D relation (see Anderson & Bower, 1973, Section 15.2, for a review). These experiments tend to find that greater interference can be obtained for those A–B pairs with A–C counterparts than for those without. These interference effects cannot be explained in terms of response-set suppression. The experiments reported in this article use a mixed-list design.

*Relation-specific search.* It seems hard to believe that the search rates uncovered by Anderson (approximately 100 msec per proposition) could generalize from the experimental context to a search of all of long-term memory. The subject must have available to him some heuristics that reduce search times. Anderson and Bower (1973) proposed that it would be possible to reduce the amount of search by using the relational information in a memory structure like Figure 1. For instance, suppose a subject learns two propositions: *Nixon kicked the student* and *The milkman tickled Nixon. Nixon* would be attached to the first proposition by an S link but to the second proposition by an O link. Suppose that when probed with *Nixon kicked the student,* the subject could selectively search the S branches out of the *Nixon* node and disregard the O branches or branches with other labels. Then the subject would be able to reduce the amount of required search. This proposal of Anderson and Bower amounts to the hypothesis that memory is capable of *second-order content addressability.* First-order content addressability refers to the limitation of search to those propositions about the probed concepts. Second-order content addressability adds the claim that the subject can further restrict his search to those propositions that bear the right semantic relation to the probed concept. It is possible to effect second-order content addressing in a computer system. The interesting empirical question is whether it is implemented in the human mind.

HAM's use of relational links to structure memory has some similarities to Gestalt ideas. Gestalt theories proposed that the association of elements depended on merging them into a higher order relation (e.g., Asch, 1969). Similarly, in HAM, ideas cannot be connected by unlabeled links; the links must be labeled with a [illegible]. The Gestalt theory proposed [illegible] memory trace formed from ele[illegible]s characterized by the emergent [illegible] It was further proposed (Koffka, [illegible] traces which were not distinc[illegible] not be segregated from one another. So, similar traces would interfere with one another. Thus, it seems that Gestalt theory would claim, as would HAM, that traces in which the elements bear distinct relations should be segregated in memory and not interfere with one another. To my knowledge, the question of whether distinct relations reduce interference has not been addressed in the Gestalt literature, but it seems that Gestalt theory should predict that interference is relation specific. Thus, the prediction of relation-specific search may not be unique to HAM, but rather may be a claim of a much larger class of memory models.

There is evidence that subjects do not search their memory in a relation-specific manner. I have performed a series of four experiments on this issue, the third and fourth of which are reported here. In the first experiment, subjects studied active or passive sentences in a first list such as *The student was kicked by the lawyer.* Then in a second list they studied further sentences about *student* in which it mantained the same semantic role (i.e., the logical object)—*The student was tickled by the fireman*—or switched logical role—*The fireman was tickled by the student.* Interest focused on a final retention test in which the subject was tested for memory of the original sentence by responding to a probe of *student.* Subjects in both conditions were worse in a recall (77%) than in a control condition (91%) in which further interpolated information had not been learned about *student.* However, there was no difference between the two interpolated conditions. This would seem to contradict the prediction that interference should occur only when the interfering information bears the same semantic relation to the probe as the original information.

One might criticize this experiment because it used class names like *student.* Note in Figure 1 that an $\epsilon$ link intervenes between the class concept *student* and the O link, which connects it to the underlying proposition. This is because the proposition is about a particular student and not the general class. Suppose subjects in the experiment had created different individuals for the original and interpolated sentences. Then there would have been two ϵ links leaving *student*, independent of whether *student* had the same logical role in both sentences. Then there would be interference of the ϵ links leaving the class node.

To deal with this possibility, the second and subsequent experiments have used proper names as the to-be-interfered-upon elements. Note in Figure 1 that the idea node for *Nixon* is directly connected to the proposition by an S link. With proper names there is not the complication of an intervening ϵ link. The second experiment was reported in Anderson and Bower (1973, Section 15.4). Some changes were introduced over the first experiment in an attempt to produce larger interfering effects. Subjects with interpolated material retained 53% of the original material, whereas subjects without interpolation retained 84%. So, very large item-specific interference was obtained. However, the interfering effects were as large when the interpolated relation was different as when it was the same.

The third experiment, to be reported here, was very similar in design to the second. It dealt with a difference between the propositional network representation of HAM and of the ELINOR system of Rumelhart et al. (1972). In HAM there is a small, fixed set of relations that one node in memory may bear to another. In contrast, ELINOR has many relations. In particular, verbs can serve as distinct relations to connect subject and object. In the sentences *Nixon kicked the student* and *Nixon touched the painter*, ELINOR could connect *Nixon* to *student* by a path labeled *kick* and to *painter* by a path labeled *touch*. If these memory structures were searched in a relation-specific manner, there would be no interference. However, the ELINOR representation, even with a relation-specific search, would expect interference if the second sentence were *Nixon kicked the painter*. In that case *Nixon* would be connected to *student* and *painter* by the same *kick* relation. Both experiments reported here looked for relation-specific interference in this one situation where the ELINOR representation would predict it; HAM predicts interference to be relation specific in this situation as well as in others. Note that the ELINOR model, with relation-specific search, would not have predicted any interference in the previous two experiments. So the experimental question is not whether interference is only obtained when predicted by the ELINOR model with relation-specific search, but whether *greater* interference is obtained in this situation. Greater interference would count as evidence for some relation-specific search.

*Meaningful learning instructions.* The first experiment reported here looked for evidence of relation-specific search in an interference design and the second, in a reaction time design. Both experiments used similar learning instructions. Subjects were encouraged to treat the sentences as meaningful entities and not as strings of unrelated words. Subjects reported imaging the sentence's referent, generating semantic elaborations of the sentences, and trying to think of circumstances in which the sentence might be true. Thus, it appeared that instructions did induce subjects to set up a representation in memory that reflected the sentence's meaning. This was critical because I wanted to determine if interference could be obtained with meaningfully learned sentences and whether search processes could use the semantic relations in these sentences.

## EXPERIMENT 1

### *Method*

Subjects studied an original list of 20 sentences for five trials. Earlier experiments had suggested that five original list trials would get most subjects to over 90% recall, but would not produce very much overlearning in the better subjects. Forgetting is expected to be maximal for material just barely learned. Following the original learning trials, subjects had two trials on each of three interpolated learning lists of 20 sentences each. It has been shown that more interference is produced with a number of distinct interfering lists than with one list even though the number of interpolated learning trials is constant (e.g., Postman, 1965; Underwood, 1945). Strong interference is optimal for obtaining a clear-cut decision on the question of relation-specific search.

Both in original and interpolated learning, half the sentences were active and half passive. Also, half of the proper names served as grammatical subject and half as grammatical object. These two factors were combined factorially to provide the four following kinds of sentences:

> Nixon kicked the milkman.
> The milkman kicked Nixon.
> Nixon was kicked by the milkman.
> The milkman was kicked by Nixon.

There were five sentences of each kind in each list.

The sentences in the original list could have any of five relations to the sentences in interpolated learning:

*Control.* The particular political figure was not reused in interpolated learning. Instead, some statement was substituted about a political figure not used in the original list. This corresponds to the A-B, C-D condition in the paired-associate paradigm. The other four conditions all correspond to various forms of A-B, A-C.

*+L, +G.* The political figure was used in a new proposition but in the same *logical* (L) and *grammatical* (G) role as before. So if *Nixon* had appeared in the original list as the grammatical object in a passive sentence, it was a grammatical object in a passive interpolated learning sentence.

*+L, −G.* The political figure was used in the same logical role but in a new grammatical role. This required a switch in the voice of a sentence. *Nixon* might occur as the surface object in a passive original sentence, but as the surface subject in an active interpolated sentence. Thus, *Nixon* would remain the logical subject of the underlying proposition, but would switch surface roles.

*−L, +G.* The surface role of the name was maintained while its logical role was changed. Again a change in the voice of the proposition was required. *Nixon* might appear as the subject of an active sentence in the original list, but the subject of a passive sentence in interpolated learning.

*−L, −G.* Both surface and logical roles were changed. No change of voice was required. Thus, *Nixon* might appear as the object of an active sentence in the original list, but the subject of an active sentence in interpolated learning.

A model that includes relation-specific search predicts interference relative to the control condition in the +L, +G condition. The predictions for the +L, −G and −L, +G conditions depend on the propositional representation for the sentences. If sentences are represented always in terms of their "logical" active form (there is evidence that they are; cf. Anderson, 1974b), then there should be interference in the +L, −G condition but not in the −L, +G condition. Just the opposite prediction obtains if the propositional presentation maintains the same subject and object relation found in the surface representation. In any case, there should be no interference in the −L, −G condition.

Twenty conditions were created by combining factorially the four types of original sentences with the five possible relations to interpolated sentences. The conditions were each represented by one of the sentences each subject had to study. An original sentence had the same relation to an interpolated sentence in each of the three interpolated lists.

Thus the experiment required four lists of 20 sentences each. Since the political figures were repeated across lists in the experimental conditions, only 32 proper names were required. Eighty distinct class names, all referring to people, were used to fill the other noun role in the sentence. Since the same verbs were used in all lists, only 20 verbs were required. However, a different set of 20 verbs was randomly selected for each subject from a master set of 80. Sentences were randomly generated for each subject by computer. This involved selecting for each subject words from lists of 32 proper names, 80 class names, and 80 verbs to realize each of the 20 conditions. The nouns and verbs were chosen such that all possible subject–verb–object combinations generated meaningful sentences. The verbs were chosen so that their logical subject would be assigned to the agentive role in a Fillmore (1971) case grammar and their logical object to the patient role.

The sentences were printed on computer cards, and each subject studied his or her own deck of cards. Within a trial, subjects first studied the sentences at a 12-sec rate and were tested at the same rate. This study–test sequencing was repeated throughout the experiment. The sentences were printed one sentence to a card for both study and test. The test probes consisted of presenting the proper name embedded in a sentence frame with subject and verb missing. Thus a subject might see:

THE ______ WAS ______ BY NIXON.

This kind of probe identifies the relation between *Nixon* and the studied proposition and permits the subjects to take advantage of this relational information in their memory search. The subjects were instructed to write in the missing noun and verb. The study order of the sentences within a trial was randomized, but sentences were tested in the same order that they had been studied so that sentences would be tested at a constant lag. Instructions urged the subjects to treat these sentences "as meaningful sentences about famous politicians." That subjects did treat the sentences meaningfully was manifest from their amusement over the various actions ascribed to the political figures.

After interpolated learning, subjects received one retention test for the 20 original sentences. Subjects were not told in advance about the final retention test. The manner of testing was identical to that used in the original list and interpolated learning. Subjects were instructed only to recall sentences from the first list.

*Subjects.* Thirty-seven subjects were recruited from the general Stanford community by means of an advertisement offering $3.50 for their services. These were run in groups ranging in size from 1 to 7. Subjects who had not achieved a criterion of at least 75% recall of the class nouns by the first original list trial were rejected. By this criterion six subjects were rejected.

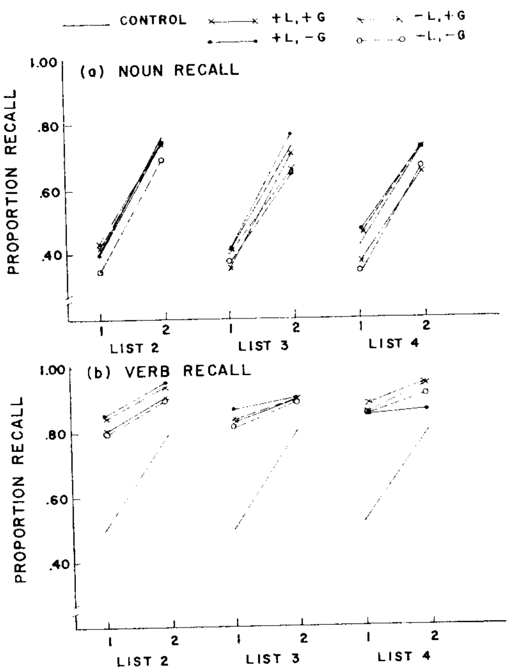


FIGURE 2. Proportion recalled on each of the three interpolated learning tests. (Plotted is performance on Trials 1 and 2 of Lists 2, 3, and 4. The L and G stand for the logical and grammatical roles of the political figures; + and − indicate whether the roles were the same or different in the original list and interpolated learning.)

*Results*

In scoring, a gist criterion was adopted for noun and verb recall. So, for instance, recall of *postman* when *mailman* was required was scored as correct. Such synonomous recall accounted for only 2% of total recall.

*Original learning.* By the fifth trial subjects were recalling an average of 94% of the nouns and subjects correctly. There were no significant differences among the conditions.

*Transfer data.* Figure 2 presents recall on each of the two trials for the three interpolated lists of Experiment 1. Subjects are significantly better at recalling the verbs in the experimental conditions than in the control condition. This is to be expected, since in the experimental conditions the proper names were being paired with the same verbs in all four lists. On the other hand, the class noun was changed for each list. There were no significant differences in recall of the class nouns among the conditions.

*Retention.* The principal concern of the experiment is the final retest for the original sentences. These retention data are displayed in Table 1, classified according to noun or verb recall and the experimental condition. The standard error of the means reported in Table 1 is .035. Verb recall was higher in the experimental conditions than in the control condition. This is to be expected, since with respect to recall of the verb to the proper name, the interpolated trials actually provided extra study. The critical retention data concern the noun recall. There were significant differences between the control and experimental conditions, $F(1, 120) = 6.14$, $p < .025$, but the variation among experimental conditions was not significant, $F(3, 120) = 1.83$. In particular, there is no evidence for relation-specific interference.

The difference between noun recall in the control versus exprimental conditions was .095. The experiment reported by Anderson and Bower (1973) found a much larger difference of .297. Anderson and Bower used different verbs in the interpo-

TABLE 1

MEAN PROPORTION RECALLED IN EXPERIMENTAL LIST 1 RETEST

| Grammatical role (G) | Logical role (L) +L | −L | *M* |
|---|---|---|---|
| **Nouns** | | | |
| +G | .734 | .750 | .742 |
| −G | .839 | .782 | .811 |
| *M* | .787 | .766 | .776 |
| **Verbs** | | | |
| +G | .911 | .927 | .919 |
| −G | .976 | .903 | .940 |
| *M* | .944 | .915 | .929 |

*Note.* Control values: nouns = .871; verbs = .887. Plus and minus signs indicate that the grammatical and logical roles are the same (+) or different (−) in the original and interpolated sentences.

ss noun was changed
were no significant
of the class nouns

ıcipal concern of the
retest for the original
ention data are dis-
assified according to
ınd the experimental
ndard error of the
ıble 1 is .035. Verb
e experimental condi-
rol condition. This is
with respect to recall
per name, the interpo-
provided extra study.
on data concern the
were significant dif-
e control and experi-
F(1, 120) = 6.14, p <
on among experimental
gnificant, F(3, 120) =
, there is no evidence
nterference.
etween noun recall in
exprimental conditions
periment reported by
r (1973) found a much
.297. Anderson and
t verbs in the interpo-

BLE 1
TION RECALLED IN
AL LIST 1 RETEST

| | Logical role (L) | |
|---|---|---|
| | −L | M |
| Nouns | | |
| 4 | .750 | .742 |
| 9 | .782 | .811 |
| 7 | .766 | .776 |
| Verbs | | |
| 1 | .927 | .919 |
| 6 | .903 | .940 |
| 4 | .915 | .929 |

nouns = .871; verbs = .887. Pl
that the grammatical and logic
r different (−) in the original an



lated lists. Otherwise there was no difference between the two experiments. The reduced retroactive interference in this experiment can be explained in terms of the verb repetition. Since verbs were so highly available in the experimental conditions, the subjects effectively had a second cue for noun recall. It has been shown (Anderson & Bower, 1971, 1972) that recall to a noun–verb cue is higher than to a noun cue. This is because the subject has a second retrieval route to the object noun. The effect of this second retrieval route was to lower the interference effects in this experiment.

### *Conclusions from Experiment 1*

Item-specific interference has now been obtained in three mixed-list interference designs. This has important consequences for the questions of whether retroactive interference is due to response set suppression (Postman & Stark, 1969) and whether it can be obtained with meaningfully learned sentential material. However, this interference was not specific to the semantic relation the stimulus bore in the to-be-retrieved proposition. If we assume a labeled network memory representation, then it appears that subjects do not take advantage of the relational information in the labels when searching memory. Some have suggested that this failure is due to the use of well-known proper names. Therefore, it is important to point to the first pilot experiment in the series. This experiment, using class names, also failed to find relation-specific interference. Subsequent experiments have used proper names because of the theoretical difficulties, discussed earlier, in interpreting such results with class names.

## EXPERIMENT 2

The preceding experiment seems to indicate that search of memory is not relation specific. However, there are two objections that could be advanced against this conclusion: First, it depends on the theoretical validity of the analysis of interference in terms of memory search. If this analysis is incorrect, the experiments are inconclusive. Second, it was not necessary for the subjects to treat the proper names as subjects and objects for successful recall. They had only to remember the triplet composed of proper name, verb, and class noun. Given subjects' reports, it seems implausible that they were not treating the proper names as occupying their assigned semantic roles. However, it would be nice to assure experimentally that subjects had committed to memory the logical role of the proper names. Experiment 2 deals with both objections.

Experiment 2 used a reaction time methodology in which subjects were asked to judge whether a probe proposition was among the set committed to memory. Here the interest is in the speed with which these judgments can be made. This allows a more direct inference about the search of memory.

### *Method*

*Design.* As in Experiment 1, political figures' names were the items chosen for experimental manipulation. A particular proper name appeared in 0, 1, or 2 propositions as subject and similarly in 0, 1, or 2 propositions as object. These two dimensions were combined factorially to yield nine combinations. The question of interest was whether the time taken to make a judgment about a probe proposition in which the name appears in one relation would be affected by number of propositions the subject has learned about the name in the other relation. Particular probe propositions in the reaction time test are classified according to the frequency of the name as subject and as object in study propositions. Orthogonal to these dimensions, probe propositions can be classified according to whether they use the proper name as subject or object and whether the probe is a studied proposition or not. Studied propositions were designated as *true*; those not studied, as *false*. These four variables combine factorially to yield 36 apparent conditions (3 numbers of subject propositions × 3 numbers of object propositions × 2 positions in probe × 2 truth values). Certain true conditions are not realizable, however. For instance, one cannot have true probes where the name is subject but the name has occurred in 0 study propositions as subject. There are, in fact, 12 realizable true conditions and 18 false conditions.

*Procedures.* The experiment was divided into two principal phases: a learning phase in which 36 propositions were committed to memory and a reaction time phase in which the subjects judged whether probe propositions were among the to-be-

TABLE 2

AN EXAMPLE OF THE MATERIALS IN EXPERIMENT 2

| No. of propositions as subject | No. of propositions as object: 0 | | 1 | | 2 | |
|---|---|---|---|---|---|---|
| 0 | Reagan gambler<br>painter Reagan<br>Reagan milkman<br>florist Reagan | F<br>F<br>F<br>F | Ehrlichman janitor<br>student Ehrlichman<br>Ehrlichman gambler<br>florist Ehrlichman | F<br>T<br>F<br>F | Connally student<br>milkman Connally<br>janitor Connally<br>Connally gambler<br>painter Connally | F<br>T<br><br>F<br>F |
| 1 | Thurmond milkman<br>florist Thurmond<br>Thurmond gambler<br>janitor Thurmond | T<br>F<br>F<br>F | Daley painter<br>student Daley<br>Daley florist<br>milkman Daley | T<br>T<br>F<br>F | Mitchell painter<br>janitor Mitchell<br>florist Mitchell<br>Mitchell student<br>gambler Mitchell | T<br>T<br><br>F<br>F |
| 2 | Muskie student<br>Muskie florist<br>painter Muskie<br>Muskie milkman<br>gambler Muskie | T<br><br>F<br>F<br>F | Agnew gambler<br>Agnew milkman<br>florist Agnew<br>Agnew janitor<br>student Agnew | T<br><br>T<br>F<br>F | Humphrey milkman<br>Humphrey janitor<br>gambler Humphrey<br>painter Humphrey<br>Humphrey florist<br>student Humphrey | T<br><br>T<br><br>F<br>F |

*Note.* The first element given was the subject of the sentence and the second was the object. The verb in each sentence was *kicked*. T = true; F = false.

remembered set. The 36 to-be-remembered sentences were printed on computer cards. The subjects proceeded through these one at a time trying to commit them to memory. They were instructed to create out loud for each sentence some logical continuation. The purpose of requiring continuations was to encourage subjects to process the sentences semantically. After the subjects had completed the 36 sentences they were asked to recall the answers to all the possible questions (there were 24) of the form *Whom did Nixon kick?* or *Who kicked Nixon?* This assured not only that the subjects could recall the items that occurred with *Nixon* but that they could recall them in their correct logical relation. The sentences had been very imperfectly learned after the initial study, and subjects made many errors in response to these questions. Questions to which the subjects could recall the correct answers were dropped out of the set. After one pass through the set of questions, subjects were re-asked those questions that remained. They continued until there were no questions left. Then all 24 questions were rerandomized and this drop-out procedure was repeated a second time. This learning session lasted from about 20 min to an hour.

A reaction time phase followed. Subjects were shown propositions of the form *Nixon kicked the janitor* and *The janitor kicked Nixon*, and they had to judge whether each sentence was true or false. The sentences were presented to the subject on a tachistoscope. The presentation of the sentence started a timer. The subject pressed one of two buttons that stopped the timer and yielded a reaction time. Subjects were told when they made an error in their decisions. They were instructed to respond as fast as they could while maintaining accuracy. The subjects were tested on a sequence of 144 trials that was repeated three times. After each block of 144 trials, they were given a rest of a few minutes. This part of the experiment lasted about one hour.

*Materials.* The sentences were composed from the last names of 18 politicians, six professions, and one verb, *kick*. Unlike Experiment 1, all sentences were in the active voice. A different set of study sentences was constructed for each subject. Table 2 indicates the politicians and the professions used in half of the materials for one subject. There are nine conditions defined by number of propositions as subject and number of propositions as object. A different politician was randomly assigned to each of the nine experimental conditions. The study and test sentences for each condition are given in the various cells of Table 2. The sentences used in the reaction time phase have indicated to their right either a T for true or an F for false. Those sentences used in study are those marked T or without a symbol. The study sentences without Ts were not tested, to keep the number of distinct test sentences about a politician at four for all conditions. As can be seen, it was not possible to have an equal number of true sentences in all conditions. For instance, in the 0–0 condition there can, of course, be no true sentences. This design in Table 2 was replicated for another set of nine politicians.

The choice was deliberate to use one verb and only six professions. Previous data (Anderson, 1974a) had indicated that verification time is principally a function of the least number of propositions attached to a concept in the probe. This is to be expected according to the Anderson and Bower race model. Searches are simultaneously proceeding from all concepts. The one most likely to finish first is the one that has the fewest propositions to search. The

TABLE 3

MEAN REACTION TIMES (IN SECONDS) AND ERROR RATES (IN PARENTHESES)

| No. of propositions as object | True | | | | False | | | |
|---|---|---|---|---|---|---|---|---|
| | No. of propositions as subject | | | | | | | |
| | 0 | 1 | 2 | *M* | 0 | 1 | 2 | *M* |
| **Politician as subject in probe** | | | | | | | | |
| 0 | | 1.587 (.176) | 1.644 (.120) | 1.616 (.148) | 1.138 (.005) | 1.501 (.000) | 1.764 (.056) | 1.468 (.020) |
| 1 | | 1.731 (.157) | 1.626 (.148) | 1.679 (.153) | 1.671 (.042) | 1.675 (.019) | 1.984 (.056) | 1.777 (.039) |
| 2 | | 1.831 (.176) | 1.993 (.204) | 1.912 (.190) | 1.703 (.014) | 1.863 (.028) | 1.778 (.037) | 1.781 (.026) |
| *M* | | 1.716 (.170) | 1.754 (.157) | 1.735 (.164) | 1.504 (.020) | 1.680 (.016) | 1.842 (.050) | 1.675 (.029) |
| **Politician as object in probe** | | | | | | | | |
| 0 | | | | | 1.258 (.000) | 1.636 (.032) | 1.653 (.014) | 1.516 (.015) |
| 1 | 1.698 (.167) | 1.610 (.138) | 1.881 (.176) | 1.730 (.160) | 1.717 (.037) | 1.639 (.037) | 1.922 (.000) | 1.759 (.025) |
| 2 | 1.636 (.148) | 1.850 (.333) | 2.000 (.204) | 1.829 (.228) | 1.811 (.028) | 1.971 (.083) | 2.009 (.056) | 1.930 (.056) |
| *M* | 1.667 (.158) | 1.730 (.236) | 1.941 (.190) | 1.780 (.195) | 1.595 (.022) | 1.749 (.051) | 1.861 (.023) | 1.735 (.032) |

interest in the experiment is the search process from the proper name. To get maximal effect of number of propositions attached to the proper name, it is desirable to create large numbers of propositions leading from the verb and profession. Therefore, it is desirable to use few verbs and nouns in many propositions. Six was the fewest number of professions with which the design could be realized. Condition 2–2 in Table 2 requires two professions for the trues, two professions for the falses, and two more professions for the studied but not tested sentences. The professions were randomly chosen to fill the various roles in each cell. All professions occurred in five to seven propositions. The exact number was randomly determined for each subject. Because the same verb was used for all sentences, relation-specific search makes the same prediction with either a HAM or ELINOR representation.

There were 72 to-be-tested sentences (4 for each of the 9 cells in Table 2 and two replications of the design). Two sentences in each cell were tested with the politician as subject and two with the politician as object. Depending on the particular condition, 0, 1, or 2 of these were true. Altogether, there were 24 true and 48 false sentences. This created a 2:1 ratio of falses to trues, which created a considerable response bias. However, we were not concerned with the performance on trues relative to falses, but rather with differences among true conditions and among false conditions. Each of the 72 probe sentences was repeated twice within a block of 144 sentences. The order of sentences within a block was randomized.

*Subjects.* Because of the considerable work involved in constructing a new set of materials for each subject, only nine subjects were used. With the large number of trials each subject received, the data should nevertheless be quite reliable. The nine subjects were graduate students in psychology at Yale University. They were paid $6 for their participation in the experiment.

## *Results*

Table 3 presents the data for the 30 experimental conditions defined in the Design section. The standard error of the individual cell means in Table 3 is 110 msec. Orthogonal to these 30 conditions, the subject's data can be classified according to whether they occurred in the first, second, or third block of 144 trials. An analysis of variance was performed, using as factors the 30 conditions, the 3 blocks, and the 9 subjects. The data for the analysis were each subject's mean reaction time for each condition × block combination. There were four or eight potential observations contributing to each mean. Trials on which a subject's truth judgments were in error were excluded in calculating mean reaction time. In determining whether there were significant differences among other conditions, or blocks, or a Block × Condition interaction, the estimate of error used was the interaction of these treatments with subjects. Because different materials were used for each subject, statistical generalizations across subjects are also valid across materials (see Clark, 1973). There was a significant effect of conditions, $F(29, 232) = 3.28$, $p < .001$, and of blocks, $F(2, 16) = 14.99$, $p < .001$. There was a marked decrease in reaction times over the experiment, with a mean of 1.89 sec in the first block, 1.72 sec in the second, and 1.57 sec in the third. This type of speed-up effect was also found in a similar experiment (Anderson, 1974a). The Condition × Block interaction was not significant, $F(58, 464) = 1.17$.

It is clear from Table 3 that there is a general increase in reaction time with the number of propositions studied about the politician, either as subject or object. A model that posits relation-specific search predicts that, when the probe is tested with the politician as subject, all that should matter is the number of propositions studied about the politician as subject. This certainly does not appear to be the case. Reaction times increase as number of subject propositions or number of object propositions increases. In the true data, subjects are 296 msec slower with two object propositions than with none. In the false data they are 313 msec slower. Both effects are quite significant, $F(1, 232) = 7.17$, $p < .01$, and $F(1, 232) = 12.02$, $p < .001$, respectively. There is a 38-msec effect for 1 versus 2 propositions on the subject dimension for trues and a 338-msec effect of 0 versus 2 propositions for falses. The first effect is not significant, $F(1, 232) = .18$, probably because the contrast involves a difference of only one proposition. The effect for falses is very significant, $F(1, 232) = 14.02$, $p < .001$.

A relation-specific search model predicts that for object probes only number of object propositions should matter. Again this prediction appears to fail. The effects of 0 versus 2 subject propositions are significant both for trues (274 msec), $F(1, 232) = 6.01$, $p < .05$, and for falses (266 msec), $F(1, 232) = 8.81$, $p < .01$. The contrast between 1 versus 2 object propositions for true sentences (99 msec) is not significant, $F(1, 232) = 12.0$, but the contrast for falses between 0 and 2 object propositions (414 msec) is highly significant, $F(1, 232) = 21.03$, $p < .001$.

Thus, it appears that search is not relation specific. This conclusion may be criticized because it rests on comparisons including conditions where there were zero propositions for one or both grammatical functions. Perhaps the search operations in these zero-proposition cases are very different. To deal with this possibility, Table 4 presents an analysis of that subset of the data in Table 3 in which the numbers of propositions for the grammatical functions are one or two. The four matrices from Table 3 have been collapsed into one. The columns indicate whether there are one or two study propositions about the politician involving the probed relation. The rows indicate the number of propositions involving the nonprobed relation. This analysis is very telling. Both the probed and nonprobed dimensions show

TABLE 4

EFFECT OF PROBED AND NONPROBED DIMENSIONS ON VERIFICATION TIME (IN SECONDS)

| No. of propositions with nonprobed relation | No. of propositions with probed relation | | |
|---|---|---|---|
| | 1 | 2 | *M* |
| 1 | 1.664 | 1.857 | 1.761 |
| 2 | 1.874 | 1.945 | 1.910 |
| *M* | 1.769 | 1.901 | 1.835 |

substantial effects: The probed has an effect of 132 msec, $F(1, 232) = 5.70$, $p < .025$, and the nonprobed dimension 149 msec, $F(1, 232) = 7.27$, $p < .01$. The difference between the two effects is in the wrong direction and is not significant, $F(1, 232) = .19$.

It should be emphasized that, just as in Experiment 1, there is more than a failure to reject the null hypothesis. The null hypothesis is that there should be no difference between the probed and nonprobed dimensions. We have also shown that increasing the number of propositions on the nonprobed dimensions slows down search. Thus, although subjects were able to use relational information correctly to recall which people were kickers and which were kickees, they were not able to use the relational information to eliminate unnecessary searching. There has been a consistent failure to support HAM's hypothesis about relation-specific search.

### *Implications of the Failure to Find Relation-Specific Search*

There was good motivation for the Anderson and Bower proposal that search of memory is relation specific. That motivation came from consideration of the enormous amount of search that would be required in real-life situations without some mechanisms for eliminating unnecessary search. However, contrary to the expectations of HAM, subjects apparently do not use relational information to segregate their memories.

It might be argued that, these results notwithstanding, subjects do use relational information to guide memory search. Just because subjects were required to commit to memory *Nixon kicked the milkman*, it does not follow that they only committed to memory one proposition with *Nixon* as logical subject. They might have elaborated the stated fact with propositions like *The milkman cheated Nixon*. That is, although the subjects were presented with one subject proposition, they might have committed to memory $n$ subject and $n$ object propositions. This argument, in its extreme, makes relation-specific search empirically unidentifiable. According to it, whenever we commit to memory one proposition we really add $n$ propositions with each relation. Then the search time should increase very rapidly with number of facts we have committed to memory, independent of their relational structure. The point of these remarks is that at some deep level we may never be able to know whether search of memory is relation specific; at a functional level, however, it clearly is not.

The failure to find relation-specific search has important implications for the use of relational information in network models. Relational information is essential to record the logical relation between one idea and another. Therefore, current network models (i.e., Anderson & Bower, 1973; Quillian, 1969; Rumelhart et al., 1972) all have relational labeling of network arcs. However, all these models have more relations than the logical minimum. Rumelhart et al. even propose a facility for creating arbitrarily many labels. What is the justification for this proliferation of redundant relational labels? The more labels that a system has the greater the interpretive mechanism that will be necessary to perform deductions with them. The only published justification is that of Anderson and Bower (1973)—these relations could be used to cut down search time. The experiments reported here have eliminated that justification.

## General Conclusions

There seems to be some validity to the associative network representation and the search algorithms that naturally go with it. Subjects do store sentences in a way that causes individual sentences to interfere with each other. This item-specific interference manifests itself either in percentage of recall measures (Experiment 1) or in reaction time measures (Experiment 2). While other mechanisms could undoubtedly be marshalled to explain these effects, they are predicted by associative network search algorithms. However, subjects do not use relational information to make their searches more efficient. It remains unclear what mechanisms subjects do use to permit efficient search of their large data bases.

It has been suggested by some that these experiments constitute a serious refutation of HAM. However, as explained in Anderson and Bower (1973, chap. 16), HAM is a model of many independent assumptions. What these experiments indicate is that the search algorithm will have to be reformulated to operate in a way that does not use relational information. Doubtless further changes will have to be made in HAM. However, it is not the data in the experiments reported here that will force these other changes, for they are otherwise consistent with the HAM model.

## REFERENCES


Anderson, J. R. Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 1974, *6*, 451–474. (a)

Anderson, J. R. Verbatim and propositional representation of sentences in immediate and long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 149–162. (b)

Anderson, J. R., & Bower, G. H. On an associative trace for sentence memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 673–680.

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

Anderson, J. R., & Bower, G. H. *Human associative memory*. Washington, D.C.: V. H. Winston, 1973.

Asch, S. E. A reformulation of the problem of associations. *American Psychologist*, 1969, *24*, 92–102.

Ausubel, D. P. *The psychology of meaningful verbal learning*. New York: Grune & Stratton, 1963.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 335–359.

Fillmore, C. J. Some problems for case grammar. In D. Steinberg & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics, anthropology, and psychology*. London: Cambridge University Press, 1971.

Koffka, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.

Miller, G. A. *Language and communication*. New York: McGraw-Hill, 1951.

Postman, L. Unlearning under conditions of successive interpolation. *Journal of Experimental Psychology*, 1965, *70*, 237–245.

Postman, L., & Stark, K. The role of response availability in transfer and interference. *Journal of Experimental Psychology*, 1969, *79*, 168–177.

Postman, L., Stark, K., & Fraser, J. Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*, 1968, *7*, 672–694.

Quillian, M. R. The teachable language comprehender. *Communications of the Association for Computing Machinery*, 1969, *12*, 459–476.

Rumelhart, D. E., Lindsay, P., & Norman, D. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

Thorndyke, P. W., & Bower, G. H. Storage and retrieval processes in sentence memory. *Cognitive Psycgology*, 1974, *6*, 515–543.

Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychological Monographs*, 1945, *59* (3, Whole No. 273).


(Received July 22, 1974; revision received October 24, 1974)