Recent studies have offered support for a constructive view of sentence memory in children, based on their preference in recognition errors for true inferences, which can be drawn from input sentences, over false inferences. However, with the materials used in these studies, this preference may reflect responding either on the basis of semantic or formal similarity to the original sentences. The present experiment separates these factors. Both semantic and formal similarity between input and test sentences are found to be significant and independent factors which together determine if children (second and fifth grade) will accept new sentences as old in a recognition test. The effect of formal similarity is greater for the younger children. Moreover, the patterning of recognition errors on contradictory true and false inferences suggests that children of both ages monitor their own decisions so as to remain internally consistent with respect to the meaning of those test sentences indicated as old. (Author)
Technical Report No. 76
The Effects of Semantic and Formal Similarity on Recognition Memory for Sentences in Children

Thomas J. Thieman and Ann L. Brown
Institute for Child Behavior and Development
University of Illinois at Urbana-Champaign
November 1977

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02138

The preparation of this manuscript was supported in part by the U. S. Public Health Service Grants Nos. HD-00244 and HD-06864 from the National Institutes of Child Health and Human Development and in part by the National Institute of Education under Contract No. MS-NIE-C-400-76-0116.
The Effects of Semantic and Formal Similarity on Recognition Memory for Sentences in Children

In a series of recent experiments, Paris and his associates (Paris & Carter, 1973; Paris & Mahoney, 1974; Paris, Mahoney, & Buckhalt, 1974) have tested the validity of a constructive view of sentence memory in children of several ages. In keeping with studies of recognition memory for underlying ideas in adults (cf., Bransford & Franks, 1971; Bransford, Barclay, & Franks, 1972), the children were read a series of short stories and then given a recognition test which included original premise (old) sentences as well as both true and false inference (new) sentences related to the original stories. If children's memory representation of the stories is limited to the syntax of the original individual sentences, then they should fail to recognize true inferences, which entail the integration of information across several sentences, and false inferences equally often, given they both involve changes in syntax. On the other hand, if children actively construct semantic descriptions of the stories while failing to maintain their exact syntactic form, then they should accept or reject inferences as old or new on the basis of their semantic congruence with the original stories rather than their syntactic similarity. In fact, Paris reports the latter holds for both normal (Paris & Carter, 1973; Paris & Mahoney, 1974) and retarded (Paris, Mahoney, & Buckhalt, 1974) children as young as seven years old.

These results suggest a direct continuity of processes in the assimilation of meaningful material between young children and adults, with no
developmental differences other than those attributable to increased memory span or the use of mnemonic skills. However, this generalization bears closer inspection in light of the particular experimental procedures on which it is based. For example, in the Paris and Carter (1973) study, second- and fifth-grade children were read seven three-sentence stories, such as: (a) The bird is inside the cage. (b) The cage is under the table. (c) The bird is yellow. The recognition items for this story included 'a' as the true premise sentence, together with "a slightly altered false premise", (d) The cage is over the table, "a permissible true inference", (e) The bird is under the table, and "an invalid false inference", (f) The bird is on top of the table.

If the children form a nonlinguistic representation of the story, they are more likely to false recognize the true inference, e, than either of the false statements, d or f. However, the same result is predicted if one assumes that the children simply store surface information about each story. It should be noted that whereas the true inference includes a relational term (under) which occurred in the original story, the two false sentences use relational terms (over and on top of) which are novel to the story. This structural distinction between true and false new sentences is present in the test items of all seven stories used by Paris and Carter. Thus, the false sentences can be rejected as new, not only because they are inconsistent with the meaning of the original stories, but also because they introduce new surface information. Paris and Carter (1973) lightly dismiss the importance of this structural feature of their materials by affirming that "the only difference between true and false inferences was the validity of the relation term, a
subtle semantic, not syntactic, difference" (p. 111). It appears this may have been a premature conclusion, at least pending further study.

The more discriminating recognition test used in the present experiment includes two types of false sentences, those which introduce new relational terms and those which use the same terms as the true premises. Paris and Carter would predict no difference in the false recognition rates for these two item types. However, if the children respond on the basis of retained surface information, then the familiarity of the relational term will be a critical factor, such that new sentences with familiar relational terms may be false recognized at a high rate, regardless of whether they are true or false with respect to the original stories. Moreover, this finer analysis may reveal developmental differences which have evaded detection in the past.

Method

Subjects

Fifty second-grade children (range of CA = 7.1 to 8.4; mean = 7.8) and 60 fifth-grade children (range of CA = 9.10 to 11.2; mean = 10.6) from three local public schools participated in the experiment. There were 58 males and 52 females tested, with approximately equal numbers of males and females at each grade level.

Task

The children were read an initial list of sentences and later given a recognition test for those sentences. The acquisition list contained seven unrelated stories, each made up of three sentences describing an event or a scene. For example, the first story was:

J1
The telephone is on top of the book (7).
The book is under the table (8).
The telephone is ringing (9).
As in the Paris and Carter (1973) study, all the sequences followed
the same design of $A \times B, B \times C$, inviting the inference $A \times C$. The remaining
six stories are presented in Table 1.

The recognition test consisted of four different sentences for each
story: the true premise (TP), $A \times B$, the invited true inference (TI), $A \times C$,
the invalid false inference (FI), $A \times C$, and a false statement (FS) similar
to Paris and Carter's false inference, which used a new relational term,
$A \times C$. For example, for the first story the test sentences were:
The telephone is on top of the book (10) TP.
The telephone is under the table (11) TI.
The telephone is on top of the table (12) FI.
The telephone is over the table (13) FS.
In the recognition list, the four sentences relating to each story
were blocked. The order of testing the seven stories was initially ran-
domized and presented to all children in the same order. However, the
order of sentence types within stories was arranged so that half the students
at each grade level received one order and the other students were tested
with the same sentences in reverse order. The within-story test orders
included the following two constraints: (a) the TI and FI items from the
same story, which were opposite in meaning with respect to the relational term, were always separated by at least one other test sentence from that story, and (b) across both test orders, a given TI sentence preceded and followed its related FI sentence equally often.

Procedure

The children were tested in large groups at each grade level. The instructions stressed verbatim memory. The experimenter asked the children to "listen carefully to exactly what I say" since they would be tested later for "how well you can remember the words in the stories." Each sentence in the acquisition list was read aloud at a normal speaking rate. After the last sentence was read, the students were instructed to work for five minutes on a hidden word puzzle which had previously been distributed and explained. The recognition test for all seven stories followed this interpolated activity. Each student recorded his or her own responses on a provided answer sheet. For each sentence, which was read, the child was told to write yes if the exact sentence was heard before and no if the sentence had not been previously heard. Then the student indicated his or her certainty in that response on the three-point scale described by Paris and Carter. A response of 1 indicated the child was "real sure", a 2 indicated "kind of sure", and a 3 indicated "not too sure". For reference, this scale was printed at the bottom of the answer sheet. The entire experiment was completed within 20 minutes.

Results

The overall percentage of recognition errors for second and fifth grade students for each of the four sentence types is shown in Figure 1.
As in the Paris and Carter (1973) study, children at both grade levels made the greatest number of recognition errors on true inferences. However, it is also clear that the error rates for the remaining sentence types are not equal, and that the second grade children, in particular, made a substantial number of errors on false inferences which were meaningfully inconsistent with, but syntactically and formally similar to, the original stories. An initial analysis of variance was performed to assess the effects of grade level, stories and sentence types on recognition performance. The main effects of grade, $F(1, 108) = 22.52$, and sentence type, $F(3, 324) = 65.48$, as well as a smaller effect for stories, $F(6, 648) = 4.99$, were significant (all $p < .001$). Grade level did not interact significantly with sentence type or stories. However, the interactions of Stories X Sentence Type, $F(18, 1944) = 9.14$, and Grade X Stories X Sentence Type, $F(18, 1944) = 2.56$, were both significant at the .001 level. Thus, while the overall pattern of errors was relatively consistent across grades, there were some systematic differences related to specific sentence types occurring in specific stories.

The large main effect for sentence type was further analyzed by means of Newman-Keuls comparisons on error rates averaged across both grades. When the sentence types are rank ordered in terms of percentage of errors, from most errors on true inferences to fewest errors on false statements, all pairwise comparisons are significant at the .05 level. In particular, this analysis confirms that recognition errors are much more likely to occur on false inferences which include familiar though inappropriate relational terms than on false statements which introduce additional surface cues.
which may aid in correctly rejecting those sentences as new.

A second analysis of variance examined the effect of test order on each sentence type. It will be recalled that two different orders of testing sentences within each story were used, so that each sentence was tested equally often early (first or second) or late (third or fourth) in its four-item question block. This order factor should be particularly crucial for recognition performance on the opposite-meaning true and false inferences which always occupied different halves of the question block. In the overall analysis of variance, Grade X Sentence Type X Half of Question Block, the main effects of grade, $F(1, 108) = 22.05$, and sentence type, $F(3, 324) = 68.53$, were again highly significant ($p < .001$), whereas the main effect of half of question block was nonsignificant. Although there were no significant interactions with grade, the Sentence Type X Half of Question Block interaction was significant, $F(3, 324) = 4.63$, $p < .005$.

---

The percentage of recognition errors across halves of question blocks for each sentence type is shown separately for second and fifth grade students in Figure 2. Of particular interest are the percent of errors made by second grade children on true and false inferences tested within the first half of a question block. In these cases, the younger children falsely recognized the false inferences nearly as often as the true inferences. On the other hand, fifth grade students showed a much lower error rate on false inferences than on true inferences, even when the sentences are tested early. Thus, under this one set of conditions, reliable developmental differences
are evident in the way the children react to the critical false inference items.

Moreover, children at both grade levels made fewer errors on false inferences when they were presented within the second half of a question block. This decline in error rate for false inferences runs counter to the general trend of increasing errors when sentences are tested later in the same block. It appears that the false inferences are initially quite attractive as recognition foils, especially to the second graders, due to their syntactic and formal similarity to the original stories. However, the children less frequently accept these sentences as old after they have been tested, within the first half of each question block, with the even more attractive and opposite-meaning true inferences. In practice, the children were reluctant to contradict themselves by false recognizing both the true and false inferences in the same story block. This suggests that the children were actively monitoring the internal consistency of their recognition decisions, and sentences were responded to both in terms of their semantic and formal relation to the original stories and their relation (consistent or inconsistent) to previous responses.

The nonindependence of errors on the two types of inferences is supported by several conditional probabilities. For example, across both grades the probability of making a false inference error, given the true inference was already correctly rejected, is .197, compared to the probability of making the false inference error following a true inference error of only .140. A similar constraint holds when true inferences are tested in the second half of the question block. Specifically, the probability of making a true
inference error, given the false inference was previously correctly rejected, is .458, compared to the probability of making the true inference error, following a false inference error of .319.

Finally, the confidence ratings which the children assigned to their recognition responses also show systematic differences across grade level in the way the students treated true and false inferences. Table 2 presents:

---

Table 2 about here
---

the percentages of certainty judgments which were assigned the highest rating of "real sure" for correct and erroneous responses to each of the sentence types by second and fifth grade children. The fifth graders exhibit the same pattern of results as that reported by Paris and Carter (1973). That is, correct rejections of true inferences are made with lowest confidence, but false recognitions of true inferences are made with highest confidence. In fact, for the older children, true inferences are the only sentence type for which confidence is higher on errors than on correct responses. However, with the second graders, this relationship holds for the false inferences as well. Both false and true inferences are incorrectly accepted as old sentences with high confidence. This again points to the difficulty which the younger children experience in deciding whether they have heard the false inferences before; a problem not nearly as pronounced with the false statements.

Discussion

Overall, the results confirm earlier reports that children quite often false recognize valid but unstated inferences as having been presented in the context of a meaningful passage. While this finding is predicted by a
A construcutive view of memory, it is also consonant with the view that recognition responses are made on the basis of syntactic and formal similarity, at least with respect to the test materials used prior to this study. This latter position is partially upheld by the differences in recognition error rates for the two types of false statements in this study. Children at both grade levels were highly proficient at correctly rejecting false statements, which were equivalent to Paris and Carter's (1973) false inferences, in which the relational terms were both incorrect (i.e., not meaning-preserving) and new. In comparison, over twice as many recognition errors were made on the false inferences introduced in this study, in which the relational terms were incorrect but old. Nonetheless, the overall error rate on false inferences is still significantly below that for true inferences, in which the relational terms are both old and correctly applied in a meaning-preserving manner. The only exception to this general finding involves second graders, who false recognized almost as many false inferences as true inferences when the sentences were tested early in a question block.

This latter finding suggests that the second and fifth grade children are equally adept at distinguishing between valid and invalid inferences which can be drawn from a simple story, but the younger children are more heavily influenced in a recognition memory test by the degree of formal similarity between the original and test sentences. If one also assumes that the testing procedure per se helps to reinstate the memory of the original story, then the different effect of test sequence on true and false inferences for the second grade children can be explained. The formal similarity of the false inferences to the original sentences is found to be in
contradiction with the meaning of the stories during the course of testing, and the false inference error rate decreases from first to second half of questioning. On the other hand, the formal similarity of the true inferences is recognized as consistent with the meaning of the original stories, and their false acceptance rate increases with later testing.

One further piece of evidence reported by Paris and Mahoney suggests that children at both grades are highly sensitive to the formal similarity between original and test sentences. Paris and Mahoney presented sentences of the type A is to the right of B and tested for recognition with the true but formally dissimilar sentence B is to the left of A. Thus, their "propositionally-similar" recognition items involved changes in both word order and the relational term. When presentation and test were both in sentential form (the Verbal-Verbal condition), children responded randomly to both true and false inferences written in the inverted syntactic form. However, when the original presentation was in the form of a picture (the Picture-Verbal condition), the children were able to correctly identify either meaning-preserving syntactic form as correctly expressing the picture verbally. Thus, the failure of the children to differentially false alarm to true and false inferences written in an inverted form reflects a recognition bias rather than a failure to comprehend.

In summary, the results show that both semantic and formal congruence between original and test sentences are significant and independent factors which together determine the probability that children will accept new sentences as old in a recognition test. Earlier studies confounded these two factors in the types of sentences used for recognition testing, and
inappropriately dismissed the importance of formal similarity. The present results further suggest that younger children are more sensitive to formal similarity than older children, as a basis for making recognition decisions. Finally, children of both age groups appeared to monitor their own responses within each block of test sentences relating to a given story, and in so doing, limited the number of meaningfully-inconsistent recognition decisions they made. While these results do not necessarily contradict a constructive view of sentence memory in children, they do point out that recognition memory decisions by children can be routinely influenced by several factors other than the degree of semantic similarity between input and test sentences.
References


This research was supported by U. S. Public Health Service Grants Nos. HD-00244 and HD-06864 from the National Institutes of Child Health and Human Development. Requests for reprints should be sent to Ann L. Brown, Institute for Child Behavior and Development, 51 Gerty Drive, Champaign, IL, 61820.

1. Our thanks to Scott Paris for making his test material available for our inspection.
Table 1
Additional Six Acquisition Stories

<table>
<thead>
<tr>
<th>The tiger is inside the cage.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The cage is behind the circus tent.</td>
</tr>
<tr>
<td>The tiger is very hungry.</td>
</tr>
<tr>
<td>The frog jumped over the bug.</td>
</tr>
<tr>
<td>The bug was sitting on a leaf.</td>
</tr>
<tr>
<td>The frog was green.</td>
</tr>
<tr>
<td>The dog is under the bed.</td>
</tr>
<tr>
<td>The bed is to the right of the chair.</td>
</tr>
<tr>
<td>The dog is named Sam.</td>
</tr>
<tr>
<td>The doll is on top of the toy box.</td>
</tr>
<tr>
<td>The toy box is in front of the TV.</td>
</tr>
<tr>
<td>The doll is Raggedy Ann.</td>
</tr>
<tr>
<td>The boy ran into the yard.</td>
</tr>
<tr>
<td>The yard is near the house.</td>
</tr>
<tr>
<td>The boy had a football.</td>
</tr>
<tr>
<td>The apple is in the bag.</td>
</tr>
<tr>
<td>The bag is next to the refrigerator.</td>
</tr>
<tr>
<td>The apple is good to eat.</td>
</tr>
</tbody>
</table>
### Table 2
Percentage of Highest Confidence Ratings of Recognition Responses

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Grade 2</th>
<th></th>
<th>Grade 5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>Error</td>
<td>Correct</td>
<td>Error</td>
</tr>
<tr>
<td>True Premise</td>
<td>87.8</td>
<td>60.9</td>
<td>87.2</td>
<td>46.3</td>
</tr>
<tr>
<td>False Statement</td>
<td>67.3</td>
<td>59.1</td>
<td>81.4</td>
<td>33.3</td>
</tr>
<tr>
<td>False Inference</td>
<td>68.7</td>
<td>74.5</td>
<td>80.2</td>
<td>59.7</td>
</tr>
<tr>
<td>True Inference</td>
<td>61.6</td>
<td>73.3</td>
<td>66.9</td>
<td>74.2</td>
</tr>
</tbody>
</table>
Figure Captions.

Figure 1. Percent recognition errors.

Figure 2. Percent recognition errors across halves of question blocks
(TI = True Inference; FI = False Inference; TP = True Premise; FS = False Statement).
HALF OF QUESTION BLOCK
CENTER FOR THE STUDY OF READING
READING EDUCATION REPORTS


No. 2: Asher, S. R. Sex Differences in Reading Achievement, October 1977.


* No. 8: Mason, J. M. *Questioning the Notion of Independent Processing Stages in Reading*, February 1976. (Journal of Educational Psychology, 1977, 69, 288-297)


No. 34: Bruce, B. C. Plans and Social Actions, April 1977.


No. 54: Fleisher, L. S., & Jenkins, J. R. Effects of Contextualized and Decontextualized Practice Conditions on Word Recognition, July 1977.


No. 59: Mason, J. M. Reading Readiness: A Definition and Skills Hierarchy from Preschoolers' Developing Conceptions of Print, September 1977.
