The Role of Imagery in Sentence Memory: A Developmental Study

Stephen M. Kosslyn and Gordon H. Bower
Stanford University

KOSSLYN, STEPHEN M., and BOWER, GORDON H. The Role of Imagery in Sentence Memory: A Developmental Study. CHILD DEVELOPMENT, 1974, 45, 30–38. This experiment tested whether children rely relatively more than adults on sensory imagery as the internal representation of sentences in memory. If so, then children should show more confusions in recognition memory between sentences which evoke very similar imagery. Adults, on the other hand, should be able to encode, remember, and distinguish sentences differing conceptually even though these sentences evoke very similar imagery. The experiment used sentence pairs rated as evoking highly similar imagery. Ss studied 1 list of sentences and then judged a test list of sentences regarding whether each was a verbatim replica of 1 studied earlier. Compared with a studied counterpart, each test sentence was either verbatim the same, imaginally similar and conceptually similar, imaginally similar but conceptually different, or imaginally and conceptually different. Recognition confusion errors supported the hypothesis that children did not discriminate between same versus imaginally similar sentences whereas adults did. Although children comprehended conceptual distinctions nearly as well as adults, they forgot such distinctions more readily, so that imaginal similarity guided their recognitions.

Several developmental psychologists (e.g., Bruner, Olver, & Greenfield 1966; Church 1961) have proposed that imagery is the predominant mode of cognitive representation of information in children before they develop facility with language. While older children and adults might habitually encode information in terms of linguistic abstractions (e.g., the “conceptual structures” described by Schank [1972]), it is alleged that children rely relatively more on concrete, sensory imagery. The present study investigates an implication of this thesis regarding children’s and adults’ memory for sentences. While previous developmental investigations have examined the role of imagery in memory for visual patterns (e.g., London & Robinson 1968, 1971), in memory for paired associates (e.g., Rohwer 1970; Wolff & Levin 1972), and in various types of transformations and problem-solving tasks (e.g., Piaget & Inhelder 1971), no previous developmental studies have considered the role of imagery in memory for sentences.

We shall be examining confusion errors in recognition memory for sentences. The logic of using recognition confusions is roughly the same as tests for “stimulus equivalence” and is as follows: It is supposed that when S studies and comprehends a particular sentence, he sets up an internal representation of the information he has extracted from it. Although S’s immediate memory may include aspects of the sentence’s exact phonetic (or literal) presentation, these “surface features” presumably decay from short-term memory, leaving as the more enduring residue something like the conceptual or imaginal representation of the sentence studied. A list of studied sentences will leave behind a set of such memory structures. Upon presentation of a test sentence in a later recognition memory test, its surface structure will be encoded into a deep-structure form; this input structure is presumed then to be matched in some way against the set of memory structures established by the study list. If the deep structure of the test sentence is sufficiently “similar” to some one or more structures in memory (the exact metric is undefined), then

The authors express their appreciation to C. J. Sengul and Sheri Robinson who assisted in collection of the data and to the staff of the Bing Nursery School. Research of the second author is supported by grant MH-13950 from the National Institute of Mental Health. Reprints may be obtained from Stephen Kosslyn, Department of Psychology, Stanford University, Stanford, California 94305.

[Child Development, 1974, 45, 30–38. © 1974 by the Society for Research in Child Development, Inc. All rights reserved.]
it is assumed that S decides that the test sentence is one he remembers from the study list.

Focusing upon a particular study sentence and its internal representation, one can then imagine a range or domain of other sentences which differ in their surface features but which are all equally confused in long-term memory with the originally studied sentence. Now, recognition confusion may occur for a variety of reasons, including, for example, simple failure to learn all of a complex proposition. But a sure cause for confusing the test sentence with the memory of the studied sentence is if the encoding mechanism maps them into the same or very similar internal representations. Given these working assumptions, one can then use systematic confusion errors in recognition memory as an index (albeit a "noisy" one) of "similarity" of the internal representations in memory of two sentences. In turn, knowledge of which pairs of surface sentences are encoded similarly and which pairs differently may provide some leads in inferring the nature of the representations and the mechanism which encodes surface strings into these representations. At least, that is the hope underlying use of confusions in recognition memory for investigating cognitive representations (see Fillenbaum [1970] for discussion of some problems with this methodology).

The present experiment inquires whether the internal representation of a sentence in memory has a substantial imagery component; more specifically, we ask whether this imagery component is more prominent for children than for adults. The logic, again, is that if a sentence is remembered in terms of an image that it evokes, then a new test sentence ("lure") which evokes about the same imagery and which is encoded and judged primarily in terms of this imagery will then be "recognized" as a repetition of the earlier sentence studied (e.g., a "false alarm").

Previous studies with adults by Begg (1971) and Anderson (1972) partially tested this imagery hypothesis. While finding an appreciable percentage of false alarms to imaginally similar lures, Begg found that adults could reliably discriminate between paraphrase lures and actual recurrences of sentences originally studied. (Paraphrases were presumed to evoke the same imagery.) Anderson also found that adults did not confuse sentences primarily on the basis of their imagery. He argued rather that sentences were confused in memory according to similarity of their semantic or conceptual content and independently of the similarity of the imagery they evoked.

A problem with these earlier studies is that neither followed procedures to insure that their sentence pairs did in fact evoke highly similar imagery. To illustrate, one of the sentence pairs which Anderson claimed to be "imaginally the same but conceptually different" differed in providing a judgmental description of an act: either "John signed his signature on the check," or "John forged his signature on the check." Adults can remember later which of these two sentences they had studied. But this may occur because they use differing elaborative features of the image to "carry" the otherwise latent judgment. Thus, in contrast to the "straight" check signer, the forger may have a handlebar mustache, a furtive manner, and do the signing with slow deliberation. Of course, trained method actors are skilled at making manifest, in auxiliary actions visible to an audience, such subtle nuances of mood, intentionality, uncertainty, and feelings. Accordingly, one might "use a good actor" to portray in internal imagery the episodes or states of affairs depicted by the study sentences. In this case, then, many of Anderson's or Begg's sentence pairs might not be imaginally identical—which may account for the degree of discrimination they observed.

To assure high imaginally similarity of study and test sentences, we used pairs of sentences which had been prerated by normative Ss as evoking identical or highly similar imagery. To contrast the relative importance of imaginical versus conceptual similarity in recognition-memory confusions, we compared false-alarm rates to imaginally similar sentences which either did or did not have the same conceptual meaning. Our main hypothesis was that adults would be able to discriminate between reoccurrence of a studied sentence and presentation of an imaginally similar lure, while children would not. Moreover, within the class of imaginally similar lures, adults should make more errors on conceptually same sentences than conceptually different sentences, whereas children should false alarm about equally to these two types of imaginally similar lures.

**Method**

Subjects heard a study list (or lists, in the case of the children) of sentences to learn. A
second test list was immediately presented, and S judged whether the exact verbatim replica of each test sentence was or was not on the study list.

**Materials**

All the sentences were concrete (see Paivio 1971) and were composed of seven to 11 words. One third of the sentences on the test list were exactly identical with some counterpart in the study list; these sentences will be referred to as the “sames.” Another one third of the test-list sentences were composed by rearranging words taken from the study-list sentences. These “different” sentences did not express meanings similar to any of the study-list sentences. These should be easily rejected if S were storing any meaningful representation of the study sentences. The final third of the test-list sentences were chosen so that they each evoked a visual image similar to that evoked by a counterpart sentence of the study list. Let us describe how these “imagery” sentence pairs were constructed and selected.

One hundred and fifty pairs of sentences were generated and then rated on a seven-point scale for similarity of the evoked images by 25 graduate psychology students. The 25 pairs of sentences which were rated as having the most similar images (see table 1) were used in the experiment. An informal panel of judges initially determined that both sentences in 10 of the imagery pairs expressed the same conceptual meaning (e.g., “The flowers in the green vase were on the table,” and, “The flowers on the table were in a green vase”). It also was determined that the sentences in 10 other imagery pairs expressed different meanings (e.g., “The lion killed the buffalo because he was hungry,” and, “The lion killed the buffalo, so he must have been hungry”). A later recheck of these “conceptual similarity” ratings with 30 Stanford undergraduates produced nearly perfect agreement with the informal panel. Unfortunately, there was one pair for which the majority of the raters disagreed with our earlier panel. Using the raters’ classifications, then, among our imaginally similar sentence pairs we had nine that had the same conceptual meaning and 11 that had different meanings. Pairs where both members expressed the same meaning will henceforth be referred to as “conceptually similar” (CS), and pairs where the meanings are different will be termed “conceptually different” (CD). These distinctions are marked in table 1. Conceptually different pairs were characterized by sentences that differed in the expression of causality, intentionality, time, judgments, and appearance versus reality.

For the adults, the 75 study-list sentences were tape recorded in a random order. Test-list sentences that were the same as study-list sentences were recorded in the same relative position on both the test list and the study list. Imagery test-list items were recorded in the same position as the corresponding study-list sentence. This procedure insures a constant lag between study and test of any target sentence. Sentences were recorded every 7 sec, and the test list followed 60 sec after the termination of the study list.

To increase performance of the children, the set of sentences was presented as three shorter blocks or sublists. Each block comprised 20 study sentences followed by 20 corresponding test sentences. These sentences also were 7 sec apart, and test sentences occurred 60 sec after the study list. Within each block approximately one third of the test sentences were same, different, and imagery tests. Also within each block approximately half of the imagery sentences were conceptually similar and half were conceptually different. The order of the sentences in the study lists was random. As in the adults’ list, the serial position of test-list items mirrored the study-list order. The same conceptually same and conceptually different imagery sentences were used with both adults and children. All 60 of the children’s sentences were used in the adult lists; both age groups received the same proportion of each type of test sentence.

**Procedure**

The original design of this experiment included two sets of learning instructions. Within each age, half of the Ss were to be instructed to learn each sentence by imaging its referent episode or state of affairs, whereas the other Ss were to be instructed to learn by “attending to the conceptual meaning” of each sentence. Two groups of adults were tested utilizing this procedure. However, no memory differences whatsoever emerged as a consequence of these instructions, and Ss’ self-reported use of imagery did not correspond well with instructions. Thus, the adult data were pooled over these instructional differences, and such differing instructions were not used with the children.

**Adults.** Adults were tested in small groups of two to seven. They received printed
TABLE 1

CHILDREN’S AND ADULTS’ PROPORTION OF “OLD” RESPONSES TO IMAGERY ITEMS

The top sentence of each pair was a member of the study list and the bottom was on the test list. To the right of each pair is the conceptual similarity designation (derived from adult ratings): CS, indicating conceptually similar, or CD, indicating conceptually different. To the right of the conceptual similarity designation is the proportion of false alarm (“same,” “old”) judgments given to that pair by children (left number) and adults (right number).

1. a) Henry scraped deep into the ground with his hoe. (CS: 50, 30)
   b) Henry gouged the ground with his hoe.
2. a) The ball rested on the box in front of the tree. (CS: 60, 80)
   b) The ball in front of the tree rested on the box.
3. a) The fly flew over the dog that was under the blanket. (CS: 40, 60)
   b) The fly flew over the blanket that was over the dog.
4. a) The dog chased after the frisbee into the bushes. (CS: 90, 75)
   b) The dog ran into the bushes after the frisbee.
5. a) The ambulance went down the street with its lights flashing. (CS: 70, 15)
   b) Using its flashing lights, the ambulance went down the street.
6. a) The flowers in the green vase were on the table. (CS: 40, 40)
   b) The flowers on the table were in a green vase.
7. a) Homer Smith was much taller than fat old Stephen. (CS: 50, 10)
   b) Fat old Stephen was much shorter than Homer Smith.
8. a) A falling raindrop was the reason Sally’s nose was wet. (CS: 60, 70)
   b) Sally’s nose was wet because a raindrop fell on it.
9. a) The glass of Coca-Cola was half full. (CS: 60, 45)
   b) The glass of Coca-Cola was half empty.
10. a) The crayon rested in the boy’s hand as he drew. (CD: 50, 35)
    b) The crayon rested between the boy’s fingers as he drew.
11. a) I have washed my new car twice today. (CD: 90, 40)
    b) I washed my new car twice yesterday.
12. a) Hank dug the hole furiously a day early. (CD: 40, 10)
    b) Hank dug the hole furiously a day late.
13. a) George picked up the pen to write on the paper. (CD: 20, 0)
    b) George picked up the pen to doodle on the paper.
14. a) Jimmy got his present because Santa knew he deserved it. (CD: 90, 55)
    b) Jimmy got his present because Santa thought he deserved it.
15. a) The lion killed the buffalo because he was hungry. (CD: 60, 15)
    b) The lion killed the buffalo, so he must have been hungry.
16. a) A large brown cow was falling asleep on the grass. (CD: 60, 55)
    b) A large brown cow was laying down on the grass.
17. a) My father’s friend Ed always looked very happy. (CD: 80, 30)
    b) My father’s friend Ed always was very happy.
18. a) The big white sack looked exactly like a pillow. (CD: 70, 15)
    b) The big white sack was really a pillow.
19. a) Pat looked like she thought she knew the answer. (CD: 60, 65)
    b) Pat looked like she really knew the answer.
20. a) The peanut butter and jelly sandwiches were made a few hours ago. (CD: 80, 45)
    b) The peanut butter and jelly sandwiches were made an hour ago.

instructions to attempt to remember the list of sentences that they were about to hear. The study list was then presented. During the minute between the end of the study list and initiation of the test list Ss turned to a page on which were listed numbers 1–75. Next to each number were two brackets. One column of brackets was labeled “old” at the top, and
the other was labeled “new.” The Ss were instructed to listen carefully to each sentence on the test list and to indicate whether they thought it had been on the study list by checking the bracket under “old” next to its number. If they thought it had not been on the study list, Ss were to check the bracket under “new.” It was emphasized that in order for a sentence to be considered “old” it must match a study-list sentence verbatim (“word for word”).

Children.—Children were tested individually. Before beginning the actual experiment it was necessary to assure that the child understood what was meant by “same” and “different” verbatim sentences. The child was tested for his concept of same and different with a pair of sentences differing by active-passive voice (“The fox chased the rabbit around the barn” vs. “The rabbit was chased by the fox around the barn”), with a pair of sentences differing by a quantifier (“The boy ate all of the ice cream” vs. “The boy ate some of the ice cream”), and with a pair of identical sentences (“They bought a lot of new furniture for their house”). If the child expressed any uncertainty, he was trained until he could correctly classify sentences as verbatim same or verbatim different. It was emphasized that for two sentences to be called the same they had to be exactly the same, word for word. After S’s understanding of “same” and “different” was established the experiment proper began.

Each S was instructed to listen carefully to each sentence. (The study block was not presented until we were sure he was attending.) In the minute following presentation of each block of study items S was told that a test list of sentences would be presented, and he was to say “same” if a sentence was exactly the same (word for word) as one on the first list and “different” if it was different. Subjects experienced no difficulty in making these judgments. For most Ss, all three study-list blocks were presented during the same session. Three Ss completed only two blocks during the first session, but finished the third several days later. Five Ss wearied and failed to complete all three blocks. Despite these losses of observations, the sublist orders were so rotated that the total number of observations per sentence type was equal, and all sentences were presented an equal number of times.

Subjects
The subjects were seven boys and five girls ages 5-2 to 5-7 (median age = 5-4) from Stanford’s Bing Nursery School population and 20 adults (12 males and eight females) from an introductory psychology class at Stanford University.

Results
Since we wish to generalize our results beyond the specific items used in this experiment, as well as the particular Ss, it is necessary to consider both items and Ss as random effects (see Clark 1973). Thus, for each comparison t tests were performed twice, once pooling over items exemplifying a given condition for each S (the standard method) and a second time pooling data from different Ss and analyzing in terms of the actual items within each condition. Unless specifically noted, reported significance levels are representative of both analyses.

The primary data are the proportions of times a test sentence of a given type is judged to be a verbatim replica of a studied sentence (old). These proportions are shown in figure 1 for the two populations and three categories of test sentences. Each proportion is based on 500 observations for the adults and 200 for the children. The adults clearly show very reliable discrimination among all three types of test sentences; separate t tests between adjacent conditions are all highly significant statistically (same vs. imagery, p < .001; imagery vs. different, p < .001). The high false-alarm rate of adults to imaginally similar sentences is important in conjunction with the otherwise very high degree of learning (discrimination) shown for same versus different sentences.

The children show a different pattern of recognition confusions. First, their overall discrimination is poorer than that of the adults; the children have a lower hit rate on same items and a higher false-alarm rate on different items. These main effects in themselves are not of much interest. What is of interest is that the children do not discriminate at all (in memory) between same and imagery sentences; observed proportions of “acceptances” were 66% and 61%, respectively, for the two sentence types. Although these proportions do not differ reliably from one another, they both exceed the false-alarm rate (18%) to different
items ($p < .001$ in both comparisons). So while the children clearly discriminate between "meanings" which were or were not presented, they do not discriminate in memory between sentences which evoke highly similar imagery. The difference in this differentiation between children and adults was assessed by two $t$ tests, one across Ss and one over times. For the first test, each S received a score reflecting his difference in old proportions to same versus imagery test sentences. These discrimination scores proved to be reliably greater for adults than for children, $t(30) = 2.90, p < .01$. The items analysis was performed by subtracting the mean proportion of old responses to the imagery items from the proportion of old responses to same items. Once again, these discrimination scores were greater for adults than for children, $t(43) = 4.47, p < .001$.

Our second interest concerned the discrimination in memory between imaginally similar sentences which are either CD or CS (see table 1). As might be anticipated from the results above, children and adults also differed in this respect. The children false alarmed more to CD items (65%) than to
CS items (56%), whereas adults showed the reverse trend, false alarming more to CS items (47%) than to CD items (36%). The adults show a drop in false alarming between CS and CD items, whereas children show a slight trend in the opposite direction. This interaction between item type and age group was tested in two ways. First, pooling across items, the directed difference in false-alarm proportions to CS and CD items was reliably greater for adults than for children, \( t(30) = 3.11, p < .01 \). Second, pooling across Ss, each item was scored according to the difference in false-alarm rates to it for adults and children. These difference scores were reliably greater for the 11 CD items than for the nine CS items, \( t(18) = 2.13, p < .05 \). The differences in false alarming to CS and CD items can also be tested within each age group separately. Such tests reveal no significant differences for children; a significant difference for adults when pooling over items, \( t(19) = 2.86, p < .01 \), but not for adults when pooling over subjects and treating items as the random effect, \( t(18) = 1.37, p < .1 \). Inspection of false-alarm rates in table 1 reveals item-specific variability or inhomogeneity within the sets of CS and CD items; beyond simple “sampling error,” this variance presumably reflects the influence on sentence memory of such potent factors as word salience, presuppositional focus, etc. Although adults more than children remember and discriminate among sentences according to their conceptual meaning, table 1 suggests that there are further subdivisions within the overall CS and CD categories which may relate to recognition false alarms in a more uniform and lawful way.

**Comprehension Control Group**

The observed age differences in memory confusions may be thought of in two ways. Consider two sentences, A and B, which are imaginally similar and which adults judge to be conceptually different. Children may confuse A and B in recognition memory either because (a) they simply do not understand or even detect the difference in meaning between the two sentences even under optimal conditions of comparison, or (b) although able to comprehend the difference, children tend to forget more readily than adults that type of information that distinguishes A from B. The issue can be settled readily by testing children for comprehension of the conceptual distinctions in our CS and CD sentence pairs. To this end, an additional 10 children from Bing Nursery School (six girls and four boys, ages 5-0 to 5-7, median 5-5) were tested individually on two types of comprehension judgments. In each task Ss listened to pairs of taped-recorded sentences, the first sentence of each pair read in a male voice, the second in a female voice, and then immediately compared the two sentences according to a specified criterion. In the first task, they judged whether the sentence uttered by the female was “talking about the exact same thing” as that uttered by the male. They were to say “different” in case the second sentence of the pair meant anything different from the first. The tape for this first task contained our 20 sentence pairs rated earlier as being highly similar in the imagery they evoked (so-called imagery pairs); these were mixed in with 10 pairs sampled from the same sentence pairs used for the children in the earlier memory experiment. As expected, the children correctly categorized 98% of the same sentence pairs as identical, whereas they categorized 5% of the imagery pairs as expressing exactly the same meaning. So on an immediate test, they can discriminate between same pairs versus imagery pairs.

Now, this discrimination among sentences of an imagery pair may reflect little more than children’s ability to detect changes in wording of the sentences. To help decide this issue, they performed a second comprehension test. The Ss heard the 20 imagery sentence pairs again and this time were asked to decide whether the two sentences were “talking about the same thing or idea but with different words.” The distinction was illustrated with several examples. As one instance, the target sentence, “I ran across the room and touched the wall,” was said to have the same meaning as “I touched the wall after I ran across the room,” whereas it has a different meaning from “I ran across the room and touched the floor.” A second illustration was the target, “The floor was dirty because paper and ashes were all over it,” which was said to mean the same as “Paper and ashes all over the floor made it dirty” but to have a different meaning from “Paper and tin cans all over the floor made it dirty.” Following these instructions, the 20 pairs of imagery sentences were read twice for each S. If S’s second judgment for any pair differed from his first, the pair was read again and S was queried regarding his certainty. Three responses were possible: “same,” “different,” or “not sure.” In order for a response to be counted as “same” or
"different," S either had to justify his decision or resist attempts to get him to change it. The scores to be reported, then, are only for certain and/or rationalized decisions; excluding uncertain judgments causes change performance to be less than 50% on the two-alternative test.

To briefly summarize the results of this "gist judgment" task, the children's judgments agreed with the adults' 69% of the time for conceptually different pairs and agreed 79% for conceptually same pairs. That is, our children noted similarities and differences among meanings of sentence pairs much as our adult raters did. The difference between CD and CS pairs in amount of child-adult agreement was not significant, $t(18) = 1.10, p < .10$. One might expect that children's tendency to say that sentences A and B "mean the same thing" would correlate with how often they false alarm by saying "same" to B when A had been presented in the study list. But this Pearson correlation (over all 20 items) was only +.13 for the children, suggesting that memory confusions were being largely controlled by the imaginal similarity of the test sentence to the study sentence. This low correlation was indicated also by the earlier facts that children judge "same meaning" much more often for CS pairs than for CD pairs, but yet their false alarms in memory to these two classes of items are somewhat in the opposite order.

Discussion

The findings of Begg (1971) and Anderson (1972) with adults have been replicated. Moreover, memory confusions for adults occur more often to conceptually similar sentences than to conceptually different sentences, even when both lures evoke the same image as a study sentence. This implies that a sentence's internal representation which endures in an adult's memory probably contains conceptual distinctions over and above those represented in the sensory imagery the sentence evokes. The fact that adults do discriminate between same (78%) and conceptually similar sentences (47%) may be explained in terms of their auxiliary encoding of and memory for surface-related features of the study sentences (e.g., the focus or presuppositions of the surface assertions).

In full contrast to these results with adults, the children's equal acceptance of same sentences (66%) and conceptually different (65%) sentences suggests that the conceptual dimensions or features along which these sentence pairs differed are either not encoded from the original study sentence or are soon lost from the memory representation (or cannot be regenerated out of it). Our comprehension data suggest that the children's difficulty was not in initially understanding the conceptual distinctions in the imaginally similar sentences; rather, the difficulty lay in the children's rapid forgetting of these conceptual distinctions.

Our present interpretation of these data is somewhat as follows: A sentence is initially encoded along several dimensions including its surface phonological form, its semantic (conceptual) relations, and the referential imagery it arouses. For simplicity, these may be thought of as three distinct memory codes that are closely tied together. In our comprehension tasks involving only very short-term memory, the codes from the two sentences of a pair were readily available, and the two sentences could be discriminated at that time on the basis of their surface phonology, their conceptual semantics, or their aroused imagery, depending on S's judgment task. But, of course, forgetting—or loss of code specificity—occurs with the passage of time and/or interpolated events. The data suggest that the forgetting rate for the semantic-conceptual code is much faster for children than for adults. Since the children later have available in memory only the "imagery codes" to compare with the test sentence, they make many false alarms to imaginally similar lures, and these are not distinguished (as they are for adults) between conceptually similar and conceptually different lures.

A final comment is that the conceptual distinctions that the child forgets would appear to be those which he learns late and finds somewhat hard to comprehend—dimensions such as causal attribution, temporal precedence, intentionality, appearance versus reality, knowledge versus belief, etc. Many other developmental psychologists have remarked upon the slow pace of cognitive development regarding these conceptual, and relatively non-sensory, distinctions. Thus, the developmental patterns of memory confusion errors found here seem quite consistent with this correlative information regarding cognitive development.
References


