The abstraction of arbitrary categories

DUANE W. SMALL

LaSalle College, Philadelphia, Pennsylvania 19141

Bransford and Franks (1971) showed that people retain the ideas expressed in linguistic input, rather than retaining the individual sentences. The extent to which this phenomenon reflects organizational processes based on co-occurrence information was investigated by using Bransford and Franks' procedure but substituting arbitrarily chosen groups of words for sentences. The subjects gave higher recognition ratings to large new groups (analogous to complex sentences) than to small new groups (analogous to simpler ones), so long as the size of the group was in the range normally presented. They also showed the same ordering for old groups; but they rated old groups higher than new groups within each group size. Hierarchical recognition response structure can be established without an ideational basis, and it does not require complete absence from memory of original input items.

Bransford and Franks (1971) have described a situation in which responses on a sentence recognition task are almost entirely independent of whether the tested sentences have actually been heard before. Their acquisition list consisted of a series of sentences expressing, in repetitious and overlapping fashion, four complex ideas. For example, the idea *The warm breeze blowing from the sea stirred the heavy evening air* was presented in sentences such as *The breeze stirred the heavy evening air*, *The warm breeze stirred the evening air*, *The warm breeze blowing from the sea stirred the evening air*, *The breeze was blowing from the sea*, etc. Sentences from the four complex ideas were intermingled so that no two sentences from the same idea were adjacent. Subsequent recognition task performance showed highest positive responses to sentences which expressed the complete complex ideas, even when these sentences had never been presented. In general, the more components of an idea represented in a sentence, the more likely that sentence was to be responded to as "old," with almost no effect of actual previous presentation.

These results contrast sharply with the well established accuracy of recognition in most situations. The authors explain the phenomenon in terms of memory for ideas as opposed to memory for sentences. This follows from a natural assumption that any differences from traditional findings result from the use of natural language, ideational organization, or something of this nature.

Analysis of the paradigm used by Bransford and Franks reveals that it is novel not only in adding complex ideas in order to organize material, but also in demanding recognition for groups rather than for individual items. In the course of their acquisition list, Bransford and Franks presented every component of every idea; in general these were presented in identical

phrases. Therefore, subjects were not asked whether a particular component idea (or phrase) had been presented, but whether a particular grouping of two or three of these components had been heard before. Recognition memory is well known to be highly accurate for individual items, but recognition for particular groupings of familiar items has not previously been tested.

To determine what effect complex idea organization of material actually has on recognition, a baseline of performance on a similar task, but without such organization, is needed. Therefore, it was decided to replicate the Bransford and Franks method insofar as possible, but to substitute words for the component ideas in their sentences. Thus sentences containing more than one such component were replaced by a group of words, rather than by a single word.

METHOD

Three experiments used one basic technique, a variant of that used by Bransford and Franks. For each study, four groups of four ordered words were established, corresponding to the four four-component complex ideas used by Bransford and Franks. These four groups of words will be designated hereafter as categories (even though they do not generally form natural categories) to distinguish them from the subgroups used in the acquisition and recognition lists.

For each category, the set of all possible combinations of one, two, three, and four members was generated. The order of words in each combination followed the order of the original category. Acquisition and recognition lists were then generated by selecting appropriate combinations. Both acquisition and recognition lists were blocked in fours, so that each block contained one word group from each category. Order within blocks was random with the restriction that a block could not begin with a word group from the category that ended the previous block.

All materials were tape recorded. Word groups were read at the rate of one word per second. On acquisition lists, the last word of the group was followed 2 sec later by a click; 10 more seconds preceded the next group. On recognition lists, there was no click, and the intergroup interval was 10 sec.

The subjects were told that the experiment involved verbal behavior, but they were not informed of the tasks to be performed except as they arose. For the acquisition list, they

This research was conducted while the author was supported by Public Health Service Training Grant HD00151 from The National Institute of Child and Human Development, David S. Palermo, project director.

Table 1
Mean Recognition Ratings in Experiment I

Word Group	Rating
Category A	
Rock Mountain Hut Woods	-1.67
Rock Hut Woods	1.75
Rock Mountain	1.34
Mountain Hut	1.42
Mountain	-3.17
Woods	-1.00
Rock (old)	1.00
Category B	
Breeze Sea Air Evening	-3.67
Breeze Sea Air	.92
Breeze Sea	.92
Air Evening	-.17
Evening	-1.25
Air	-2.58
Sea (old)	2.92
Category C	
Ants Kitchen Jelly Table	-3.09
Ants Jelly Table	1.42
Kitchen Table	-.09
Ants Kitchen	-.92
Table	-.33
Ants	-1.92
Kitchen Jelly (old)	2.25
Category D	
Man Couch Story Newspaper	-2.75
Couch Story Newspaper	.00
Story Newspaper	-2.00
Man Couch	.42
Man	-2.00
Story	-1.42
Man Couch Story (old)	2.00

were instructed via tape to listen to each of the groups in the list, to wait for the click, and then to copy the group of words on a sheet of paper provided. This departs from Bransford and Franks' procedure, which involved a color-naming task after each acquisition-list sentence, followed by a content question (e.g., "did what?") about the sentence. Content questions were not appropriate for word groups, so the procedure used was that of a successful replication of Bransford and Franks (Cofer, 1973) in which the color-naming task was omitted and the subjects copied the sentences rather than answer questions about them.

After presentation of the acquisition list, the subjects were given a 5-min break. They were then given a mimeographed form for their recognition list responses, consisting of places to check "yes" and "no" and the numbers 1-5 for each of the response items. Taped instructions for this list told them for each group of words to check "yes" if they had heard it in the previous list, and to check "no" if they had not. Then, on a scale of 1 to 5, they were to rate their confidence in the correctness of their answer, and to check the appropriate number on the form. No items were to be skipped.

A subject's recognition score for an item consisted of an integer from -5 to 5, with 0 excluded. This was achieved by taking the confidence value, and assigning a negative sign if the subject reported that he had not heard the word group before. Thus, low scores represent high-confidence "new" responses. The scores were averaged across subjects for each item.

All the subjects were volunteers from introductory psychology classes at The Pennsylvania State University and they received course credit for their participation.

EXPERIMENT I

The first experiment sought simply to establish

whether the Bransford and Franks phenomenon could be extended to nonsentential groups of words. To this end, material was kept as close as possible to the original. The actual word groups were drawn from Bransford and Franks' sentences. Thus, the words *breeze*, *sea*, *air*, and *evening* replaced the component ideas in *The warm breeze blowing from the sea stirred the heavy evening air*. The other complex ideas from Bransford and Franks' Experiment I were replaced by *rock mountain hut woods*, *ants kitchen jelly table*, and *man couch story newspaper*.

Because the Bransford and Franks phenomenon has been shown to be sensitive to the order in which items appear, at least in the recognition list (Cofer, 1973), it seemed desirable to use an already successful order. Therefore a correspondence was established between each word in a category and a component of the complex idea it replaced. The order of lists used in an exact replication of Bransford and Franks' Experiment I (Cofer, 1973) was then copied, substituting words for corresponding components.

The resulting acquisition list consisted of six word groups from each of the four categories: two single words, two groups of two, and two groups of three. The recognition list consisted of the remaining sententially possible groups: the remaining singles, the two pairs and one triple corresponding to possible sentences, and the complete category. (Excluded pairs correspond to *The breeze was warm—the evening air was heavy* and *The breeze was blowing from the sea—the evening air was heavy*; the excluded triple corresponds to *The warm breeze was blowing from the sea—the evening air was heavy*; with similar exclusions for the other idea sets.) In addition, there were two singles, one pair, and one triple from the acquisition list, each from a different category.

There were two conditions, one in which the lists were presented in what was arbitrarily designated as the "original" order and one in which both lists were reversed. Six subjects served in each condition, and each condition was run in a single group session. Results were averaged across the two conditions.

Results

Mean recognition ratings for each group are presented in Table 1. Following Bransford and Franks' procedure, triples were compared only to pairs and singles that were present in the triple (i.e., *couch story newspaper* could be compared to *story newspaper* and to *story*, but not to *man couch* nor to *man*, since the triple does not contain *man*); and pairs were compared only to their component singles.

It was expected a priori that quadruples would not show the same high ratings as the corresponding sentences of Bransford and Franks because the obvious rhythmic structure of the lists encourages counting or an equivalent, and there were no four-rhythms in the acquisition list. Of the remaining possible comparisons, 18 were in the direction predicted by the assumption

that the Bransford and Franks phenomenon would extend to word groups, three were in the opposite direction, and there was one tie. Excluding comparisons involving old word groups, the figures are 13, 1, and 1. (Available statistical tests are not appropriate since the comparisons are not independent, but the expectation is equal numbers of differences in the two directions.) Of seven comparisons involving old groups, three had the old as the larger group. Of the remaining four, the larger, new group was rated higher in two cases and the smaller, old group was higher in two.

In general, the results conform to expectation that the Bransford and Franks phenomenon could be extended. The exception for quadruples was expected, as has been noted. There is some indication that presence in the acquisition list may be more important in the present situation than in that of Bransford and Franks: In all cases the "olds" were rated higher than any other group of the same size from any category. Four items are not sufficient to draw firm conclusions, and the comparisons do not meet Bransford and Franks' criteria for acceptable comparisons. The result is nonetheless suggestive, and the question will be dealt with again in Experiment III.

EXPERIMENT II

The second experiment sought to extend the phenomenon to associatively unrelated words. Sixteen concrete nouns were chosen from the first hundred words of Palermo and Jenkins' (1964) norms. Selection was random with the restriction that no pair of words have an associative frequency in either direction of more than nine for either males or females on the college norms. Once selected, the 16 words were randomly assigned to four categories of four words each. The categories selected were *king whiskey scissors citizen*, *stem music table sheep*, *fruit butterfly hand eagle*, and *child moon blossom thief*.

Once again words were assigned correspondences to component ideas. List structure and order were copied this time from an exact replication of Bransford and Franks' Experiment II (Cofer, 1973). The resulting acquisition list was identical in form to that of Experiment I. The recognition list included 30 items instead of 28, the four "old" groups being replaced by six "news": three cross-category quadruples and three cross-category triples. Again there were two conditions. In the first, the acquisition and recognition lists were presented in an "original" order. The order of the acquisition list was reversed in the second condition; the recognition list was presented in reversed halves (i.e., 15-1, 30-16). Fourteen subjects served in the first condition and were run in groups of 4 and 10; 15 served in the second, in groups of 8 and 7. Results were again averaged across conditions.

Results

The mean ratings for the word groups of this experiment are presented in Table 2. Once again ignoring

Table 2
Mean Recognition Ratings in Experiment II

Word Group	Rating
Category A	
King Whiskey Scissors Citizen	-4.21
King Scissors Citizen	2.69
King Whiskey	-.17
Whiskey Scissors	1.86
Whiskey	-1.89
Citizen	-.34
Category B	
Stem Music Table Sheep	-3.90
Stem Music Table	2.79
Stem Table	2.59
Table Sheep	-.03
Sheep	-2.59
Table	-.38
Category C	
Fruit Butterfly Hand Eagle	-3.59
Fruit Hand Eagle	.86
Butterfly Eagle	1.24
Fruit Butterfly	.55
Eagle	-2.34
Fruit	-3.45
Category D	
Child Moon Blossom Thief	-3.66
Moon Blossom Thief	3.10
Blossom Thief	1.72
Child Moon	1.59
Child	-1.79
Blossom	2.31
Cross-Category	
Stem Butterfly Scissors Blossom	-4.38
Fruit Music Scissors Thief	-4.79
Whiskey Moon Hand Sheep	-3.48
Fruit Moon Scissors	-.34
King Music Hand	.07
Fruit Table Thief	.41

the quadruples, there were 15 allowable comparisons, of which 14 were in the predicted direction. The cross-category quadruples were very low, as were cross-idea four-component sentences in the Bransford and Franks study. The effect was not dramatic here, as it was in Bransford and Franks, however, since the quadruples which did not cross categories were also low. The cross-category triples were all lower than the lowest within-category triple. However, they were considerably higher than many smaller groups, contrasting with corresponding results for Bransford and Franks' sentences. In general, it appears to be possible to extend Bransford and Franks' basic findings to ordering of recognition within new items, but cross-category items show some indication of acting differently.

EXPERIMENT III

The preceding experiments provide little evidence on the role of the old and the new in this recognition task. There is an indication in Experiment I that it is much greater with word groups than Bransford and Franks found it to be for sentences, but the data are sparse; and

Experiment II included no old groups in the recognition list. Also, as has been noted, the quadruples were not responded to as positively as one would expect from Bransford and Franks' results, perhaps because the rhythm of presentation excludes all quadruples.

The third experiment included both old and new groups in the recognition list in equal numbers, and for two of the categories the quadruples appeared in the acquisition list. The acquisition list consisted of half the possible singles, pairs, and triples from each category, plus the quadruples from half the categories. The recognition list consisted of the full set of sixty possible combinations. No correspondences were established between words and component ideas, so no pairs or triples were excluded as being sententially impossible.

The 16 words from Experiment II were randomly reassigned to three new category structures: (a) *blossom scissors citizen whiskey, thief sheep child moon, hand eagle table butterfly, music fruit stem king*; (b) *eagle stem whiskey music, butterfly child scissors fruit, king blossom music sheep, hand thief table citizen*; and (c) *stem table citizen hand, sheep blossom child eagle, thief moon scissors fruit, king music whiskey butterfly*. For each category organization, separate acquisition and recognition list orders were generated. Thus there were three category structures, each with its own order. Running the acquisition and recognition lists in forward and reverse orders as before yielded six conditions. Fifteen subjects served in each condition and they were run in groups ranging in size from individual subjects to, in one case, the full number for the condition.

Results

Recognition ratings for old and new word groups were tabulated separately, but ratings were combined across word groups of the same size. For new groups, the mean ratings were $-.41$, 1.50 , $.19$, and $-.90$ for quadruples, triples, pairs, and singles, respectively; for old groups the corresponding figures were 2.32 , 2.62 , 1.95 , and 1.03 . The general ordering observed for triples, pairs, and singles in the previous experiments still held, and it held for old as well as for new groups. Including two quadruples in the acquisition list did not serve to boost the ratings of new quadruples above that of new triples, although they are higher than in the earlier experiments. In fact, even for the old groups the rating is slightly lower for quadruples than for triples. In addition to placing triples rather than quadruples at the top of the recognition hierarchy, the present results contrast with Bransford and Franks' in showing significant differences between ratings for old and new groups at all group sizes. A modification of the Mann-Whitney U-test was used to test these differences. Each word group within a category structure was treated as a single observation, and mean ratings for old and new groups of a given size were entered as the scores for two independent conditions. The sum of ranks was taken in the usual

manner, the procedure was repeated for the other two category structures, and the resulting rank sums were added together. Taking advantage of the normal approximation to the Mann-Whitney statistic, the mean and variance of the combined sum were calculated and a z-score obtained. The normal curve function then provided approximate probability values. For word groups of Size 1 there were three replications with eight observations per condition (old and new) in each, $z = 2.97$, $p = .003$. There were 12 observations per condition in each replication for word pairs, $z = 5.57$, $p = .0002$. For triples, with again eight observations, $z = 3.79$, $p = .0002$. For quadruples, the normal approximation again yields significance, $z = 2.68$, $p = .008$; but since there were only two observations per condition in a replication, the exact probability was calculated as $p = .0092$.

DISCUSSION

The results of the three experiments may be briefly summarized. Among word groups which had not previously been presented, a recognition hierarchy was observed similar to that obtained by Bransford and Franks, differing in the failure of complete "categories" (quadruples) to appear at the top. Even the presence of occasional quadruples in the acquisition list was inadequate to raise quadruples to a position of primacy. Furthermore, while the hierarchical ordering was obtained and appeared to be similarly shaped for old and new word groups, there was a strong effect of prior presentation of specific groups. Bransford and Franks' finding for sentences (no differences between old and new at more complex levels of three- and four-component sentences) was not replicated with word groups.

The similarities and differences between the present results and those of Bransford and Franks (which provided the performance baseline needed) raise some questions about the kind of interpretation offered by the latter. Since the subjects produced a hierarchical structure on the basis of co-occurrence information alone, ideational organization cannot be a necessary condition for this aspect of the Bransford and Franks phenomenon. If the nearly total loss of old-new differentiation observed with sentences had been duplicated with word groups, it would have been possible simply to substitute co-occurrence for ideational integration as the organizing principle, and thus leave the rest of the interpretation unquestioned. (This assumes, of course, acceptance of a rhythm-based response bias as the reason for the failure of complete categories to stand at the top of the hierarchy.) Since this did not occur, additional problems must be considered.

According to Bransford and Franks, the hierarchical structure and the loss of old-new differentiation are two

reflections of a single memory characteristic. The memory system does not store input items directly; rather it abstracts elements from them which it uses to construct a single, organized representation of the information from several related items. Old-new information is lost because input items were not stored in their original form; and hierarchical structure emerges because the structure that is stored is better matched by complex than by simple recognition items. Thus the two phenomena occur together, hierarchical structure demonstrating what is stored, and loss of old-new differentiation clinching the argument by showing the absence of original input from memory.

Since, with arbitrary categories, the hierarchical structure and the presence of old-new information occur simultaneously, the argument is weakened. Perhaps organized memory and event memory coexist, with the subjects using the former when the latter fades or becomes inaccessible. However it is possible that the hierarchy is produced by a response strategy in the face of a confusing recognition task and that the strategy has little to do with what is stored. While other available evidence supports the claim that actual input strings are stored, there is little basis for choosing between hierarchy as response strategy and hierarchy as reflection of an additional memory representation. For example, recognition list position has been shown to play a role in performance: New complex sentences receive low ratings at the beginning and high ratings at the end of the recognition list; while old simple sentences (at least if last on the acquisition list) show the opposite effect (Cofer, 1973). While this opens the possibility of a role for the recognition task itself in producing the hierarchy (perhaps creating response bias), it is equally compatible with a model in which event information (actual input-string memory) fades with time, and/or with retroactive inhibition, to reveal an organized store. All that is clear is that event information does not disappear immediately. Only slightly more suggestive is a free

recall experiment (also reported by Cofer, 1973) in which 58% of the subjects' productions were classified as actually having been presented. Since in that experiment none of the nonpresented sentences which were produced contained all four components of a complex idea, there is a possibility that the hierarchical structure disappears as a function of response-mode shift. This would argue for response mechanisms and against an abstracted memory representation. However, the subjects were able to produce four-component sentences when instructed. This does not demonstrate that these sentences represent memory as opposed to on-the-spot construction, but it does leave open the possibility that they were in memory. Their not being produced could be the product of a bias against producing long sentences, just as the failure of complete arbitrary categories to appear at the top of the recognition hierarchy in the present experiments was presumably caused by a response bias against four-rhythms.

It is clear that, whatever causes hierarchical responding, co-occurrence is sufficient to induce it. Ideational integration need not be invoked, except possibly (if the assumption of response bias against four-rhythms is rejected) in a facilitative role. It is also clear that information about the actual input strings is stored and persists at least for a few minutes. Whether an abstracted, reorganized memory representation is also stored remains open to question, along with any other possible causes of the hierarchical structure of recognition responses.

REFERENCES

- Bransford, J. D. & Franks, J. J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, 2, 331-350.
 Cofer, C. N. Constructive processes in memory. *American Scientist*, 1973, 61, 537-543.
 Palermo, D. S. & Jenkins, J. J. *Word association norms: Grade school through college*. Minneapolis, University of Minnesota Press, 1964.

(Received for publication July 31, 1974,
 revision accepted October 15, 1974.)