ORGANIZATIONAL FACTORS IN FREE RECALL

GORDON H. BOWER
STANFORD UNIVERSITY

1. A colloquium speech delivered at several Psychology Departments during February and March 1968. Research reported here was supported by a grant, MH-13950, to the author from the National Institute of Mental Health.
In free recall, the subject is asked to produce items that belong to some class, and he can do this in any order he pleases. For example, I might ask you to recall the names of all the people you met yesterday, or to recall all the bird names you can think of. In experiments in the laboratory, the class of words to be produced are defined arbitrarily as those that the experimenter has just shown you in a list. The task is similar to that of recalling a shopping list, or a list of guests you've invited to a party, or a list of students in your class, and so forth.

Mandler & Tulving have pointed out that most such tasks involve a minimum of learning in the sense of storing new words but a maximum of learning in the sense of developing strategies for accessing those words already in storage which have been named as members of the list to be recalled. There are at least several distinct components to what is learned in such experiments. One thing that surely is learned is the ability to discriminate list membership — that is, to recognize a word that was mentioned on the list from another word that was not mentioned. In terms of basic processes, this discrimination of list membership is probably resolved, first, in terms of semantic classification information, where this is appropriate, and, second, in terms of a judgment of recency of experience which is always appropriate. An example of the first factor is the fact that we do not consider non-grocery items when we are trying to remember a grocery shopping list. An example of the second factor is my use of the phrases "first factor" and "second factor" to refer to the classification and recency information I mentioned in that order several sentences ago. By some means, items are tagged according to their arrival time; and although this information surely decays with time, it is one of the
sources of information enabling me to remember this week's shopping list as distinct from last week's.

So, to return to our prior thought, discrimination of list membership is surely one of the things that is manifest in free recall learning. The other thing that is learned is some strategy for accessing or retrieving the individual words of the list. For most cases I am familiar with, this accessing strategy usually reduces in the final analysis to the subject in some way providing himself with a cue which is associated with one or more words of the set to be recalled. Thus, in most cases, accessing would seem to operate by way of associations, in this case between retrieval cues typically provided by the subject and list words to which they are associated.

These retrieval cues can come in a variety of disguises, so let me enumerate just a few of the more obvious kinds to indicate in a general way how they work. One way by which subjects can improve their free recall is by using a known, ordered set of pegwords as access cues for the individual words of the list. This is the basis for a well-known mnemonic system whereby numbers are first associated with pegwords, such as one is a bun, two is a shoe, three is a tree, etcetera. These pegwords are then associated with corresponding words on any new list, thus effectively converting a free recall task into a paired-associate task. At the recall test, the subject can systematically generate the pegwords and these act as retrieval cues for the individual words of the list. We've studied this device a little bit, enough to know that it is enormously effective and will boost recall by a factor of 2 or 3 over subjects who are not using such peglists. We found too that it's not necessary to have a
unique one-to-one pairing between pegwords and list words; the person can
do nearly as well with fewer pegwords but more list-words being hooked onto
each one.

So that is one retrieval scheme. Another one that works on a si-
milar principle is to use the letters of the alphabet as cues, as Tulving
has shown. The alphabet is a well-known list, of course, and the subject
associates the first letter of each list word with the word itself, so that
a comes to mean apples, b means bananas, c means catfood, and so forth. Of
course, in recall, consideration of the letter b does not always lead one
uniquely to the word bananas. Rather it may lead to consideration of a re-
latively large number of b words, arranged perhaps in a hierarchy, and the
one of these which is output is that one which the executive routine re-
cognizes as having been on the list, performing this discrimination on the
basis of the word's recency. This mechanism doesn't always work success-
fully, of course; banana may be too low in the hierarchy to the letter b
to be considered at first, so it won't be recognized and output; but then,
that is one of the functions of repeated trials, to temporarily strengthen
the b-ball association so that it is considered for recognition and output
on the later trials.

Another scheme that works on a similar principle is the use of
superordinate categories that classify the list words. Thus, my shopping
list may be subdivided into sublists of similar things, like vegetables,
meats, bakery goods, and so on. This categorization can be carried through
several levels, so that a larger shopping list can be divided into sublists
according to groceries, dry goods and hardware items, with these simply
being more superordinate categories in a hierarchy. In either event, if
the person associates a category label to each word as he studies it, these category labels can then serve as accessing cues for recall of the words if these category labels are provided to him or if he can recall them at the time. This has been neatly demonstrated in experiments by Tulving 
& others. An elementary classroom demonstration of this is to read category labels to subjects, have them write down an association to each, and later ask them to recall their associates; you get about 50 to 60 percent recall. You then provide them with the original category labels, and you now get 100% recall of their associates. We've also run the reverse of this; unrelated words are read to the person which he classifies under some superordinate categories; when tested later for recall of the original words, the score is about 60% with no cues and about 97% when they're provided with the category associates they gave. All this means is that category labels are very effective pegwords that can be used as accessing cues because they are associated with the list words.

Another variety of accessing cue for a word is one or more other words in the list to be recalled. That is, by some means the output of word A has come to serve as a retrieval cue for word B, and word B may in turn access word C, and so on. The form of this relationship is most easily seen in the consecutive output of word-groups which have high normative associations even before appearing in the same list.

When lists of allegedly unrelated words are used, the exact groupings or clusterings of the words can not be predicted in advance; but Tulving's measures of subjective organization leave little doubt that subjects are nonetheless forming stable clusters of words on some basis even in this case. The basis for the groupings in these cases are largely
ideosyncratic and vague, determined sometimes by rather loosely defined categories that the person invents to include several words, and other times by more or less accidental contiguities of words in the input list or in the subjects output list. This simply shows that adults can be terribly inventive in solving ambiguous problems; in this case, classifying a list of words that the experimenter has carefully chosen so as to be unrelated. However, there is little in all of this subjective organization to suggest that simple association matrices will not handle most of it. That is, the memory list would be represented in theory by a matrix of interword associations of varying strengths. Contiguity of words in the input list or in the output list or common grouping of those words would temporarily increase the association between them. In outputing, the model might start with the last 2 or 3 items in short-term memory, and then follow associative chains or branches through the matrix, with output eventually stopping, for example, when associations to words already output are below some threshold or criterion.

If this account is correct, then by the end of free recall learning one should be able to show fairly strong interassociations amongst items in the list. Leo Postman has just told me about an experiment he's recently done which shows this effect in a very elegant way. He had his subjects learn a free recall list of 20 unrelated nouns. Thereafter, the subject learned a set of 10 paired associates which either involved new words or involved 10 arbitrary pairings of the 20 words from the free recall list. The important result was that subjects showed negative transfer in learning pairs involving words of the free recall list. In terms of interference theory, if the interitem associations de-
veloped in free recall are represented as A-B, then the arbitrary paired-associate list bears an A-B'r relation to the original free recall associations - that is, the same responses but paired with different stimuli. And the A-B, A-B'r paradigm is known to produce negative transfer. For example, in free recall the person may have come to cluster the words basket and desk together and the words house and flower together. But the later paired-associate task may require him to associate desk with flower, and basket with house, and his previous associations would interfere with this new learning.

We have completed an experiment which can be viewed in much the same way, although there are alternative interpretations. This experiment involved free recall of a list of 24 unrelated nouns. These were presented to the subject grouped into arbitrary sets of four words; for each four-tuple the subject was instructed to make up an elaborate mental image or picture which included the four objects interacting in some sensible way. Leaving aside the specific details of the imagery-mediation, the intent was to strengthen interassociations amongst the words comprising each four-tuple. There were 6 such 4-tuples, each studied for 12 seconds, or 3 seconds per word. This technique proved very successful as a means for producing interitem associations since practically all the output in recall could be described as clustered by means of short runs of 2 to 4 responses from each four-tuple.

The point of interest in this experiment concerned the effect of repeated study trials either when we kept the 4-tuple groupings the same from trial to trial, or randomly varied the word groupings from trial to trial. Thus, word A may be grouped with words B, C, D on trial 1, with
words E, H, Q on trial 2, and with words G, R, and V on trial 3. The results of this maneuver are shown in Figure 1. Recall in the two conditions is approximately the same on the first trial, as it should be, but the effect of subsequent trials is quite different in the two cases. Subjects with the same 4-tuple groupings improve rapidly, whereas those for whom the 4-tuple groupings are changed improve hardly at all in their total recall. In fact, the change in recall over the 3 trials for this group doesn't even approach statistical significance. Their poorer performance shows up both in a higher forgetting rate on words they once recalled, and a lower rate of recalling new words they hadn't previously recalled.

One can interpret this result in either of two ways, although in the final analysis they may be the same thing. One way, due to Mandler and Tulving, would suppose that recall is a function of the stability of the organization of the list words, and that our procedure of changing 4-tuples prevents the development of the stable organizations that are required for doing well at free recall. An alternate interpretation is that when word A is first associated with words B, C, D, and then with E, H, Q the conditions of an A-B, A-Br paradigm for negative transfer are being imposed upon the interword associations from trial to trial, and this is responsible for the fact that the recall performance shows very little improvement. It is not apparent to me that either of these views explains everything, but they are all I've thought of for this experiment. I do not see at the moment how one could experimentally separate them.

I would like to make a brief digression at this point to describe another experiment we've done on this business of the stability of organi-
zation and recall. This new experiment involves serially ordered recall rather than free recall, and it was done not with words but with the random strings of digits so popular in testing the memory span. The experiment used a paradigm introduced by Hebb several years ago in which, during an apparently conventional series of tests for digit span, a particular digit-sequence is repeated several times. Table I diagrams a sample series of trials, where the letters a, b, c, each stand for a particular permutation of the digits 1 to 9. The digit string labeled c here is being repeatedly given and recalled every third trial. Hebb found that recall of such a recurrent series improved over trials, which led him to certain conclusions regarding the involvement of long-term storage processes even in a task ostensibly involving only immediate memory.

We repeated this experiment more or less exactly but manipulated whether the organization of the recurrent string, c, was kept the same from trial to trial, or whether its organization was changed from trial to trial. What corresponds to organization in a string of digits? Prior work by Muller & Schuman, and speculations by Katona, lead to the belief that digit strings are organized in terms of grouping of subseries of digits, these groups being emphasized by vocal stress and pause patterns, or in the actual words used to denote digits in terms of units, tens, hundreds and thousands. At the bottom of this slide, I have illustrated just a few of the many ways by which a digit string can be organized into groups.

\[ (4 \ 7) \(1 \ 3 \ 5\)\(9\)\(6 \ 8 \ 2) \]

The first would be read "forty-seven, a hundred thirty-five, nine, six hundred eighty-two" whereas the second would be read "four hundred seventy-one, 
\[ (4 \ 7 \ 1)\(3\)\(5 \ 9 \ 6\)\(8 \ 2) \]
TABLE I

Series:  a b c d e c f g c h i c

c:  (47) (135) (9) (682)
c':  (471) (3) (596) (82)
three, five hundred ninety-six, eighty-two" and so on. If you prefer, you may think of these as different ways of parsing the syntactical structure of the digit string. With this as background, our experiment is easy to describe. Subjects had two different types of Hebb series, in one of which the recurrent string was always read with the same syntactical organization, and in the other of which the recurrent string was read each time with a new syntactical organization. Recall was always by writing down the proper string of digits. We expected a large improvement over trials in the former case, but relatively little in the latter case. The results are shown in Table II in terms of the mean number of errors per string, averaging together the first two versus the last two trials. The nonrecurrent or control items presented once increase slightly in errors as the series goes on. The recurrent string for which the same organization is repeated shows a quite significant drop in errors over trials. On the other hand, the recurrent string which has its organization altered every trial shows only a very small change over trials; in fact, one can not reject the null hypothesis of no improvement in this case.

In order to get a bigger effect, we're currently repeating this experiment with strings of 12 digits; these produce many more errors initially, so that a much larger learning effect can be shown. The results to date are shown in Figure 2. The results couldn't be prettier: the recurrent string that's repeated with the same organization shows a large progressive drop in errors over trials, whereas the recurrent string with altered organization shows no improvement. The error rate for the control strings is approximately at the same level as that of the string with changing organization.
TABLE II

Errors per String

<table>
<thead>
<tr>
<th>Trial</th>
<th>Trials</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2</td>
<td>3, 4</td>
<td>1.81</td>
</tr>
<tr>
<td>Same Org.</td>
<td>2.05</td>
<td>1.15</td>
</tr>
<tr>
<td>Changed Org.</td>
<td>1.94</td>
<td>1.87</td>
</tr>
</tbody>
</table>
In conclusion, these studies have shown a profound detrimental effect of changing the organization of the material upon a recall process which would otherwise improve with repetition. A rather direct interpretation would suppose that the organization of the material affects its coding, and this affects which atomic units are activated in the storage system; and furthermore, that repetition of a series that is coded in a different way simply references different locations in the storage system. Thus, the second experience is not making contact with or strengthening the trace of the first experience. The effect is similar to that which we experience with different syntactical parsings of an ambiguous sentence; the same string of words references quite different semantic structures depending upon the phrase-structure of the sentence. This interpretation can be specified further by supposing that the mind has a small number of serial-position slots into which it assigns or associates the successive groups of the string. Thus, the first group is associated with the first slot, the second group with the second slot, and so on. According to this view, when the string is regrouped differently, entirely different units are being assigned to the successive slots. Thus, considering each slot as a hypothetical stimulus, an A-B, A-C paradigm of negative transfer is implicit in the experiment with changing organization. This approach has certain testable implications; one example is that it expects improvement in, say, a constant third chunk even though the groupings around it in the string are being changed from trial to trial.

To return to our general summary on free recall, the present reconstruction of free recall performance is that by some means the words become grouped into several clusters of interassociated words, and that the
person cues off his recall of these clusters. The next few experiments I wish to report are concerned with this latter business, the methods by which the person can move from one cluster to another in recalling. In many ways, this is just the original question of how free recall occurs, except that now our units are large clusters rather than single words. There are two complementary factors that can be mentioned in regard to this cluster recall. The first factor, which Mandler has stressed in his writings, is the subject's memory span, which we suppose has a limited capacity of 5 to 7 or so units. Thus, if each cluster comes to be coded as a single unit, this brute memory span would permit the recall of 5 or so clusters. If the number of clusters exceeds this limit, then recall of more than this must be mediated by further associations - either by associations between elements of two clusters or by a mediating association to a superordinate category which includes both of the lower-level clusters. By either means, the number of higher-order units is thus reduced to approximately 5 or so units. Or so the argument runs.

In the next experiment, thinking along these lines we have tried to aid and abet the subject in moving from one cluster to the next. We used a rather long list, of 48 words, and these were presented in groups of 4 words at a time for integration into a visual image. We know that one trial of this produces very stable recall clusters, so the place where the subject is most likely to fail, when he does, is in not recalling any words of a given cluster. If we wish to prevent this from happening, then we should try to provide the person with some linkage from one cluster to the next in his memory. Figure 3 shows the way we tried to achieve this. It shows four 4-tupl groupings of letters, where the letters stand for words of the list. The
FIGURE 3

LINKED LIST

CONTROl LIST

A  D  G  J
B  C  E  F
H  I  K  L
A  D  G  J
circles have been drawn around the presented 4-tuples - A, B, C, D, then D, E, F, G, and so on. The structure of the groupings for the experimental list is such that the groupings are linked via common elements, in this case A, D, G, and J. The idea here was that the word common to two groups would enable the subject to move from one to the other in recall. Although the figure shows only four word groups, in the actual experiment there were 16 groups, the entire 16 linked in one continuous chain that looped back upon itself. These groups were presented in random temporal order within the trial, although the list had this logical structure. Examining this list, we find that the elements common to two groups are presented twice during the trial, whereas the unique elements like B, C, and E, F are presented once. We'll keep separate track of these double and single-presentation items, since any theory expects the doubled items to be better recalled.

Looking now at the list structure in the control condition, it shares with the linked list the fact that the same items are presented once or twice, except that there are no linking common elements between groups. Thus the subject is left to his own devices in getting from one group to the next in his memory.

These lists were run for 3 trials with results as shown in Table III. This gives recall percentages for twice-presented and once presented words for the two conditions over the 3 trials on each list. First there is the expected advantage of twice-presented over once-presented items in both conditions, and although recall of these in the Linked condition is slightly better on trials 1 and 2, this is an insignificant difference. The real difference between the two conditions shows up in the first trial recall for the once-presented items; the 42 percent recall in the Linked
## TABLE III

<table>
<thead>
<tr>
<th>Trials</th>
<th>Single Linked Control</th>
<th>Single Linked Control</th>
<th>Double Linked Control</th>
<th>Double Linked Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.42</td>
<td>.28</td>
<td>.69</td>
<td>.67</td>
</tr>
<tr>
<td>2</td>
<td>.72</td>
<td>.70</td>
<td>.87</td>
<td>.81</td>
</tr>
<tr>
<td>3</td>
<td>.84</td>
<td>.86</td>
<td>.94</td>
<td>.94</td>
</tr>
</tbody>
</table>
condition is significantly higher than the 28 percent recall in the Control condition.

I would attribute this to the better recall of clusters in the Linked than in the Control condition. It is a bit difficult to devise an appropriate measure of cluster-recall to compare the two conditions because of the double-functioning role of the twice-presented items in the Linked condition. One score we have devised simply considers the unique elements, BC, etc. as a cluster, and recall of either of these gives recall credit to the cluster. The results of this scoring are shown in Table IV, and it gives a very decided advantage to the Linked-list condition.

The recall advantage of the Linked condition is relatively short-lived, since the Control condition catches up after 2 or 3 trials. The technical problem is that this procedure of having subjects visually-image with sets of 4 words simply produces very fast learning. By trials 2 and 3, these subjects are getting out at least one word in practically all the 4-tuple groupings, and thus no advantage can be shown any longer for the Linked condition. Obviously, if one wants to show a longer persisting advantage for the Linked condition, one has to use even more word-groups than the 16 we had in this experiment.

To return to the main point, I was talking about the means by which the subject can move from one cluster to the next in free recall, and have now shown that this can be achieved by common linking elements. It would probably also work with highly associated pairs of words included in the two groups of words, but we haven't shown this yet. The other method of improving recall in this sense is by the use of superordinate categories which subsume or include members of two or more lower-level categories. The
<table>
<thead>
<tr>
<th>Trial</th>
<th>Linked</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.53</td>
<td>.32</td>
</tr>
<tr>
<td>2</td>
<td>.82</td>
<td>.74</td>
</tr>
<tr>
<td>3</td>
<td>.86</td>
<td>.88</td>
</tr>
</tbody>
</table>
superordinate category then serves as a retrieval cue, as a guideline which
the person can use to search for list words within the restricted set iden-
tified by that category. If one imagines embedding this process within it-
self several times, having general categories that include subordinate ca-
tegories, which themselves include still lower-level categories, one arrives
at a hierarchical tree of categories. And if our reasoning about grouping
and retrieval cues for recall is correct, then presentation of list words
in some kind of natural hierarchy ought to be an extremely compelling and
easy way to learn those words for free recall. In his writings, Mandler has
stressed the efficiency of such hierarchical classification schemes for fa-
cilitating free recall, but to my knowledge no research has been done on it.
As Miller, Galanter & Pribram have noted in their book, a lot of our plans
about daily living are arranged hierarchically. If I think, for example,
about the plan for one of my typical Saturdays, it is organized into sub-
plans, for example, to do some garden work, to have some fun-activity with
my kids, and to go shopping for my wife. The shopping is further divided
into sublists of things to get at the hardware store and grocery store; and
the grocery list is further divided into sublists of vegetables, meats, baked
goods, and so on. This larger plan then organizes and controls the sequence
of actions I might go through in getting through a typical Saturday after-
noon.

We've done a little work on this kind of hierarchical organiza-
tion in free recall. It's difficult to know exactly where to start to get
a hold on the thing, but our preliminary efforts may prove interesting to
you.
First of all, in these experiments we have used word lists which appear to us to have a natural hierarchical organization, and for our experimental subjects we have specified this organization completely rather than letting them hit upon it themselves. We have used specially constructed word lists which embody hierarchical schemes. An example of one of our word-hierarchies is that for Minerals, shown in Figure 4. Minerals include metals and stones, stones are classified as precious or masonry stones, and limestone, granite, marble and slate are instances of masonry stones. I've labeled the levels or depth of the hierarchical tree by the numerals 1, 2, 3, 4. We've made up about 10 of these word hierarchies which we've used in our research.

In our first experiment, we simply compared the effect on recall of structuring or randomizing the organization of the lists. Subjects learned 4 different hierarchies at once, each hierarchy containing about 28 words, so that the total number of words in the list was 112. For one group, the 4 conceptual hierarchies were presented in completely organized fashion like the Minerals list. One hierarchy was shown in this tree form on each of 4 cards, the hierarchy being studied for about 56 seconds, or 2 seconds per word. After studying the 4 hierarchies, the subject then recalled as many of the 112 words as he could in any order. The subjects in the Random condition had the same set of 112 words, except the words from the 4 conceptual hierarchies were randomly intermixed on a given study card. They studied 4 cards, each laid out in this tree form, but the words within a given tree didn't fit together conceptually. The randomization was restricted within levels; that is, a level-3 word in this hierarchy always appeared as a level-3 word, but the words above and below it might come from conceptual
FIGURE 4

MINERALS

METALS

ALLOYS

COMMON

RARE

PRECIOUS

MASSONY

STONES

Limestone
Granite
Marble
Slate
Sapphire
Emerald
Diamond
Ruby

Aluminum
Copper
Lead
Iron

Platinum
Silver
Gold
categories corresponding to, say, types of instruments, articles of clothing, and parts of the body. By analogy to traditional studies, the variable we're manipulating here is most analogous to that of blocked vs. random presentation of categorized word lists. Blocking means that all instances of a given category are presented in adjacent input positions, whereas random means that they occupy random positions in the input list. In those traditional studies, blocking typically has a relatively small effect, improving recall by perhaps 5 to 8 percent, although in some of Cohen's more recent and extensive work there has been no effect at all on total output.

By way of contrast, the effect of this hierarchical blocking of the input words is huge. The results are shown in Figure 5, giving mean words recalled over two trials for the Organized and Random conditions. The Organized subjects recall 2 to 3 times as much as subjects in the Random condition. Moreover, the nature of the recall protocols is quite different in the two cases. Subjects in the Organized conditions often give back the list structured in tree-form much as they saw it, leaving blanks where they can't get a word. On the other hand, subjects in the Random condition tend to alternate between writing down words in the spatial tree-form presented to them versus listing clusters of categorically associated words in a column, essentially showing a conflict between the crazy spatial organization we imposed on their input list and the categorical organization existing in their semantic memory.

Compared to the traditional variable of Blocked vs. Random presentation then, this hierarchical blocking has had an enormously larger effect. I suspect that this is due to our method of complete presentation, which permits the subject to notice and exploit more organization in the
input than with the traditional memory-drum method of showing one word at a time. Moreover, the traditional studies have not studied hierarchical organization, nor required the category names themselves to be recalled. The important factor in terms of our analysis is that recall of a category name serves as a retrieval cue for getting out most of the words included within that category. And of course this advantage for the Organized condition just snowballs as the subject recalls down through successive levels of the tree.

We have also checked out these two conditions on recognition tests for memory. Although this hierarchical arrangement of words is efficient for generating retrieval cues in recall, one might suppose that this advantage would be absent if memory for list-membership were tested by the recognition method. This same experiment was therefore run with recognition tests. After being shown the 112 list words, the subject was handed a sheet of paper on which were printed the 112 list words plus 112 distractors, all being mixed together in random order. The distractors consisted half of unrelated words and half of words that belonged to the same conceptual categories as did the list words, but which had not appeared on the list. The subject was told to check those words which he thought had been on the list.

The results for two trials of this are shown in Table V. For comparison, I've also given the recall percentages from the previous slide. Recognition is scored in terms of the probability that words of a given type were checked by the group of subjects. Thus, the second line reports hit rates, and the third and fourth lines report false-alarm rates for related and unrelated distractors, respectively.
### Table V

<table>
<thead>
<tr>
<th></th>
<th>Organized $T_1$</th>
<th>Organized $T_2$</th>
<th>Random $T_1$</th>
<th>Random $T_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>.58</td>
<td>.83</td>
<td>.21</td>
<td>.44</td>
</tr>
<tr>
<td>Recog. + For Old</td>
<td>.80</td>
<td>.92</td>
<td>.60</td>
<td>.80</td>
</tr>
<tr>
<td>Related New</td>
<td>.07</td>
<td>.08</td>
<td>.09</td>
<td>.19</td>
</tr>
<tr>
<td>Unrelated New</td>
<td>0</td>
<td>0</td>
<td>.02</td>
<td>.07</td>
</tr>
<tr>
<td>$d^r$</td>
<td>2.28</td>
<td>2.82</td>
<td>1.55</td>
<td>1.70</td>
</tr>
</tbody>
</table>
The salient features of these data are as follows: first, that recognition hit rate exceeds recall probabilities; second, that related distractors are checked more often than are unrelated distractors; and finally, that subjects in the Organized condition perform better than those in the Random condition in recognition as well as in recall; they have a higher hit rate and a lower false alarm rate, and these are reflected in the d' transforms of these two percentages. Assuming that recognition tests measure only the amount of information stored, and do so independently of retrieval processes, then this result means that more information is stored about words in an organized list than about the same words in an unorganized list. It would seem reasonable to suppose that item recognition here involves not only recency discrimination but is also affected by the structure and number of associations converging upon the test item from other verbal units which have been recently strengthened. This at least is a view which also explains the higher false alarm rates on the associatively related distractors.

The next experiment I want to talk about was concerned with a different question, namely, whether by judicious use of the principle of category-cueing, we could improve recall of a first list by interpolated learning of a second list. The literature abounds with instances of the opposite effect, whereby interpolated learning of a second set of material produces unlearning and retroactive interference in recall of a first set of material. This has been shown too for free recall with unrelated lists of words, so it is not idle to search for conditions which produce retroactive facilitation. This we tried to do by using these word hierarchies. The design and results of this experiment are shown in Table VI. We used two hierarchies, A and
TABLE VI

<table>
<thead>
<tr>
<th></th>
<th>Second-List</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Relevant</td>
<td>Irrelevant</td>
</tr>
<tr>
<td>1. Learn $A_4^+B_4^-$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial 1</td>
<td>27.8</td>
<td>24.2</td>
</tr>
<tr>
<td>Trial 2</td>
<td>35.6</td>
<td>34.7</td>
</tr>
<tr>
<td>2. Second List</td>
<td>$A^+B^-$</td>
<td>$C^+D^-$</td>
</tr>
<tr>
<td>3. Recall All</td>
<td>44.4</td>
<td>34.2</td>
</tr>
<tr>
<td>4. Rel. Probe</td>
<td>--</td>
<td>41.1</td>
</tr>
</tbody>
</table>
B. The first list consisted of the 48 level-4 words of these two hierarchies. After two trials on this, subjects went on to a second list, which consisted either of relevant supraordinate categories, the levels 1, 2, 3 of the A and B hierarchies, or irrelevant categories, levels 1 to 3 of hierarchies C and D. These hierarchies were presented in organized tree-form until the subject recalled them perfectly, which usually took 2 or 3 trials. At the end of this interpolated learning, subjects were then asked to recall everything they could from both the first and second lists together. We see in this case that the group recalling relevant supraordinate categories now improves in their recall of the 48 level-4 words, going from an average of 35 up to 44 words recalled; the controls with irrelevant interpolated categories show little change in recall. [Table VI]

The obvious interpretation of this result is that recall of the supraordinate words serve as effective retrieval cues to help subjects recall more level-4 words than they were able to do before, although they hadn't studied the level-4 words in the context of these supraordinate words. That this is all that is involved here is shown by the probed recall in the control condition. After their total recall, the controls were given the level 1 to 3 words of hierarchies A and B and told to recall again the A_4 and B_4 words, their first-list words, in the context of these relevant supraordinate words. Their recall now improved to about the same level as the first group which had the relevant interpolated categories. Thus, whether the supraordinate categories are recalled from memory or are provided directly to the subject, in either event they seem to act as equally efficient retrieval cues.

An incidental finding of no less importance in this experiment comes from the control condition, namely, that there is very little loss in
their first-list recall by interpolated learning of an organized list of words belonging to categories differing from the first. This suggests that the amount of forgetting of one set of words produced by learning of a second set is a function of the degree to which the two sets become organized into distinctly different associative structures. We have another experiment that bears directly on this point. It concerns the effect of thematic organization upon recall of serially-ordered sets of unrelated words. The organization in this case required the subject to generate a narrative story, woven around the critical words to be remembered. The details were as follows: the subject was handed a list of 10 unrelated nouns, and told to make up some meaningful story that used these words in precisely the order written. They were given as long as they needed to achieve this, typically taking 1 to 2 minutes. The time taken was recorded and the same amount of time was given to a yoked control subject who was merely told to study and learn the words in the order written.

An immediate test for recall showed that both subjects were essentially perfect, so there was no difficulty there. The difficulty stemmed from the fact that 12 different sets of 10 words were put into the subject in this manner within an hour. At the end of the hour, recall of all 12 lists was tested by cueing the subject with the first word of each list and asking him to recall the remaining 9 words of the list in their correct order. The results of doing this are shown in Figure 6, which shows percentages of words in each list recalled in their correct serial order. The narrative condition is 6 to 7 times better than the controls; the median percentages are 93 versus 14 percent, and there's no overlap among the groups on any list. The same result holds if one simply counts all words recalled regardless of seri-
al order or list in which they're recalled. The items simply become unavailable as the controls learn successive interfering lists, whereas this is prevented by the thematic organization imposed upon the material by subjects in the narrative condition.

A few examples of stories given by subjects are shown in Figure 7, where the 10 critical words are capitalized. These have a sort of "stream of consciousness" sense to them; and while they're obviously not in line for a literary prize, they aren't bad solutions to the task of connecting in order 10 unrelated nouns chosen at random.

I think the effect here is due to thematic organization; that is, the words are woven into sentences, and the sentences related to one another by a theme. Themes from successive lists have little in common, are kept distinct from one another in memory, and so do not interfere with one another, and so are easily recalled. And recall of the theme is sufficient for the person to begin reconstructing the sentences and then the critical words. The operative word here is reconstruction; the subject doesn't remember the word-lists directly, but rather remembers the organizational plan he constructed, and that enables him to generate, to reconstruct, the critical words.

One of the baffling features of this performance is the subject's ability to discriminate in recall between the critical words and all the context words he used to string these together; they make practically no intrusion errors, although I can't understand why they don't make some since both critical and noncritical words have been recently primed in their memory. In general, however, the memory organization of the word lists here is similar to that of an actor who's learned his lines for a play: the play is hierarchically organized into three acts, those into the scenes in which he appears, the
FIGURE 7

S1. a LUMBERJACK DARTed out of a forest, SKATEd around a HEDGE past a COLONY of DUCKs. He tripped on some FURNITURE, tearing his STOCKING while hastening toward the PILLOW where his MISTRESS lay.

S2. a LUMBERJACK prepares DARTS and uses his SKATEs to cut down the HEDGES in order to clear the COLONY of DUCKs. He also sells FURNITURE so he can buy STOCKINGS and a PILLOW for his MISTRESS.
scenes into subplots, themes, or episodes, and those are related in turn to the specific cues for his actions and speeches.

ENDING

It's getting time that I be drawing this talk to a close. The view that we've been examining about free recall is, in simplified overview, that subjects form higher-order units or clusters of words that go together, and then learn some retrieval scheme for accessing these word-clusters in recall. I've talked about several types of retrieval schemes, one of the most efficient of which is a hierarchical arrangement of categories, but have argued that the basic ingredient of most of these schemes consist of associations between the retrieval cues and the words to be recalled. I have illustrated some of the ways by which words in a free-recall list come to be grouped together into higher-order units, and have argued that one effect of repeated trials is to provide more opportunities for these subjective clusters to increase in size and in their stability. I've shown too that if these word-clusters are prevented from being strengthened, then there is little or no benefit from repetition.

None of these are particularly novel views, since a sympathetic reading of Mandler and Tulving would find that they've said more or less the same thing, perhaps with less emphasis on associationistic interpretations than I've given. The aim of this research may be viewed primarily as an extension of their ideas into somewhat different experimental paradigms, probing the heuristic value of their ideas. As I said at the outset, these experiments were instigated by my reading Mandler's paper, and I think we've made some interesting openings for research. In any event, the research has been exciting for my research group, and we continue to have more ideas that
we can possibly follow up. And of most importance, I think, the research has been fun, and I can think of no better justification for working on anything in particular in psychology. I'm finished now, and I thank you for listening so patiently.