Representation of Order Information: An Analysis of Grouping Effects in Short-Term Memory

Don McNicol and Andrew Heathcote
University of Tasmania, Hobart, Tasmania, Australia

SUMMARY

Presenting short sequences of items in temporal or spatial groups is known to improve recall of their order. Theories about this effect propose either that groups of items are represented in a hierarchical structure in which the positions of items in a group are nested under codes denoting the groups themselves, or in a matrix structure where each item is directly tagged for its group and position. In a matrix system, position codes are independent of group codes, and the retrieval of a code near the bottom of a hierarchy depends on the successful retrieval of the codes above it. Experiments 1–4 tested these dependence relationships with a probed recall procedure in which subjects were presented with a grouped sequence of items and were then required to recall the position and group of one of the items. This technique provided information about how well subjects correctly recalled both the group and position, the group only, the position only, and neither the group nor the position of an item. When the items in a group were letters, digits, or musical notes, the data conformed to a hierarchical structure. When the nonalphanumeric characters #, $, %, &, ., +, @, T, <, were used, a matrix structure emerged.

Experiments 5–7 required subjects to judge the dissimilarity of two sequences of grouped items, the second of which (the variation) was a reordering of the first (the original). The variation was made by reordering the groups in the original, reordering positions within groups, or reordering both groups and positions. In a matrix system, the dissimilarity of a variation that involves reordering of groups and positions will be greater than the dissimilarity of one that involves the reordering of only one of these attributes. In a hierarchical system, the dissimilarity of a variation that is reordered in both respects will be equal to one or other of the single types of reordering, depending on the exact nature of the hierarchy. When the members of a group were able to be encoded as single verbal units, the data supported a hierarchical system. When this was not possible, a matrix system fitted the data best.

We conclude that there is no general code for representing the order of grouped sequences. Grouping effects sometimes arise from verbal strategies that involve the recoding of a group's items into a single higher order unit. Such strategies produce hierarchical structures. Grouped sequences of musical notes also produce hierarchical structures, but not necessarily for the same reasons as for groups of verbal items. In other circumstances, items appear to be coded as elements in a spatial array, in which one dimension reflects the position of an item in its group and the other, the position of the group in the sequence. The results are thus more compatible with a theory that postulates a number of specific subsystems in short term memory, each with its own format for preserving order, than with one that assumes a generalized order code.

The immediate recall of temporally presented sequences is improved if the members of the sequence are divided into subgroups by pauses, stresses, or marker signals. This improvement in recall comes mainly from a reduction of order errors (Ryan, 1969a, 1969b). Order errors in grouped sequences also differ by type as well as frequency from those in ungrouped sequences. In the ungrouped case, adjacent items are most likely to have their serial positions confused (Aaronson, 1968; Ryan, 1969b). In the grouped case, items that occupy equivalent positions in different groups are also likely to be confused (Ryan, 1969b; Wickelgren, 1964, 1967). Three classes of model have been proposed to account for grouping effects. These are tagging models, associative models, and recoding models.

Tagging Models

One interpretation of grouping effects arises from the idea that serial positions are represented in memory as points on a line, with adjacent positions being close together, and nonadjacent positions being further apart. This dimensional model has been used on several occasions to explain the pattern of order errors in ungrouped sequences (Crossman, 1955; McNicol, 1975;
The dimensional model's extension to the grouped case is obvious; serial positions can be treated as points on a plane, with one dimension representing an item's position within its group and the other dimension, the groups position within the message. Thus an item can be close to one within its group, and to another that occupies the same position in another group.

The dimensional model assumes that information about serial positions is preserved by assigning position tags to items. For ungrouped sequences, a single tag shows the item's position in the message; for grouped sequences, two tags are needed. Errors in recall of an item's serial position occur because the values of the tags are subject to random perturbations that make it likely that an item's position will be confused with that of a neighbor in its own group or with that of an item in the same position in a neighboring group. Estes' (1972) and Lee and Estes' (1981) reverberatory process, by which the serial order of sequences is held in short-term memory, is an example of this approach. Their model comprises a hierarchical system in which information about the positions of items within groups is nested below information about the positions of groups within the sequence. At the lower level of the hierarchy, each item in a group is held in a delay loop that recycles the item after a period of time has elapsed. All items have the same average recycle time, and the delays before the recurrence of items are lagged, so that each item reappears in the order in which it originally occurred in the group. At the next level up in the hierarchy, the groups are represented by special coding elements that are also held in delay loops and are recycled in the same manner as the items. The delays in the loops thus act as tags that denote each item's position in its group and its group's position in the sequence. If the recycle times of items within groups and groups within the sequence are constant, the serial positions of the items in the sequence will be correctly preserved. Order errors arise because of random variations in the recycle times of items and groups. If a perturbation of the coding elements for the first two groups of a sequence results in the element for the second group preceding that for the first, then the items in the sequence will be recalled in the wrong group but in correct positions within groups. If perturbations affect the sequencing of within-group coding elements, then within-group transpositions will be observed. These predicted error patterns will resemble those observed in the previously cited studies.

Non-directional and Directional Associations

In Murdock's (1983) distributed memory model, the order of the sequence ABCD is preserved by a set of overlapping associations A*B, B*C, C*D, between adjacent items. The associations are non-directional and are formed by performing the mathematical operation of convolution on vectors of features representing the items. These vectors are then added together to form the total memory vector for order information, A*B + B*C + C*D. The serial position of any item can be retrieved from this vector.

Pike's (1984) matrix distributed memory system proposes a number of means by which associations between vectors of features representing the items could be used to preserve their order. However, in this case the mathematical operation used to form the association is the multiplication of vectors representing items to be associated. For example, the items in the sequence ABCD may be associated by forming the matrix ab'c'd", where a is a row vector of features representing item A; b' is a column vector representing item B; c' is a block vector representing item C, etc. This matrix preserves directional associations that record that A preceded B, B preceded C, and C preceded D in the sequence.

According to Pike, the grouped sequence AB–CD could be represented by ab' + cd", the sum of two matrices, each preserving associations between items within a group. However, the resulting memory trace needs some additional information in order to preserve the positions of the groups in the sequence. The manner in which this could be achieved and a suggestion as to how Murdock's model might be extended to cover grouping effects, will be discussed after coding element models have been described in the following section.

Coding Elements

Murdock’s (1974, 1976) nesting model proposes that items are not associated with one another but with coding elements. The sequence ABCD is coded as

\[ ** + X_4(D + X_3(C + X_2(B + X_1(A + *)))) \], \hspace{1cm} (1)

where X_1, X_2, X_3, and X_4 are coding elements, and * and ** are special items denoting the beginning and the end of the list, respectively. The grouped sequence AB–CD is encoded as

\[ [X_4(** + X_3(B + X_2(A + *)))][Y_3(** + Y_2(D + Y_1(C + *)))]. \hspace{1cm} \ldots [** + Z_4(Y_3 + Z_1(X_3 + *))]. \hspace{1cm} (2)\]

Here the items within each group are first associated with their own coding elements, X_1, X_2, and X_3 and Y_1, Y_2, and Y_3, respectively, thus coding within-group serial positions. The positions of the two groups are then coded by associating X_3 and Y_3 with the group coding elements Z_1 and Z_2.

A similar scheme has been proposed by Estes (1972) for the coding of serial position information in long-term memory, but the idea could equally well be applied to short-term memory. For the grouped sequence AB–CD the items in each group are associated with a coding element denoting the group—namely, X*A, X*B, Y*C, Y*D. The coding elements themselves are associated with a higher order element denoting the sequence itself—namely, Z*X, Z*Y. Within- and between-group order is
preserved by a set of inhibitory connections so that when the
group AB is being retrieved, the link A ---< B inhibits the
response B until A has been made. Similarly the link X ---< Y
inhibits the group Y until the group X has been retrieved.

Both coding element models adequately preserve within- and
between-group serial positions and suggest a means by which
Murdock's distributed memory model might be extended to
cover grouping effects. This is done by including two coding
elements X and Y, which denote the groups. Thus the coded
representation of the grouped sequence AB-CD is the sum of
the associations X*A.B + Y*C*D. To ensure that each group
is retrieved in its correct order, it is necessary to assume that the
coding element X contains contextual information that identifies
the start of the sequence, and that Y contains contextual in-
formation identifying the end of the sequence. If the sequence con-
tains more than two groups, it would be necessary to include
some additional information that explicitly preserves associations
between the groups. Thus for the grouped sequence AB-CD-
EF, memory may contain the following associations:

\[ X*A.B + Y*C.D + Z*E*F + X*Y + Y*Z. \]  

where A, B, C, D, and E are collections of features representing
the items, and X, Y, and Z are collections of features representing
the groups.

Coding elements can also be used to preserve the correct order of
groups in Pike's model. However, as this model generates di-
rectional associations, all that needs to be done is to assume that
the composite memory trace is the sum \( x*a*b' + y*d' + x*y \), where
\( x \) and \( y \) are vectors of features representing the groups, and \( a' \),
\( b' \), \( c' \), and \( d' \) are vectors of features denoting the items. The
positions of the groups in the sequence are preserved by the
directional association \( x*y' \).

All of these models could be thought of as hierarchical ar-
rangements in which within-group serial positions are nested
beneath between-group serial positions. In principle they could
account for many of the errors reported in recall of grouped
sequences due to loss of information about between-group and
within-group coding elements or about the associations between
items in a group.

Recoding Models

The models reviewed so far assume a distinction between
components of the memory trace that represent the items in a
sequence and components that represent their order or serial
positions. The serial positions of items may also be preserved by
using a code book or a set of rules that recodes the separate items
in a group into a single item. The same code book or rules can
be used to decompose the coded item into its constituent items
in their correct serial positions. However, the coded item does
not contain order information as such—only features that denote
the item itself. Miller's (1956) "chunking" strategy, in which
groups of three binary numbers are recoded as a single octal
number, is an example of how order information can be preserved
as a recoded item. Bower and Winzenz's (1969) phoneme chain-
ing hypothesis, by which the grouped sequence 23-17-94 is coded
as "twenty-three, seventeen, ninety-four," is another. Johnson's
(1972) theory of opaque containers is a more general statement of
a recoding model. Each group of items in a sequence is rep-
resented by a single abstract high-level code whose contents (the
items themselves) can only be retrieved by performing the ap-
propriate decoding operation. Like the coding element models
described, this class of model is hierarchical, as the items and
their positions within their group can only be retrieved via the
high-level codes. This suggested dependency of within-group po-
position information on the successful retrieval of the group codes
is an important feature of this class of model, as will be shown
in the next section.

Code Dependence and Code Independence

A number of the previously mentioned models, such as Estes' short- and long-term memory models and Johnson's opaque
container model, have been described as hierarchical. For grouped
sequences, the topmost nodes of the hierarchy correspond to
coding elements for groups and nested below them are nodes
denoting positions within a group. The Murdock nesting model
can also be represented as a hierarchical structure. Its topmost
nodes are the coding elements for groups, and nested below them
are right-branching trees, the nodes of which are the coding ele-
ments that preserve within-group positions. These trees have an
arrangement similar to that of a push-down stack, with the last
item in the group available at the top of the tree.

Figure 1 compares a hierarchical structure for representing a
grouped sequence with that suggested by the dimensional model.
Broadbent (1981) calls this a matrix structure (not to be confused
with Pike's [1984] matrix distributed memory model). The dif-
fERENCE between the matrix and the hierarchy is that for the for-
mer, group and position coding elements are independent of one
another, and loss of one does not imply loss of the other. We
refer to schemes of this type as examples of the independence
model. In the hierarchical model, some codes are accessible only
via other codes. If it is assumed that information is retrieved by
starting at the top of the hierarchy in Figure 1, the group codes
may be considered as containers that hold the items and their
position codes. Loss of a container will make its position codes
unavailable. We will refer to this as the higher order code model.
Alternatively, if retrieval is assumed to start from the bottom of
the hierarchy, loss of an item's position code will deny access to
the group code above it. We will refer to this as the item search
model. The distinction between systems in which codes are in-
dependent of one another, or share various kinds of dependency,
follows that of Cumming and Coltheart (1969) in their study of
the relation between item and position information in visual
short-term memory. We will show, as they have shown, that the
models can be evaluated by goodness-of-fit tests applied to data
from probed recall tasks.

Models for recall of grouped sequences can be classified either
as cases of the independence model, or as hierarchical, in that
they imply code dependencies. The dimensional model, as stated
here, has the property of code independence, as does the asso-
ciative model in Formula 3. The opaque container model clearly
fits the case of a hierarchy that is accessed via grouping codes.
However, some models, called hierarchical by their authors, have
the property of code independence. The reverberatory model, as
presented by Estes (1972), and Lee and Estes (1981), is code
independent, because perturbations in the timing loops that hold
the group coding elements do not affect the processes that main-
tain the order of items within groups, or vice versa. However, one modification would give it hierarchical properties as well. This would be to assume that the loop responsible for circulating the coding element for a group can sometimes stop completely, thus denying access to the loops that are attached to it and that are responsible for preserving the serial positions of items within the group.

The nesting model may be code dependent or independent. Murdock (1974) describes the unpacking of a coded sequence as a stochastic process with a higher probability of error if the previous step was incorrect than if it was correct. If the probability of correctly unpacking the position information is zero when there was an error in retrieving grouping information, but some greater value when there was no error, then the model behaves as a hierarchy. In theory the model will show code independence if the probability of correctly unpacking the position information is the same, whether or not the grouping information was correctly unpacked. However, we suspect that because the unpacking operation is strictly hierarchical, this limiting case is nonsensical.

Model Equivalences

If the data are good enough there is no difficulty in distinguishing the independence model from either of the hierarchies. There is a problem in distinguishing between hierarchical systems, however.

The hierarchical system of Figure 1 has nodes denoting position of an item within its group nested below those denoting position of the group within the sequence. It is equally possible to conceive of a system in which the group nodes are nested beneath position nodes. It has been suggested that there may be two ways of accessing a hierarchy; either from the top down or from the bottom up. A system that is accessed from the top down and in which position nodes are nested below group nodes, is indistinguishable from one that is accessed from the bottom up and in which group nodes are nested below position nodes. Model fitting, therefore, is restricted to deciding whether position codes depend on group codes, or vice versa, and it cannot of itself resolve these additional issues.

Experiment 1: Recall of Grouped Digit Sequences

In recalling the serial position of an item in a grouped sequence, four possible responses can be made. The item's group and position can be correctly recalled; the group may be recalled but the position forgotten; the position may be recalled but the group forgotten; or both may be forgotten. For the independence model these four possible outcomes can result because both the grouping and position codes are present; one or the other is absent; or both are absent. However, for a hierarchical model in which position codes are accessed via the group codes above them, failing to recall the group, but recalling the position, is no different from failing to recall both. Once the group code is lost, no position information is available, and correct recall of a position is no more than a lucky guess. If retrieval were to have begun from the bottom of the hierarchy, and group codes accessed by the position codes below them, then failing to recall a position while getting the group correct is the same type of error as getting both wrong.

By developing some mathematical models, we can use the above ideas to predict errors in the recall of serial positions. The models presented here are too simple to capture all the features of the data, but comprise the minimal assumptions necessary to assess the relative adequacies of the independence and hierarchical models. The development of the models begins with a consideration of the pattern of response frequencies expected in a confusion matrix under the various conditions of code dependence and independence outlined above. The confusion matrixes for these studies consisted of $6 \times 6$ tables in which the rows represented stimuli (the actual positions of the probe item), and columns represented responses (the positions the subject thought the probe occurred in). Luce's Choice Model (Luce, 1959, 1963; Smith, 1980) offers a convenient way of deriving the expected frequencies for the models. The derivations are given briefly here, but a more complete description, with an example, appears in the Appendix.

Starting with the independence model, assume that $\gamma$ is the tendency to recall an item's group correctly, and that $\pi$ is the tendency to recall its position within the group correctly. Then,
Expressions for Expected Frequencies of the Independence, Higher Order Code, Item Search, and Dependence Models

<table>
<thead>
<tr>
<th>Response</th>
<th>Independence</th>
<th>Higher order code</th>
<th>Item search</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP</td>
<td>(N\gamma\pi/1)</td>
<td>(N\gamma\pi/G)</td>
<td>(N\gamma\pi/P)</td>
<td>(N\gamma\pi/D)</td>
</tr>
<tr>
<td>G*</td>
<td>(N(p-1)\gamma/1)</td>
<td>(N(p-1)\gamma/G)</td>
<td>(N(p-1)\gamma/P)</td>
<td>(N(p-1)\gamma/D)</td>
</tr>
<tr>
<td>P*</td>
<td>(N(g-1)\gamma/1)</td>
<td>(N(g-1)\gamma/G)</td>
<td>(N(g-1)\gamma/P)</td>
<td>(N(g-1)\gamma/D)</td>
</tr>
<tr>
<td>**</td>
<td>(N(g-1)(p-1)/1)</td>
<td>(N(g-1)(p-1)/G)</td>
<td>(N(g-1)(p-1)/P)</td>
<td>(N(g-1)(p-1)/D)</td>
</tr>
</tbody>
</table>

Note. \(I = \gamma\pi + (p - 1)\gamma + (g - 1)\pi + \sum_{i=1}^{p-1} \sum_{j=1}^{g-1} r_{ij} + (g - 1)(p - 1)\); \(G = \gamma\pi + (p - 1)\gamma + (g - 1)\pi + (g - 1)(p - 1)\); \(P = \gamma\pi + (p - 1) + (g - 1)\pi + \sum_{i=1}^{p-1} \sum_{j=1}^{g-1} r_{ij} + (g - 1)(p - 1)\); \(D = \gamma\pi + gp - 1\).

from Luce’s Biased Choice Axiom (Luce 1959, 1963; Smith 1980), if the tendency not to recall an item in its correct group or its correct position is \(1\) then the tendency to recall both its group and position is \(\gamma\pi\), the tendency to recall its group only is \(\gamma\pi\), and the tendency to recall its position only is \(\pi\). Expected response frequencies can be derived from the response tendencies for the independence model and are shown in Table 1. These frequencies are labeled \(GP, G^*, P^*, \text{ and } **\), denoting the number of times group and position are recalled, group only is recalled, position only is recalled, and neither is recalled. The expressions for response frequencies take into account that in a sequence with \(n\) groups and \(p\) positions in each group, there will be \(p - 1\) ways of getting the group right and the position wrong, \(g - 1\) ways of getting the position right and the group wrong, and \(p - 1\) ways of getting both wrong.

The derivation of response tendencies for the two hierarchical models in which position codes are accessed via group codes and group codes via position codes, is now straightforward, and the expected response frequencies are shown in Table 1. In the higher order code model (positions accessed via groups), the tendency to recall position but not group is the same as the tendency to recall neither, so that the \(\pi\) parameter drops out of the expression for the response frequency \(P^*\). In the item search model, the tendency to recall group but not position is the same as the tendency to recall neither, so that the \(\gamma\) parameter drops out of the expression for the response frequency \(G^*\).

Table 1 also shows the expected response frequencies for a fourth model so far not discussed. This is the complete dependence model, which assumes that grouping information can only be retrieved if position information is also retrieved, and that the converse is true. In this case, forgetting either the group or the position of an item is tantamount to forgetting both, and the parameters \(\gamma\) and \(\pi\) are not distinguishable from one another. Thus the model is a single-parameter one, unlike the others, which have two parameters. This model is redundant in that it would be true if both versions of the hierarchical model were true.

Method

Subjects. The subjects were 20 students from an introductory psychology course who volunteered for the experiment as partial fulfillment of course requirements. Ten served in each of the two grouping conditions.

Stimuli and procedure. The stimuli were 416 permutations of the digits 1 through 8. A different set of permutations was presented to each subject, with each item in a permutation presented one at a time on a computer-controlled visual display, so that an eight-item sequence took 4 s to present. Sequences were either grouped in two groups of four by a .5-s pause containing a 50-ms marker tone between the fourth and fifth digits or grouped in four groups of two, with the pause and marker tone coming after the second, fourth, and sixth items. The task was self-paced, and subjects began a trial by pressing a special key. This resulted in the presentation of the digit sequence, followed by a 1-s pause, and then a rehearsal-preventing task of eight random letters presented one at a time at a rate of one letter per 700 ms. Subjects were required to repeat each letter aloud three times as it appeared on the display; A probe digit, which had occurred in the sequence, was then shown, and subjects indicated its position in the sequence on a keyboard of eight horizontally arranged, equally spaced keys. These keys represented the eight serial positions of the sequence, and the subjects were instructed to press the key corresponding to the position occupied by the probe. When a key was pressed, the response was recorded, and the probe item was erased from the display. Feedback was then given, indicating whether the response had been right or wrong.

Testing was carried out over a 4-hr period, with the trials divided into four blocks of 109 sequences each. This allowed for rest periods of approximately 25 min between blocks. As the task was self-paced, the exact length of the rests varied from subject to subject, but none was shorter than 20 min. The first 5 trials in a block were practice, and the following 104 were test trials. Each serial position was probed 13 times for recall in each block.

Results and Discussion

The data from each subject in the two grouping conditions were used to find \(GP, G^*, P^*, \text{ and } **\), the frequencies of correctly reporting an item’s group and position, correctly reporting position only, and incorrectly reporting both. The mean values of these frequencies for the grouping-by-twos condition are shown in Table 2. Models were fitted to individual subjects. Observed and expected frequencies, and parameter estimates, are mean values for the group of 10 subjects. Individual chi-squares and degrees of freedom were summed to give group values.

The independence model, the two hierarchical models, and the dependence model were fitted to each subject’s data by finding the values of \(\gamma\) and \(\pi\), which minimized the chi-square calculated from observed and expected frequencies. The Hooke and Jeeves pattern search procedure was used for this purpose (Adby & Dempster, 1974). Mean values of \(\gamma\) and \(\pi\), and mean expected frequencies for the 10 subjects in the grouping-by-twos condition, are shown in Table 2. The table also shows the number of subjects whose data failed the goodness-of-fit test at the .01 level of sig-
GROUPING EFFECTS IN SHORT-TERM MEMORY

81

Table 3
Fit of the Independence, Hierarchical, and Dependence Models to Data From the Grouping-by-Fours Condition of Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Observed frequencies</th>
<th>Independence</th>
<th>Higher order code</th>
<th>Item search</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expected frequencies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GP</td>
<td>195.1</td>
<td>168.1</td>
<td>192.9</td>
<td>148.0</td>
<td>191.5</td>
</tr>
<tr>
<td>G*</td>
<td>41.4</td>
<td>68.3</td>
<td>41.2</td>
<td>21.0</td>
<td>32.1</td>
</tr>
<tr>
<td>P*</td>
<td>104.1</td>
<td>125.2</td>
<td>91.0</td>
<td>140.3</td>
<td>95.2</td>
</tr>
<tr>
<td>**</td>
<td>75.4</td>
<td>34.5</td>
<td>91.0</td>
<td>62.9</td>
<td>95.2</td>
</tr>
<tr>
<td><strong>Group values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>162.5</td>
<td>184.4</td>
<td>327.3</td>
<td>321.6</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>$N^*$</td>
<td>9</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td><strong>Parameter estimates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>3.25</td>
<td>1.53</td>
<td>3.26</td>
<td>6.39</td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>4.46</td>
<td>5.61</td>
<td>2.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. GP, G*, P*, and ** denote, respectively, the number of times group and position are recalled, group only is recalled, position only is recalled, and neither is recalled. Models were fitted to individual subjects. Observed and expected frequencies and parameter estimates are mean values for the group of 10 subjects. All of the overall chi-square values were significant at the .01 level. Thus the data differed significantly from the predictions of all of the models. However, this is not unexpected, given the simplicity of the models and the fact that the total chi-square values provide a powerful test of goodness of fit. What is important is that some did better than others in accommodating the data. The dependence model and the hierarchical system involving an item search performed poorly. The hierarchical system, which accesses items via higher order codes was satisfactory.

Experiment 2: Recall of Grouped Letter Sequences

The success of the higher order code model in accounting for the results of Experiment 1 may reflect the fact that digit pairs grouping may have been used, such as representing grouping and position information in the manner hypothesized by the dimensional model. The support for this idea is not very strong, as the independence model fits the data from the fours-grouping condition no better than the higher order code model. Another possibility is that subjects in the grouping-by-fours condition ignored the fact that the sequence had been divided into two groups of four items each and treated it as if it were grouped by twos, creating higher level codes from successive pairs of items. This possibility was checked by fitting the higher order code model to the fours-grouping data but treating the data as if it had come from the grouping-by-twos condition. The results of this analysis are shown in Table 4. The independence, higher order code, item search, and dependence models were fitted to individual subjects. Observed and expected frequencies and parameter estimates are mean values for the group of 10 subjects. Individual chi-squares and $df$s were summed to give group values. Although the overall chi-square value for the higher order code model was still significant, it gave a much better fit than the other three possibilities and was satisfactory in describing the data from most of the subjects. The results of Experiment 1 therefore support the hypothesis that grouping effects arise from a higher order coding strategy.

Table 2
Fit of the Independence, Hierarchical, and Dependence Models to Data From the Grouping-by-Twos Condition of Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Observed frequencies</th>
<th>Independence</th>
<th>Higher order code</th>
<th>Item search</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expected frequencies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GP</td>
<td>195.1</td>
<td>168.1</td>
<td>192.9</td>
<td>148.0</td>
<td>191.5</td>
</tr>
<tr>
<td>G*</td>
<td>41.4</td>
<td>68.3</td>
<td>41.2</td>
<td>21.0</td>
<td>32.1</td>
</tr>
<tr>
<td>P*</td>
<td>104.1</td>
<td>125.2</td>
<td>91.0</td>
<td>140.3</td>
<td>95.2</td>
</tr>
<tr>
<td>**</td>
<td>75.4</td>
<td>34.5</td>
<td>91.0</td>
<td>62.9</td>
<td>95.2</td>
</tr>
<tr>
<td><strong>Group values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>287.0</td>
<td>81.8</td>
<td>518.8</td>
<td>153.6</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>$N^*$</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td><strong>Parameter estimates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>4.10</td>
<td>3.15</td>
<td>3.26</td>
<td>6.33</td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>2.67</td>
<td>6.13</td>
<td>2.47</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. GP, G*, P*, and ** denote, respectively, the number of times group and position are recalled, group only is recalled, position only is recalled, and neither is recalled. Models were fitted to individual subjects. Observed and expected frequencies and parameter estimates are mean values for the group of 10 subjects. All of the overall chi-square values were significant at the .01 level. Thus the data differed significantly from the predictions of all of the models. However, this is not unexpected, given the simplicity of the models and the fact that the total chi-square values provide a powerful test of goodness of fit. What is important is that some did better than others in accommodating the data. The dependence model and the hierarchical system involving an item search performed poorly. The hierarchical system, which accesses items via higher order codes was satisfactory for half the subjects. This outcome makes sense in the light of some theories of grouping. According to Bower and Winzenz's (1969) phoneme chaining hypothesis, subjects will attempt to recode the sequence 34-51-79-82 into "thirty-four, fifty-one, seventy-nine, eighty-two." This is, of course, a specific example of a higher order coding strategy, and it would lead to the expectation that this model should fit the twos-grouping data best.

A similar analysis for the grouping-by-fours condition is shown in Table 3. In this case, the higher order code model did not fit the data particularly well. This is perhaps not surprising if subjects attempted to use a recoding strategy of the sort suggested by the phoneme chaining hypothesis. Recoding 3451-7982 as "Three thousand four hundred and fifty-one, seven thousand nine hundred and eighty-two" may be a much more difficult task than using the same strategy on the same sequence grouped 34-7982. This raises the possibility that a different method for...
are like words. On presentation their names are "looked up" in a lexicon, and it is these names that are stored in memory, rather than the individual digits. The lexicon thus performs the function of the code book referred to earlier. Experiment 2 used grouped sequences of random consonants, which did not make up familiar units that might have some prior lexical status.

The change in the type of stimulus item necessitated two other changes from the method used in Experiment 1. To obtain a reasonable level of accuracy it was necessary to abandon the rehearsal preventing task, which had been interpolated between the presentation of a sequence and the appearance of the probe, and also to reduce the length of the sequence to be remembered from eight to six items.

Method

Subjects. The subjects were 23 students from an introductory psychology course who volunteered for the experiment as partial fulfillment of course requirements.

Stimuli and procedure. The stimuli were 192 permutations of six different alphabetical letters, with the vowels and Y excluded. A different set of permutations was presented to each subject. The items in each permutation were presented on a computer-controlled visual display, with each letter appearing in the same position on the display for .5 s and disappearing before the appearance of its successor. The six items were divided into two groups of three by a 1-s pause between Items 3 and 4. A trial began when the subject pressed a key. A message giving the trial number then appeared on the display, followed 1 s later by the six items. There was then a pause of 1 s followed by the appearance of a probe item. This was a letter that had occurred in one of the six serial positions in the sequence. Subjects were required to indicate the position of the probe in the sequence by pressing one of six response keys, which were arranged in a 2 × 3 matrix in which the first row of keys indicated Serial Positions 1–3, and the second row, immediately beneath the first, represented Positions 4–6. After the response was made, there was an intertrial interval of 3 s before a message appeared asking the subject to press the key that started the next trial.

Testing was carried out in two blocks of 96 trials each, with a 10-min rest period between blocks. Each serial position was probed 13 times for recall in each block.

Results and Discussion

Each subject's data were used to find GP, G*, P*, and **, the frequencies of correctly reporting an item's group and position, correctly reporting group only, correctly reporting position only, and incorrectly reporting both. The mean values of these frequencies are shown in Table 5. Models were fitted to individual subjects. Observed and expected frequencies, and parameter estimates are mean values for the group of 23 subjects. Individual chi-squares and df's were summed to give group values.

The independence model, the two hierarchical models, and the dependence model were fitted to each subject's data by finding the values of γ and π, which minimized the chi-square calculated from observed and expected frequencies. Mean values of γ and

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Fit of the Independence, Hierarchical, and Dependence Models to Data From the Grouped Letters Task of Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed frequencies</td>
</tr>
<tr>
<td>Response</td>
<td></td>
</tr>
<tr>
<td>GP</td>
<td>196.2</td>
</tr>
<tr>
<td>G*</td>
<td>42.8</td>
</tr>
<tr>
<td>P*</td>
<td>95.4</td>
</tr>
<tr>
<td>**</td>
<td>81.5</td>
</tr>
</tbody>
</table>

Group values

Statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>χ²</th>
<th>df</th>
<th>N*</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP</td>
<td>348.0</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>G*</td>
<td>35.6</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>P*</td>
<td>723.1</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>**</td>
<td>162.2</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>

Parameter estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>γ</th>
<th>π</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP</td>
<td>4.65</td>
<td>1.62</td>
</tr>
<tr>
<td>G*</td>
<td>3.07</td>
<td>2.36</td>
</tr>
<tr>
<td>P*</td>
<td>2.32</td>
<td>5.22</td>
</tr>
<tr>
<td>**</td>
<td>6.49</td>
<td></td>
</tr>
</tbody>
</table>

Note. GP, G*, P*, and ** denote, respectively, the number of times group and position are recalled, group only is recalled, position only is recalled, and neither is recalled. Models were fitted to individual subjects. Observed and expected frequencies and parameter estimates are mean values for the group of 23 subjects. Individual chi-squares and df's were summed to give group values.

* Number of subjects out of 23 with p < .01.
The outcome of the model fitting for this experiment was different from that of Experiments 1 and 2. Although the overall chi-square values for the four models were significant at the .01 level, the independence model gave the best account of the data and satisfactorily fitted all but two of the subjects. This result suggests that when subjects are denied an easy verbal recoding strategy they represent grouped data in memory more in the manner expected by the nonhierarchical matrix model. Discussion of why such a structure should emerge in a situation in which verbal recoding is difficult to achieve will be left until the data from Experiments 5, 6, and 7 have been presented.

Experiment 4: Recall of Grouped Sequences of Tones

One of the questions posed by Experiments 1 and 2 was whether there are means of forming hierarchical codes that do not depend on a verbal recoding strategy. Memory for melodic sequences would seem to be a promising place to look for such codes. Dowling (1973) presented temporally grouped tone sequences followed by probes consisting of five notes. When the five-note probes fell within one of the previously presented groups they were detected more accurately than when they fell across the boundary of two groups. Deutsch (1980) found that when the grouping of a sequence by temporal gaps coincided with grouping by melodic rules, recall of the sequence was better than when the gaps and the rules produced different groupings. Temporal gaps and the rules produced different groupings. Temporal gaps and the rules produced different groupings.

Table 6: Fit of the Independence, Hierarchical, and Dependence Models to Data From the Grouped Non-Alphanumeric Characters Task of Experiment 3

<table>
<thead>
<tr>
<th></th>
<th>Observed frequencies</th>
<th>Independent code</th>
<th>Item search</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GP</td>
<td>92.0</td>
<td>85.2</td>
<td>89.1</td>
<td>88.7</td>
</tr>
<tr>
<td>G*</td>
<td>48.0</td>
<td>53.7</td>
<td>48.1</td>
<td>38.0</td>
</tr>
<tr>
<td>P*</td>
<td>26.1</td>
<td>32.0</td>
<td>18.2</td>
<td>27.1</td>
</tr>
<tr>
<td>**</td>
<td>25.8</td>
<td>20.9</td>
<td>34.4</td>
<td>38.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistics</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^2$</td>
<td>49.4</td>
<td>83.6</td>
<td>86.3</td>
<td>144.3</td>
</tr>
<tr>
<td>$df^*$</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>N*</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter estimates</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>2.79</td>
<td>1.40</td>
<td>3.25</td>
<td>2.21</td>
</tr>
<tr>
<td>$\pi$</td>
<td>3.30</td>
<td>3.92</td>
<td>1.59</td>
<td></td>
</tr>
</tbody>
</table>

Note. GP, G*, P*, and ** denote, respectively, the number of times group and position are recalled, group only is recalled, position only is recalled, and neither is recalled. Models were fitted to individual subjects. Observed and expected frequencies and parameter estimates are mean values for the group of 11 subjects. Individual chi-squares and $df$ were summed to give group values.

* Number of subjects out of 11 with $p < .01$. 

Experiment 3: Recall of Grouped Nonalphanumeric Characters

Experiments 1 and 2 raise two questions. First, Are there other means of forming hierarchical codes that do not depend on a verbal recoding strategy? Second, Are there situations in which strategies other than the generation of higher order codes are used to preserve the serial positions of grouped sequences? This third experiment addressed these questions by using material that was less amenable to a strategy that would allow the members of a group to be coded as a single verbal unit. A subject may be able to name symbols such as #, &, and $ but it is difficult to do for groups of digits and letters.

Method

Subjects. The subjects were 11 students from an introductory psychology course who volunteered for the experiment as partial fulfillment of course requirements.

Stimuli and procedure. The experiment was identical to Experiment 2 except that the stimuli were 192 permutations of six different nonalphanumeric characters drawn from the set #, $, %, &, ,, +, @, T, <. Subjects.

Results and Discussion

Each subject's data were used to find $GP$, $G^*$, $P$, and ** the frequencies of correctly reporting an item's group and position, correctly reporting group only, correctly reporting position only, and incorrectly reporting both. The mean values of these frequencies are shown in Table 6. Models were fitted to individual subjects. Observed and expected frequencies, and parameter estimates, are mean values for the group of 11 subjects. Individual chi-squares and degrees of freedom were summed to give group values.

The independence model, the two hierarchical models, and the dependence model were fitted to each subject's data by finding the values of $\gamma$ and $\pi$ that minimized the chi-square calculated from observed and expected frequencies. Mean values of $\gamma$ and $\pi$, and mean expected frequencies for each model, are shown in Table 6. The table also shows the number of subjects whose data failed the goodness-of-fit test at the .01 level of significance, and the total chi-square and degrees of freedom for each model. These latter values were found by summing the chi-squares and degrees of freedom for the individual subjects.
poral grouping would thus appear to be an important factor in memory for musical sequences, and Deutsch (1982) has proposed that this is because it allows a chunking strategy to be used. If these chunks are like those generated by verbal recoding strategies then, as in those tasks, the retrieval of within-group position information should depend on the retrieval of the grouping information. By using musically naive subjects who cannot name the notes in the sequence, it is possible to preclude the use of verbal recoding to generate higher order codes.

Method

Subjects The subjects were 37 students from an introductory psychology course who volunteered for the experiment as partial fulfillment of course requirements. None had had any formal musical training.

Stimuli and procedure The experiment was identical to Experiments 2 and 3 except that the stimuli were 192 permutations of six different musical tones drawn from the eight notes in the scale of C Major for the octave starting at middle C.

Results and Discussion

Each subject's data were used to find two sets of frequencies: (a) $G_P$, $G^*$, $P^*$, and $**$, the frequencies of correctly reporting the group and position, group only, position only, and neither group nor position of an item in the group of tones presented first in the sequence; and (b) $G_P$, $G^*$, $P^*$, and $**$, the frequencies of correctly reporting the group and position, group only, position only, and neither group nor position of an item in the group of tones presented second in the sequence.

This was a slight departure from the method for presenting the results of the previous experiments, in which data for recall were pooled over groups. The decision to present data separately for each group in this experiment was motivated by the fact that error rates in this task were sufficiently high to examine the behavior of the model parameters, $\gamma$ and $\pi$, as a function of their group in the sequence. The mean values of the observed frequencies and the expected frequencies and parameter estimates for the four models, are shown in Table 7. Parameters were fitted by the minimum chi-square procedure used in the previous studies. The overall chi-square for each model and the numbers of subjects for whom each model could be rejected at the .01 level of significance are also shown in the table.

All of the overall chi-squares in Table 7 were significant at the .01 level, but the higher order code model gives the best fit overall and was satisfactory for 24 of the 37 subjects. As none of the subjects in the experiment could read music and none could name the notes in the sequences, it appears unlikely that they used a verbal strategy to create high-level codes for groups of tones. Subjects in Experiments 1 to 2 almost all confirmed that they had made extensive use of verbal recoding techniques. By contrast, subjects in this task claimed to have rehearsed the tone sequences by singing them silently. Thus the experiment provides some evidence for a hierarchical memory structure that is not produced by a verbal recoding strategy.

For a model to be acceptable it must not only provide a reasonable fit to the data but its parameters should also behave sensibly. It would have been expected that the second group in the sequence, which produced the larger number of completely correct responses, should have produced the larger values of $\gamma$ and $\pi$. This was the case for the independence and dependence models, and for the best fitting model, the higher order code model.

Experiment 5: Dissimilarity Judgments of Letter Sequences

Experiments 1–4 tested the independence model, the two hierarchical models, and the dependence model with methods that relied on an analysis of the frequencies of different error types in probed recall of grouped sequences. Models can, of course, be uninterestingly specific to the experimental procedures developed to test them. Experiment 5 involved a very different method for evaluating the models. The task involved presenting a grouped sequence for memory and afterwards showing a second sequence consisting of a reordering of the items from the first sequence. Subjects were then required to judge the dissimilarity of this variation from their memory of the original. The original and the variations consisted of two groups of two items each. If the items in the original are denoted 12–34, then the variations used are those shown in Table 8. The first of these has the same as...
Table 8
The Eight Reorderings of the Original Sequence Used to Create the Variation

<table>
<thead>
<tr>
<th>Variation</th>
<th>Type of transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>12–34</td>
<td>Same order as original</td>
</tr>
<tr>
<td>34–12</td>
<td>Groups reordered</td>
</tr>
<tr>
<td>21–43</td>
<td>Positions within groups reordered</td>
</tr>
<tr>
<td>43–21</td>
<td>Groups and positions within groups reordered</td>
</tr>
<tr>
<td>t12–34</td>
<td>Same as original but with a nonordering difference</td>
</tr>
<tr>
<td>t34–12</td>
<td>Groups reordered plus nonordering difference</td>
</tr>
<tr>
<td>t21–43</td>
<td>Positions reordered plus nonordering difference</td>
</tr>
<tr>
<td>t43–21</td>
<td>Groups and positions reordered plus nonordering difference</td>
</tr>
</tbody>
</table>

Note: The symbol t preceding variations 5–8 denotes that these variations involved an additional change from the original that was unrelated to order.

The perceived dissimilarity, \( d \), between the original and a variation can be expressed in the following equation:

\[
d = gG + pP + kK + T,
\]

where \( G = 0 \) if the groups in the variation are in the same order as those in the original, and \( G = 1 \) if they have been transposed, but within-group position has not been altered in either case; \( P = 0 \) if the within-group positions in the variation are in the same order as those in the original, and \( P = 1 \) if they have been transposed, but position of the groups has not been altered in either case; \( K = 0 \) if either group, or positions within groups in the variation are in the same order as in the original, and \( K = 1 \) if both have been changed; \( T = 0 \) if the variation is only a reordering of the original, and \( T = 1 \) if some other change has been made to it; and \( g, p, k, \) and \( t \) denote the weights each of \( G, P, K, \) and \( T \) assume in determining the value of the dissimilarity.

The expression for the perceived dissimilarity thus has the form of a linear regression equation in which \( G, P, K, \) and \( T \) are dummy variables, and \( g, p, k, \) and \( t \) are unstandardized regression weights. Dunn (1983) has used similar equations for testing whether dissimilarity judgments were fitted by various spatial metrics. It is assumed that the components of the dissimilarity combine additively to determine the overall judgment. This assumption need not be taken on faith, however. As in any regression model, it is possible to include additional terms that test for interactions between variables.

Equation 4 provides a general framework for assessing the four models, independence, higher order code, item search, and dependence via dissimilarity judgments. Each makes its own prediction about the relation between \( k \), which reflects the importance of a change in both group and position, and \( g \) and \( p \), which reflect the importance of the separate changes. Each model is now developed.

Higher Order Code Model

This model assumes that each possible ordering of items in a group will generate its own higher order code. For the task we are considering there are four possible reorderings of items, which will give rise to higher order codes denoted A, B, C, and D, according to these rules: 12 → A, 21 → B, 34 → D, 43 → E.

The encoded version of the original 12–34 will thus be AD, and the encoded versions of the variations will be (1) 12–34 → AD, (2) 34–12 → DA, (3) 21–43 → BE, (4) 43–21 → EB.

According to this model, Variations 3 and 4, which involve transposing positions within groups only, and transposing both groups and positions, will be seen as equally dissimilar to the original. In terms of Equation 4, \( k = p \).

Additionally, it is reasonable to expect that as the encoded version of Variation 2 produces the same higher level codes as the original, albeit in a different order, there will be less dissimilarity between it and the original than between the original and either of Variations 3 and 4.

Item Search Model

The simplest way of computing the dissimilarity between a variation and the original for this model is to assume that the judge computes the number of links each item in the variation has been moved along a schematic representation of the hierarchy from its position in the original. It is easy to see that in terms of Equation 4, this assumption produces the equality \( k = g \). That is to say, changing both groups and positions will produce no more dissimilarity between a variation and the original than changing only the groups.

Dependence Model

If this model is developed as it was in the case of the confusion data it would produce the equality \( k = g = p \). This is tantamount to saying that any change is as dissimilar as any other, and as before, this model will be true when both the higher order code and item search models are true.

Independence Model

The easiest way to derive this model’s predictions is to consider the dimensional version in which the items in a grouped sequence are considered to be positions on a plane, one dimension being the position of the group in the sequence, and the other the position of an item in its group. In terms of Equation 4, therefore, transposing an item’s group will move it a distance \( g \) across the plane, whereas transposing its position will move it a distance \( p \). The distance moved when an item’s group and position are simultaneously altered will depend on the spatial metric assumed. Two Minkowski metrics are commonly used to scale dissimilarity judgments; the Euclidean and the city-block metrics.

For a Euclidean metric the equality will be \( k = (g^2 + p^2)^{1/2} \).

For a city-block metric the equality will be \( k = g + p \). A third, less commonly used metric, the dominance metric, is also worth
mentioning here. It predicts the equality \( k = g, \) or \( k = p, \) whichever of \( g \) or \( p \) is the larger. Depending on which parameter is the larger, this is equivalent to one or other of the hierarchical models.

**Testing the Equivalences**

Tests of the higher order code, item search, and dependence models are straightforward and do not even require solution of the regression equation implied by Equation 4. It should be noted, however, that the same equivalences between the higher order code and item search models apply in this situation as for the analysis of confusion data in Experiments 1-4. For example, a higher order coding system in which items within groups are recoded into single units is equivalent to an item search in which grouping information is nested under position information.

If one can be satisfied that ratings of dissimilarity are linear with respect to the perceived dissimilarity parameter in Equation 4 then it is sufficient to compare the mean dissimilarity for a simultaneous change of group and position to the mean dissimilarities for the separate changes. The assumption of a linear relation between observed ratings of dissimilarity and the underlying model parameter should not be taken on trust, as its violation could result in incorrect rejection of these models and the independence model. A method for testing this assumption will be described presently.

Observed ratings of dissimilarity cannot be used to test the Euclidean and city-block versions of the independence model. As Dunn (1983) has indicated, the relation between the dissimilarity parameter \( d \) and an observed dissimilarity rating, \( D, \) is, at best, a linear one. The expression relating rated dissimilarity to the model parameters is given thus:

\[
D = a(gG + pP + kK + tT) + c + \text{error},
\]

where \( a \) and \( c \) are constants associated with the rating scale used.

The constant \( a \) can be ignored, as it only defines the unit of measurement. With its removal, Equation 5 can be solved by normal linear regression procedures to obtain estimates of \( g, p, \) and \( k. \) These can then be used to test the equalities for the Euclidean and city-block versions of the independence model.

**Response Scale Linearity**

Tests of all of the equivalences are valid only if an observed rating of dissimilarity, \( D, \) is linearly related to its underlying parameter, \( d. \) Anderson's (1982) procedures for checking for additivity in information integration tasks provide a means of checking response scale linearity. The dissimilarity between a variation and the original can be affected not only by the manner in which the variation is reordered but also by other independent means. As mentioned at the beginning of this section, if the original consisted of two groups of upper-case letters, the variation could have its letters in the same or in a different case. Equation 4 assumes that this type of change combines additively with the changes in ordering. If this assumption is correct, and if \( d \) is a linear function of \( D, \) then the four dissimilarity ratings obtained with the case change should differ by a constant from the four obtained without the case change. Stated in terms of the analysis of variance (ANOVA) this implies the absence of an interaction between changes due to reordering of the variation's items and any other transformation that affects dissimilarity between variations and the original. Of course the lack of an interaction cannot be used as a test of linearity if the transformation has no effect on the dissimilarity judgments. There must be a main effect contribution of this variable, which there was in all of the experiments conducted, and which can be seen in the significant contribution it made to all the regression equations.

The presence of an interaction effect cannot be interpreted unambiguously. It may be due to a nonlinear relation between \( d \) and \( D, \) or it may have been caused by a nonadditive relation between the reordering variables, \( G, P, \) and \( K. \) The other transformation variable, \( T, \) in Equation 4.

**Method**

**Subjects.** The subjects were 20 students from an introductory psychology course who volunteered for the experiment as partial fulfillment of course requirements. Ten served in the grouped condition, and 10 in the ungrouped condition.

**Stimuli.** The stimuli consisted of two sequences of letters presented on a visual display. Each sequence consisted of the same four letters drawn from the set B, D, E, G, M, N, Q, T. Two methods of presentation were used. In the grouped condition the letters in a sequence were presented as two groups of two letters each. The first pair of letters appeared side by side just to the left of the center of the display for 100 ms, accompanied by a 500-Hz marker tone. The letters were then erased, and 500 ms later the second pair appeared side by side just to the right of the center of the display. They were also accompanied by a 500-Hz tone and were erased after 100 ms. In the ungrouped condition the four letters appeared simultaneously in a single line on the display for 200 ms.

The first sequence, the original, always consisted of upper-case characters. It was followed 500 ms later by the variation, whose letters were all either in upper or lower case. The letters in the variation either were in the same order as in the original or were a reordered achieved by reversing the order of presentation of the groups, reversing the order of presentation of items within the groups, or reversing the presentation of both group and position order. These reorderings were used in both the grouped and ungrouped conditions. These changes were combined factorialily with a change in case to give the eight different trial types shown in Table 8 above. During the course of the experiment, each type of trial was presented six times in a random order that was different for each subject.

**Procedure.** Each trial involved the presentation of the original sequence and its variation and following these a 500-Hz tone was presented, accompanied by the message RESPOND on the visual display. Subjects then rated the similarity between the original and the variation on a visual-analogue scale on which the right-hand end of the scale was labeled SAME and the left-hand end, DIFFERENT. The scale consisted of a printed circuit board, approximately 30 cm long, which produced a voltage indicating the point on the scale touched by a subject's finger. The voltage was converted to an integer value between 1 and 100 indicating the degree of dissimilarity between the original and the variation. This value was then displayed on the experimenter’s console and stored in a file for subsequent analysis. It did not appear on the subject's monitor. If the subject responded before the variation had been presented, the message RESPONSE TOO QUICK was shown on the display. If a response had not occurred 6 s after the presentation of the variation, the message RESPONSE TOO SLOW appeared. Following the response, and before the next trial, there was an intertrial interval of 3 s.

Before doing the task, subjects were told that they would be shown pairs of sequences and would have to judge the similarity between them. It was explained that the variation would always consist of the same letters as the original but that these might be rearranged, might be in a different case, or both. They were told that they would have to take all
Table 9
Mean Dissimilarity Judgments of Each Variation From the Original in the Letters Tasks

<table>
<thead>
<tr>
<th>Variation</th>
<th>Grouped condition</th>
<th>Ungrouped condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-34</td>
<td>6.3</td>
<td>5.0</td>
</tr>
<tr>
<td>34-12</td>
<td>35.5</td>
<td>51.7</td>
</tr>
<tr>
<td>21-43</td>
<td>40.5</td>
<td>43.7</td>
</tr>
<tr>
<td>43-21</td>
<td>47.3</td>
<td>52.6</td>
</tr>
<tr>
<td>t12-34</td>
<td>50.7</td>
<td>39.1</td>
</tr>
<tr>
<td>t34-12</td>
<td>75.1</td>
<td>82.5</td>
</tr>
<tr>
<td>t21-43</td>
<td>84.7</td>
<td>83.5</td>
</tr>
<tr>
<td>t43-21</td>
<td>88.0</td>
<td>84.8</td>
</tr>
</tbody>
</table>

Note. The symbol t preceding variations 5-8 denotes that these variations involved a change in case from that used for the originals.

Results and Discussion

Mean dissimilarity ratings to each of the eight trial types were computed for each subject by averaging ratings from the 48 test trials. The mean dissimilarities for the grouped and ungrouped conditions are shown in Table 9. The data from the grouped condition were analyzed as follows.

A 4 x 2 repeated measures ANOVA was carried out on the mean ratings, the factors being type of reordering of the variation and the case of the letters in the variation. As explained earlier, the interaction of Type of Reordering X Case can be used to check for nonlinearity of the response scale. This interaction was not significant, $F(3, 27) = .54, MSe = 56.66, p > .05$, and the dissimilarities for Variations 1-4 appear to differ by a constant from the dissimilarities for Variations 5-8 in Table 9. Therefore the assumption of response scale linearity did not appear to have been grossly violated, and the equalities suggested by each model were not significant, $F(3, 27) = .54, MSe = 56.66, p > .05$. As these means were not equal, the data differ significantly from the predictions of the item search and the dependence models.

Tests of the independence model first required solving the regression equation of Equation 5, which relates dissimilarity ratings to the model parameters $g$, $p$, $k$, and $t$. A regression analysis was carried out for each of the 10 subjects, the values of the dependent variable being the 48 dissimilarity ratings for the separate trials. The independent variables were $G$, $P$, $K$, and $T$, which were coded as 1 or 0 for each value of the dependent variable according to whether its variation was a reordering of groups only, positions only, both groups and positions, and whether the variation’s case had been changed. In addition, the set of independent variables included the additive constant, $c$, associated with the origin of the rating scale. The mean values $c$, $g$, $p$, $k$, and $t$, the unstandardized regression coefficients, are shown in Table 10.

The same analysis can be carried out for the entire group of subjects, with the values of the dependent variable being the eight mean dissimilarity ratings for each subject. Such an analysis will give the same estimated values of $c$, $g$, $p$, $k$, and $t$ as averaging their values from the individual subjects’ regressions. However, the analysis is of interest because it provides an overall estimate of the proportion of variance accounted for by the regression equation and overall tests of significance of the contributions of the independent variables to the regression. The regression gave $F(5, 74) = 107.2, p < .05$, and accounted for 90.1% of the variance in the dissimilarity judgments of the group. The values of $t$, shown in Table 10, for the four independent variables were also significant at the .05 level.

The Euclidean version of the independence model was tested by computing the expected value of $k$ from the values of $g$ and $p$ for each subject from the expression $k' = (g^2 + p^2)^{1/2}$. Its mean value for the group of 10 subjects was 43.2. This expected value differed significantly from the mean value of $k$, which had been estimated from the regression analyses of the data of individual subjects and which is shown in Table 10, $F(1, 9) = 4.29, MSe = 35.52, p < .05$. The data thus differ significantly from the predictions of this version of the model.

The city-block version was tested by computing $k' = g + p$ from the values of $g$ and $p$ for each subject. Its mean value for the group of 10 subjects was 60.1. This expected value differed

Table 10
Summary of the Regression Analysis on Dissimilarity Judgments From the Subjects in the Grouped Condition of the Letters Task

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unstandardized regression weight</th>
<th>$t^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>7.4</td>
<td></td>
</tr>
<tr>
<td>$g$</td>
<td>26.4</td>
<td>7.47</td>
</tr>
<tr>
<td>$p$</td>
<td>33.7</td>
<td>9.52</td>
</tr>
<tr>
<td>$k$</td>
<td>39.1</td>
<td>11.05</td>
</tr>
<tr>
<td>$t$</td>
<td>42.8</td>
<td>16.89</td>
</tr>
</tbody>
</table>

Note. $c =$ intercept of regression line; $g =$ weight for group transposition; $p =$ weight for within-group position transposition; $k =$ weight for double transposition; $t =$ weight for case change. $t_{df} = 74$. 
significantly from the mean value of $k$, which had been estimated from the regression analyses of the data of individual subjects, $F(1, 9) = 44.70, M^{2}S_{E} = 103.20, p < .05$. The data thus differ significantly from this version of the model.

Although none of the models gives a completely satisfactory fit to the data, the truth seems to lie somewhere between the higher order code model and the Euclidean version of the independence model. Changing both groups and positions in the variation produced a bigger change than changing positions alone, contrary to the predictions of the higher order code model. However, the change was not as great as that predicted by the Euclidean version of the independence model. It is quite possible that a subject's data from a series of trials reflect the use of two different strategies for encoding grouped sequences, one that relies on the generation of higher order codes, and the other that results in independent group and position codes being assigned to items.

Another possibility is that the similarity judgments conform to a Minkowski metric other than the Euclidean or city-block metrics. The discovery of the metric that best describes the data involves finding the value of $r$ that produces the least discrepancy between observed values of $k$ and expected values obtained from the expression $k' = |g' + p'|^{1/r}$. The best fit was obtained for $r = 3$, giving a mean value of 39.3 for $k'$ for the group of 10 subjects. This did not differ significantly from the mean value of $k$, which had been estimated from the regression analyses of the data of the individual subjects, $F(1, 9) = .01, M^{2}S_{E} = 28.383, p > .05$. Thus, there is a version of the independence model that gives a good fit to the data. However, it is not easy to argue strongly for the model unless some psychological sense can be given to a Minkowski metric with $r = 3$.

Analysis of the data from the ungrouped condition depends on how subjects are assumed to compute the dissimilarities between the ungrouped original and its variations. Three possibilities were considered.

First, if the items in ungrouped sequences are not assigned separate codes to designate their groups and positions, subjects' judgments may correspond to the complete dependence model, and all variations may be treated as equally dissimilar from the original.

Second, subjects may behave in the manner proposed by the item search model and compute the number of positions moved by each item in the variation from its position in the original. The dissimilarity judgment may then be proportional to the sum of these changes. On this hypothesis, the dissimilarities of Variations 2 and 4 in Table 8 should be equal, and both should be greater than the dissimilarity of Variation 3.

Third, subjects may perform an operation like correlation on the order of the items in the original with that of the items in the variation and base their dissimilarity judgments on the size of the correlation. According to this view, Variation 2 should have the smallest dissimilarity from the original; Variation 4 should have the largest; and Variation 3 should have an intermediate dissimilarity.

The dissimilarity of Variation 4 did not differ from that of Variation 2, $F(1, 9) = .15, M^{2}S_{E} = 172.40, p > .05$, nor did it differ from that of Variation 3, $F(1, 9) = 1.09, M^{2}S_{E} = 238.93, p > .05$. Also, the dissimilarity of Variation 2 did not differ from that of Variation 3, $F(1, 9) = 1.45, M^{2}S_{E} = 83.31, p > .05$. The interaction of Type of Reordering × Case of Letters in the variation was not significant, $F(3, 27) = 1.60, M^{2}S_{E} = 49.55, p > .05$.

These analyses suggested that response scale linearity was not grossly violated in the ungrouped condition. They also suggested that the pattern of dissimilarities was more consistent with the assumptions of the dependence model than with the ideas that subjects judged the dissimilarity of an ungrouped sequence, either on a count of the number of positions that items had moved in the variation relative to the original, or on the correlation between the order of items in the original and the variation. Note that the pattern of dissimilarities in the ungrouped condition was different from that in the grouped condition, suggesting that in this task, the memory representation of ungrouped items is different from that of grouped items.

**Experiment 6: Syllabic Versus Nonsyllabic Sequences**

This experiment attempted to encourage subjects to consistently use, or to refrain from using, a higher order coding strategy. Because the evidence from previous studies suggests that such a strategy will be used when the items in a group can be recoded as a single verbal unit, the items in a group were selected so that they either consisted of pronounceable consonant-vowel-consonant trigrams, the CVC condition, or unpronounceable consonant-consonant-vowel-consonant trigrams, the CCC condition. Clearly, the dissimilarity judgments in the CVC condition were expected to conform to the predictions of the higher order code model, with changing the ordering of both positions and groups in the variation allowing the same degree of dissimilarity as changing the positions alone. The outcome for the CCC condition was less certain, although the results of Experiments 3 and 5 suggested that when the items in a group cannot be easily recoded as a single syllabic unit, subjects may resort to a strategy that assigns independent codes to an item's group and position, in which case either the Euclidean or city-block metrics may provide the better fit to the dissimilarities.

**Method**

**Subjects.** The subjects were 20 students from an introductory psychology course who volunteered for the experiment as a partial fulfillment of course requirements. Ten received sequences in which the middle letter of each group was the vowel I, and 10 received sequences in which the middle letter of each group was the consonant R.

**Stimuli.** The stimuli consisted of two sequences of letters presented on a visual display. Each sequence consisted of the same six letters, presented as groups of three letters each. In the CVC condition, the middle letter of a group was always I; in the CCC condition it was always R. The first and third letters of a group were consonants and made up the permutations B+V, B+Z, T+V, T+Z, V+B, Z+B, V+T, and Z+T, where • denotes the middle letter of the group. The two groups in the first sequence were chosen so that the first and third consonants of the second group were different from those of the first. As in Experiment 5, the first sequence, the original, always consisted of upper-case characters. The variation that followed it consisted either of all upper- or lower-case letters. The letters in the variation were either in the same order as in the original or were a reordering of it. This was done by reversing the order of presentation of the groups, reversing the order of presentation of items within the groups, or reversing the presentation of both group and position order. These changes were combined factorially with a change in case to give the eight different trial types shown in Table 8. During the course of the experiment, each type of trial was presented six times in a random order.
that was different for each subject. The method of presentation of the original and the variation was as described for Experiment 5.

Procedure. The procedure was the same as for Experiment 5, with each trial involving the presentation of the original sequence and its variation, followed by a similarity rating using the same equipment as that for the previous experiment. Subjects were given the same instructions and the same number of practice trials as in Experiment 5, and again there were 48 test trials in which all eight different reorderings of the variation were presented six times each.

Results and Discussion

Mean dissimilarity ratings to each of the eight trial types were computed for each subject by averaging ratings from the 48 test trials. The mean dissimilarities for the CVC and CCC conditions are shown in Table 11.

A 4 × 2 repeated measures (ANOVA) was carried out on the mean ratings for the CVC condition, the factors being type of reordering of the variation and the case of the letters in the variation. As explained earlier, the interaction of Type of Reordering × Change in Case can be used to check for nonlinearity of the response scale. This interaction was not significant, \( F(3, 27) = .22, MS_e = 58.393, p > .05 \).

A similar analysis carried out on the data from the CCC condition also produced a nonsignificant interaction of Type of Reordering × Change in Case, \( F(3, 27) = .88, MS_e = 61.735, p > .05 \). The dissimilarities for Variations 1–4 appear to differ by a constant from the dissimilarities for Variations 5–8 in both the CVC and CCC conditions in Table 11. Therefore the assumption of response scale linearity did not appear to have been grossly violated, and the equalities suggested by each model could be tested.

The higher order code model predicts that the dissimilarity caused by changing both group and position ordering (Variations 4 and 8) will be the same as that caused by changing group ordering alone (Variations 2 and 6). A comparison between the means for Variations 4 and 8 with the means for Variations 2 and 6 for the CVC condition gave \( F(1, 9) = 43.39, MS_e = 674.5, p < .05 \). In the CCC condition the same comparison gave \( F(1, 9) = 86.02, MS_e = 207.5, p < .05 \). The data thus differ significantly from the predictions of the item search and dependence models for both the CVC and CCC conditions.

The independence model was tested in the manner described for Experiment 5 by carrying out a regression analysis on the 48 dissimilarity ratings for each of the 20 subjects. The mean values of \( c, g, p, k, \) and \( t \), the unstandardized regression coefficients, are shown in Table 12. The regression equation for the CVC condition gave \( F(5, 74) = 66.85, p < .05 \), and accounted for 85.4% of the variance in the subjects' judgments. For the CCC condition, the regression gave \( F(5, 74) = 66.85, p < .05 \), and accounted for 88.4% of the variance in the subjects' judgments. The values of \( t \), shown in Table 12, for the four independent variables, were also significant at the .05 level.

The Euclidean version of the independence model was tested by computing the expected value of \( k \) from the values of \( g \) and \( p \) for each subject from the expression \( k' = (g^2 + p^2)^{1/2} \). Its mean value for the 10 subjects in the CVC condition was 52.1, which differed significantly from the mean observed value of \( k \) in Table 12, \( F(1, 9) = 9.55, MS_e = 58.393, p > .05 \).

The regression equation for the CCC condition gave \( F(5, 74) = 23.49, p < .05 \). In the CCC condition the same comparison gave \( F(1, 9) = 43.39, MS_e = 674.5, p < .05 \). The data thus differ significantly from the predictions of the item search and dependence models for both the CVC and CCC conditions.

Table 11

<table>
<thead>
<tr>
<th>Variation</th>
<th>CVC condition</th>
<th>CCC condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>12–34</td>
<td>4.3</td>
<td>7.5</td>
</tr>
<tr>
<td>34–12</td>
<td>25.1</td>
<td>25.6</td>
</tr>
<tr>
<td>21–43</td>
<td>49.9</td>
<td>41.2</td>
</tr>
<tr>
<td>43–21</td>
<td>50.8</td>
<td>51.2</td>
</tr>
<tr>
<td>t12–34</td>
<td>31.3</td>
<td>40.5</td>
</tr>
<tr>
<td>t34–12</td>
<td>53.4</td>
<td>64.0</td>
</tr>
<tr>
<td>t21–43</td>
<td>81.1</td>
<td>79.2</td>
</tr>
<tr>
<td>t43–21</td>
<td>80.8</td>
<td>83.1</td>
</tr>
</tbody>
</table>

Note. CVC = consonant-vowel-consonant; CCC = consonant-consonant-consonant. The symbol \( t \) preceding variations 5–8 denotes that these variations involved a change in case from that used for the originals.

Table 12

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unstandardized regression weight</th>
<th>( t^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVC condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( c )</td>
<td>3.3</td>
<td>4.34</td>
</tr>
<tr>
<td>( g )</td>
<td>21.4</td>
<td>9.66</td>
</tr>
<tr>
<td>( p )</td>
<td>47.7</td>
<td>9.73</td>
</tr>
<tr>
<td>( k )</td>
<td>48.0</td>
<td>8.33</td>
</tr>
<tr>
<td>( t )</td>
<td>29.1</td>
<td>11.92</td>
</tr>
<tr>
<td>CCC condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( c )</td>
<td>6.3</td>
<td>4.98</td>
</tr>
<tr>
<td>( g )</td>
<td>20.9</td>
<td>8.65</td>
</tr>
<tr>
<td>( p )</td>
<td>36.2</td>
<td>10.32</td>
</tr>
<tr>
<td>( k )</td>
<td>43.2</td>
<td>11.92</td>
</tr>
<tr>
<td>( t )</td>
<td>35.3</td>
<td></td>
</tr>
</tbody>
</table>

Note. CVC = consonant-vowel-consonant; CCC = consonant-consonant-consonant. \( c \) = intercept of regression line; \( g \) = weight for group transposition; \( p \) = weight for between-group position transposition; \( k \) = weight for double transposition; \( t \) = weight for chase change. *df = 74.
Mean Dissimilarity Judgments of Each Variation From the Original in the Brackets Task

<table>
<thead>
<tr>
<th>Variation</th>
<th>Dissimilarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-34</td>
<td>8.0</td>
</tr>
<tr>
<td>34-12</td>
<td>29.3</td>
</tr>
<tr>
<td>21-43</td>
<td>38.1</td>
</tr>
<tr>
<td>43-21</td>
<td>48.6</td>
</tr>
<tr>
<td>12-34</td>
<td>52.8</td>
</tr>
<tr>
<td>34-12</td>
<td>74.9</td>
</tr>
<tr>
<td>21-43</td>
<td>84.0</td>
</tr>
<tr>
<td>43-21</td>
<td>89.6</td>
</tr>
</tbody>
</table>

Note. The symbol t preceding variations 5-8 denotes that those were presented in inverse video, unlike the originals which were presented in normal video.

The mean dissimilarity judgments of each variation from the original in the brackets task are presented in Table 13. The dissimilarity ratings for each of the eight trial types were computed for each subject by averaging ratings from the 48 test trials. These are shown in Table 13.

Results and Discussion

Mean dissimilarity ratings to each of the eight trial types were computed for each subject by averaging ratings from the 48 test trials. These are shown in Table 13.

A 4 X 2 repeated measures ANOVA was carried out on the mean ratings, the factors being type of reordering of the variation and the video mode of the variation. As explained earlier, the interaction of Type of Reordering X Video Mode can be used to check for nonlinearity of the response scale. This interaction was not significant, F(3, 27) = .74, p > .05. Therefore, the assumption of response scale linearity did not appear to have been grossly violated, and the equalities suggested by each model could be tested.

The higher order code model predicts that the dissimilarity caused by changing both group and position ordering (Variations 4 and 8) will be the same as that caused by changing position ordering alone (Variations 3 and 7). A comparison between the means for Variations 4 and 8 with the means for Variations 3 and 7 gave F(1, 9) = 7.58, MSe = 349.2, p < .05. The data thus differ significantly from the predictions of the higher order code model.

The item search model predicts that the dissimilarity caused by changing both group and position ordering (Variations 4 and 8) will be the same as that caused by changing group ordering alone (Variations 2 and 6). A comparison between the means for Variations 4 and 8 with the means for Variations 2 and 6 gave F(1, 9) = 15.44, MSe = 736.8, p < .05. The data thus differ significantly from the predictions of the item search and dependence models.

The independence model was tested in the manner described for Experiment 5 by carrying out a regression analysis on the 48 dissimilarity ratings for each of the 10 subjects. The mean values of c, g, p, k, and t, the unstandardized regression coefficients, are shown in Table 14. The regression equation gave F(5, 74) = 70.04, p < .05, and accounted for 88.8% of the variance in the subjects' judgments. The values of t, shown in Table 14, for the four independent variables, were also significant at the .05 level.

The Euclidean version of the independence model was tested by computing the expected value of k from the values of g and p for each subject by the expression k' = (g^2 + p^2)^(1/2). Its mean value for the 10 subjects was 38.1, which did not differ signifi-
Judgments From the Subjects in the Brackets Task

$p = \text{weight for within-group position transposition}; k = \text{weight for double transposition.}$

Summary of the Regression Analysis on Dissimilarity

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unstandardized regression coefficient</th>
<th>$t^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>8.2</td>
<td></td>
</tr>
<tr>
<td>$g$</td>
<td>21.7</td>
<td>4.85</td>
</tr>
<tr>
<td>$p$</td>
<td>30.6</td>
<td>6.85</td>
</tr>
<tr>
<td>$k$</td>
<td>38.7</td>
<td>8.65</td>
</tr>
<tr>
<td>$t$</td>
<td>44.3</td>
<td>14.03</td>
</tr>
</tbody>
</table>

Note: $c = \text{intercept of regression line}; g = \text{weight for group transposition}; p = \text{weight for within-group position transposition}; k = \text{weight for double transposition}; t = \text{weight for case change}.$

$c = 8.2$, significantly from the mean observed value of $k$ in Table 14, $F(1, 9) = 0.00, MSe = 71.16, p > .05$.

The city-block version was tested by computing $k' = g + p$ from the values of $g$ and $p$ for each subject. Its mean value was 52.3. The expected values differed significantly from the values of $k$, which had been estimated from the data of individual subjects, the mean of which appears in Table 14, $F(1, 9) = 11.10, MSe = 178.26, p < .05$. The data thus differ significantly from the predictions of the city-block model.

The data from this experiment support the Euclidean version of the independence model, which fits the dissimilarity judgments closely. The other models were all able to be rejected. Taken together with Experiment 3, this study supports the notion that when visually presented items cannot be recoded in a convenient verbal form, grouping and position information are presented in a matrixlike form, in which the two types of information are independent of one another, rather than in a hierarchical form, in which retrieval of one type of information depends on successful retrieval of the other. It might have been argued that the particular symbols used as items in Experiment 3 and the manner of their presentation (one at a time), made a higher order coding strategy impossible. In this study the items in a group were presented together, and symbols were chosen that might have been expected to be seen as total configurations. These manipulations failed to evoke any higher order coding strategy.

General Discussion

These experiments suggest that the serial positions of grouped sequences are represented in two formats in memory. The first of these consists of a hierarchical structure like that proposed by Johnson's (1972) theory of opaque containers, in which information about the positions of items in a group is only accessible via high level codes representing the groups themselves. The second of these corresponds to Broadbent's (1981) matrix system, in which information about the position of an item in a group and the group's position within the sequence are independent of one another. Which of these structures is used seems to depend on the amenability of the stimuli in the sequence to one or another coding operation. If the items in a group can be recoded as a single verbal unit, then a hierarchical structure emerges, and the data are best fitted by the equations for the higher order coding model. If they cannot be thus recoded, the result is usually a matrix structure, and the data are best fitted by a version of the independence model. The data of Experiment 4, which used grouped sequences of musical notes, are an exception to this generalization. This experiment produced a hierarchical structure that was best fitted by the higher order code model.

Taken in isolation, each of these experiments can be accounted for by one or another of the theories described in the introduction to this article, but most theories cannot embrace all of the results. Experiments 1, 2, and 4, which produced hierarchical structures, fit well with the Johnson (1972) opaque container theory, as does the CVC condition of Experiment 6. The opaque container theory cannot accommodate the matrixlike structures found in Experiments 3 and 7, and in the CCC condition of Experiment 6, however. Conversely, the dimensional model and the directional and nondirectional associative models cope well enough with the matrix structures but not with the hierarchical ones. The only model that could predict both of these is the modified version of the reverberatory model, which postulates that an item's serial position can be lost either by perturbations in the timing loops used to hold items and coding elements for groups (the explanation given by Estes, 1972, and by Lee & Estes, 1981), or by the complete loss of a group's coding element (as suggested in the introduction to this article). Perturbation of timing loops produces results consistent with the independence model; loss of a group's coding element produces the dependency of within-group position information on the retrieval of a group's coding element as predicted by the higher order code model. However, the reverberatory model has no intrinsic properties that would predict the circumstances under which retrieval will fail for one reason or the other. It will be suggested presently that the data from these experiments are consistent with matrixlike or hierarchical structures because of the way in which serial position information was encoded, rather than because of failures in its retrieval. An attempt to develop a theory that accounts for the data from all of the experiments is therefore better left until after some discussion of the likely encoding strategies used in the different tasks.

There are a number of ways in which groups of items may have been recoded into higher order units in Experiments 1, 2, and 4, and in the CVC condition of Experiment 6. In Experiment 1, which used digit sequences, the recoding seemed to involve naming successive pairs of digits and remembering the names rather than the digits themselves. In other words, the digits in each pair were treated like the letters making up a word. This would explain why, even when a sequence of eight items was divided into two groups of four, subjects treated it as if it had been presented as four groups of two items each. The groups of three consonants used in Experiment 2 were unfamiliar to subjects and thus had no names. However, it appears from the subjects' own reports that these were recoded as pronounceable units by interpolating vowels. In the CVC condition of Experiment 6, a vowel was provided between a pair of consonants to make the group of items pronounceable. Leaving aside the tones task of Experiment 4 for the present, it is tempting to suggest that these cases of higher order coding depend on the use of a phonemic recoding strategy of the sort proposed by Bower and Winzenz (1969), and that the recoded phonemes are stored in a memory system like the articulatory loop of Baddeley and Hitch's (1974)
working memory. There are two reasons for hesitating before accepting this explanation. The first is Broadbent and Broadbent's (1981) report that the letters of meaningful trigrams are grouped, even under conditions of articulatory suppression, and the second, albeit weaker evidence, is the claim of Bower and Winzenz's own subjects that they did not phonemically recode groups of digit sequences. A contrary piece of evidence comes from a study by McNicol (1971), which showed that runs of repeated digits were less well grouped under conditions of concurrent articulation than when they were watched silently. It appears, therefore, that two mechanisms may be needed to account for the generation of higher order verbal codes. One of these may rely on the fact that higher order codes already exist in long-term memory for some sequences of items and that these codes are not articulatory in form. Visually presented words can be recognized without the need to articulate (Coltheart, 1978), and it is likely that digit pairs, and meaningful trigrams, such as FBI and BBC, are represented in the same lexical system as words, and can be automatically recognized as single units on the basis of their orthographic characteristics. On the other hand, the strategy for recoding unfamiliar strings of digits, such as 11000111, may use the articulatory loop to store the recoded sequence.

Despite differences as to whether higher order coding involves the generation of phonemic or orthographic units, the above examples suggest that this form of recoding is a verbal process. The tones task of Experiment 4 suggests that something similar occurs in domains unrelated to verbal processing. According to Deutsch and Feroe (1981), information about the pitch of tones in a sequence is encoded hierarchically by rules which conform to Gestalt principles such as figural goodness, proximity, and continuation. However, temporal grouping of tones in a sequence, which is considered to be a relatively low-level process preceding the application of these rules, constrains their use so that the tones in different groups of the sequence are likely to be represented under separate nodes in the hierarchy (Deutsch, 1980). Figure 2 depicts a possible representation of a series of notes, abc-def, temporally grouped by a pause between c and d. The six notes are represented at the bottom level of the hierarchy and are accessed via the dominant note for each group, which reoccurs at the next level up in the hierarchy. Deutsch and Feroe (1981) point out that failure to retrieve a dominant note at an upper level will result in loss of information about members of the group below it. This property is also shared by the higher order code model. However, Deutsch and Feroe's theory is not an example of higher order coding, because the elements in the upper levels of the hierarchy are not abstract codes containing the members of a group but representations of specific notes that occur in the sequences below them. Both the theory of pitch representation and the higher order code model are specific examples of the general principle that when a hierarchy is accessed from the top down, elements at lower levels can be produced only if those above them have been retrieved successfully.

The results of Experiments 3 and 7, and the CCC condition of Experiment 6, require a different interpretation. All of these results were better fitted by the independence model than by the higher order code model. Moreover, the dissimilarity judgment tasks were generally best fitted by the particular version of the independence model, which assumed Euclidean distances between items and their groups. This raises the possibility that a quasi-spatial code is used to store the group sequences.

Accepting this possibility raises some problems for some theories of how serial position information is represented in memory. As has been shown, the directional association model and the reverberatory model both assume the independence of position and group codes. Neither contains the slightest suggestion that these codes behave as points in a Euclidean space. Only a dimensional model, which assumes that grouped sequences are stored in a quasi-spatial fashion, has this property. However, the emergence of a Euclidean metric for the dissimilarity judgments is not in itself powerful evidence for a visual-spatial code. Euclidean metrics have been found to accommodate a variety of dissimilarity judgments, many of which involve stimuli that are unlikely to be encoded in a visual-spatial fashion (Carroll & Wish, 1974). Nevertheless, the possibility that subjects in Experiments 3 and 7 did use spatial codes to represent groups of items merits further consideration in the light of considerable evidence that serial position information is sometimes stored in a spatial form. Healy (1975, 1977, 1978, 1982) has shown that the serial positions of spatially presented items are represented differently in memory from those of items presented in temporal order only. She has suggested that the code used to preserve spatial order is not verbal but of the analogical form proposed by Anderson (1978) and Kosslyn and Pomerantz (1977). Peterson, Rawlings, and Cohen (1974) presented temporally grouped auditory sequences of dots and dashes and asked subjects to imagine each group as a row in a two-dimensional spatial matrix. Their results suggested that this coding strategy enabled subjects to manipulate this representation as if it were a visual-spatial image.

Accepting that grouped sequences are represented spatially would seem to implicate the visual-spatial subsystem of the Baddeley and Hitch (1974) working memory model. Taken in conjunction with our earlier suggestions about the processes needed to carry out higher order coding operations for verbal material and to generate hierarchical structures for sequences of musical notes, it is apparent that the data from these studies are not well accommodated by theories that propose a general code for serial position information, be it based on associations or on the use of positional information. Rather we appear to be dealing with a memory system in which the format of the code for serial...
position depends on the subsystem used to hold the sequence of items. There is therefore no more a single type of code for serial position than there is a single type of code for items. Nor is there a single grouping effect in short-term memory. Grouping appears rather to be the consequence of system-specific coding strategies.

Finally, a warning should be given about the limitations of the independence, higher order code, item search, and dependence models as they have been presented here. A complete theory of how order is coded in, and retrieved from, short-term memory must account for the accuracy and latency of retrieval of an item's order. The model parameters $\gamma$ and $\pi$, which are measures of accuracy of retrieval, can be considered as reflecting how much components of the memory trace are affected by noise. Thus something like Estes' (1972) reverberatory process must be basic to all models for the coding of order. More recently, Ratcliff's (1981) has proposed a version of the reverberatory model in which information from the noisy memory trace is retrieved over time in the manner described by Ratcliff's (1978) random walk process. The marriage of the random walk process to the reverberatory model allows predictions to be made about the time needed to retrieve order information from memory.

The independence, higher order code, item search, and dependence models do not address themselves to the issue of how the memory trace becomes noisy or how information is retrieved from a noisy trace. These models are concerned primarily with the possible structures used to preserve the order code and should be considered as components of a more complete model of order storage and retrieval. The complete model for a short-term memory subsystem will thus need to specify the particular kinds of data structures used by that subsystem, the manner in which those structures are perturbed by noise, and the process by which samples of information from the noisy trace are integrated over time to obtain a more faithful representation of the original data structure.

References


Appendix

Using Luce's Choice Model to Derive Expected Frequencies for Grouping Models

Experiments 1–4 involved presenting a sequence of items that had to be remembered, and then testing for memory of one of its serial positions by showing the subject a probe item that had appeared somewhere in the sequence. The subject then had to identify the position occupied by the probe. The basic data from such an experiment are the cells of a confusion matrix, a square table in which a row represents the serial position of the probe item and a column represents the position reported by the subject. The cell entries are the frequencies with which the responses are given to the stimuli. The expected frequencies for the independence, higher order code, item search, and dependence models can be derived with the aid of Luce's (1959, 1963) Choice Model, which assumes that each of the responses has a certain tendency to be evoked by each of the stimuli. These tendencies are called response strengths. The first step in deriving a model's expected frequencies is to make a table that shows the relative likelihood of each stimulus evoking each response. This is the relative response strength matrix. For the purpose of illustration, this matrix will be constructed for the independence model.

Because the response strengths in the table are relative ones, one of the responses must be used as a reference point. Its strength is set equal to 1, and the values for the other responses will then indicate how much stronger they are than this reference response. A convenient reference response is when the subject fails to identify both the group and the position of an item. These cells are those set equal to 1 in Table A1.

The total response strength to any stimulus may consist of a number of components whose effects combine multiplicatively. In Experiments 1–4, the memorial representation of a grouped sequence was assumed to contain two components, \( \gamma \), which is the strength with which a probe item correctly evoked the group, and \( \pi \), which is the strength with which the probe correctly evoked the within-group position. Thus the relative strength of any response by which the subject correctly identified both an item's position and its group is \( \gamma \pi \). In the response strength matrix of Table A1, the cells on the major diagonal correspond to those cases in which the probe item evokes both the correct group and the correct position, so that these cells all have the expected response strength of \( \gamma \pi \).

According to the independence model, either the group or the position component can be lost from the memorial representation without affecting the other component. Thus the strength of giving and item's group correctly but its position incorrectly, will be \( \gamma \), and the strength of giving an item's position correctly but its group incorrectly will be \( \pi \). These strength values are given in the appropriate cells of Table A1.

The next step is to obtain expressions for response probabilities. This is achieved simply by dividing each response strength by the sum of the response strengths for its row. By this means, the row totals are made to sum to 1, and the table has been converted to one of expected conditional response probabilities.

The model illustrated in Table A1 assumes that the values of \( \gamma \) and \( \pi \) are constant for all stimulus positions in the sequence and that there are no biases to any of the response positions. Under these assumptions only four response types need be considered. These are responses by which (a) both the group and position of an item are recalled correctly; (b) the group is recalled correctly and the position is recalled incorrectly; (c) the position is recalled correctly and the group is recalled incorrectly; and (d) both the group and the position are recalled incorrectly. The expected probability of each type of response can be found by combining all of its entries in the response strength matrix and then dividing by the sum of all the response strengths in the matrix. Thus in a response strength matrix in which there are \( g \) groups and \( p \) positions

\[
P_{\text{of recalling group and position}} = \frac{\gamma \pi}{\gamma \pi + (p-1)\gamma + (g-1)\pi + (g-1)p \pi}.
\]
GROUPING EFFECTS IN SHORT-TERM MEMORY

95

P of recalling group only
\[ p = \frac{\gamma}{\gamma r + (p - 1)\gamma + (g - 1)r + (g - 1)(p - 1)} \]

P of recalling position only
\[ p = \frac{\gamma}{\gamma r + (p - 1)\gamma + (g - 1)r + (g - 1)(p - 1)} \]

P of recalling neither group nor position
\[ p = \frac{1}{\gamma r + (p - 1)\gamma + (g - 1)r + (g - 1)(p - 1)} \]

The expected probabilities for the other models can be found by modifying Table A1 to give the appropriate response strengths. In the case of the higher order code model loss of information about an item's group implies loss of information about its position within the group. This means that the \( r \) component of a response strength cannot exist unaccompanied by the \( \gamma \) component, so that all of the single values of \( r \) in Table A1 must be replaced by values of 1. In the case of the item search model, loss of information about an item's position implies loss of information about its group. This means that the \( \gamma \) component of a response strength cannot exist unaccompanied by the \( r \) component, so that all of the single values of \( \gamma \) in Table A1 must be replaced by values of 1. In the case of the dependence model, loss of one type of information implies loss of the other and vice versa, so that all of the off-diagonal entries of the response strength matrix become one. Once a model's complete matrix of response strengths has been constructed, then the probabilities of the four types of response used in fitting the model can be derived in the manner just illustrated.

Received January 29, 1985
Revision received July 22, 1985

Instructions to Authors

Authors should prepare manuscripts according to the Publication Manual of the American Psychological Association (3rd ed.). All manuscripts must include an abstract of 100–150 words. In addition, the Editor encourages authors to prepare a summary of up to 600 words. Abstracts and summaries must be typed on separate pages. Typing instructions (all copy must be double-spaced) and instructions on preparing tables, figures, references, metrics, and abstracts appear in the Manual. Authors should also refer to the Manual for guidelines on the use of nonsexist language in APA journals. For further information on content, authors should refer to the editorial in the March 1983 issue of the Journal (Vol. 112, No. 1, pp. 136–137). For information on the other three JEP journals, authors should refer to editorials in those journals.

APA policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more journals. APA policy also prohibits duplicate publication, that is, publication of a manuscript that has already been published in whole or in substantial part in another journal. Also, authors of manuscripts submitted to APA journals are expected to have available their raw data throughout the editorial review process and for at least 5 years after the date of publication.

Authors will be required to state in their initial submission letter or sign a statement that they have complied with APA ethical standards in the treatment of their sample, human or animal. A copy of the APA Ethical Principles may be obtained from the APA Ethics Office, 1200 17th Street, N.W., Washington, DC 20036.

Anonymous reviews are optional, and authors who wish anonymous reviews must specifically request them when submitting their manuscripts. Each copy of a manuscript to be anonymously reviewed should include a separate title page with authors' names and affiliations, and these should not appear anywhere else on the manuscript. Footnotes that identify the authors should be typed on a separate page. Authors should make every effort to see that the manuscript itself contains no clues to their identities.

Manuscripts should be submitted in quadruplicate (the original and three copies), and all copies should be clear, readable, and on good quality paper. Authors should keep a copy of the manuscript as a guard against loss. Mail manuscripts to the Editor, Sam Glucksberg, Department of Psychology, Princeton University, Princeton, New Jersey 08544.

For the other JEP journals, authors should submit manuscripts (also in quadruplicate) to one of the editors at the following addresses: Journal of Experimental Psychology: Learning, Memory, and Cognition, Henry L. Roediger III, Department of Psychology, Purdue University, West Lafayette, Indiana 47907; Journal of Experimental Psychology: Human Perception and Performance, William Epstein, Department of Psychology, University of Wisconsin, W. J. Brogden Psychology Building, 1202 West Johnson Street, Madison, Wisconsin 53706; Journal of Experimental Psychology: Animal Behavior Processes, Michael Domjan, Department of Psychology, University of Texas, Austin, Texas 78712. When one of the editors believes a manuscript is clearly more appropriate for one or another journal of the American Psychological Association, the editor may, with the approval of the author, redirect the manuscript.