A MODEL FOR THE PROCESSING OF SEQUENTIAL INFORMATION

by

JAMES NORMAN MACGREGOR

M.A., Glasgow University, 1970
M.Sc., University of Victoria, 1975

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department

of

Psycholog[redacted]

We accept this dissertation as conforming
to the required standard

[redacted]
Dr. Alex Bavelas

[redacted]
Dr. Otfried Spreen

[redacted]
Dr. Charles W. Tolman

[redacted]
Dr. Bruce R. Johnson

[redacted]
Dr. Gary G. Miller

[redacted]
Dr. Richard Gottwald

UNIVERSITY OF VICTORIA

June 1979

Supervisor: Dr. Alex Bavelas

ABSTRACT

The dissertation proposed and tested a quantitative model of the processing of sequential information.

In the first chapter, criticisms were made of the most generally accepted current quantification of sequential patterns. First, the argument was forwarded that the high correlations which support the measure are not truly representative of its predictive power. It was demonstrated that the measure typically takes on a highly limited range of values, and for this reason cannot account for much of the systematic variation found in behaviour. Second, it was argued that the measure is conceptually limited, and cannot account for the results of a particular class of sequential processing studies concerned with temporal patterning.

In the second chapter a new quantitative model was proposed, which provides a measure of sequential patterns called "structural embeddedness" or "embeddedness-of-runs." The measure is sensitive to the relative positions of the larger and smaller subjective structures within a sequential pattern. It takes on a minimum value when the magnitudes of structures vary inversely with their degree of embeddedness within a pattern. This means that the lowest value obtains when the largest subjective structures are in the outermost positions, the next largest structures in the next outermost

positions, and so on. It takes on a maximum value when the opposite is true, and the largest structures are in the innermost positions, the next largest in the next innermost, and so on. Comparisons of the structural embeddedness of sequences with behavioural data in the literature indicated that subjects appear to prefer patterns with minimal structural embeddedness. Other results examined indicated that the subjective complexity of sequences varies directly with the degree of structural embeddedness. Two main hypotheses were proposed: that the complexity of sequences varies as a direct function of structural embeddedness, and that sequences tend to be organized into forms which minimize their structural embeddedness.

In the third chapter psychological interpretations of the model's two parameters were considered. The embeddedness of structures was proposed to reflect the order in which they are processed, with outermost structures being processed first, next outermost, next, and so on, resulting in an "ends-inward" type of processing. The magnitude of structures was given two possible interpretations: *either* larger structures required more effort, in the process of forming *or*, once formed, they represented more salient features of a sequence. The order-of-processing hypothesis, and the effort *or* salience hypothesis were proposed as subsidiary hypotheses.

The fourth chapter described the general method used in testing the experimental hypotheses. The method involved tracking subjects' anticipations during processing, and classifying responses into correct anticipation, wrong anticipation and no anticipation (copy response).

The fifth chapter described two experiments. Both experiments found significant differences in the complexity of sequences across levels of structural embeddedness, under conditions where other features of sequences were controlled. This supported the first hypothesis. The second experiment demonstrated a significant tendency for sequences to be reorganized towards minimizing structural embeddedness. This result supported the second hypothesis. In addition, the first experiment confirmed that embeddedness reflects order-of-processing, and indicated that the magnitude of structures affects their salience, with larger structures being more readily processed. The structural embeddedness measure was therefore interpreted as reflecting the degree to which the most readily processed structures of a sequence are in the most readily processed positions. Such an arrangement appears to maximize subjective pattern "goodness." Hypotheses were proposed to explain this finding, and a more general psychological theory outlined.

The final chapter considered possible extensions of the theory into the related areas of serial learning and verbal processing. Finally, an issue more central to the basic theory was discussed.

Examining Committee:

[redacted]

Dr. Alex Bavelas

[redacted]

Dr. Otfried Spreen

[redacted]

Dr. Charles W. Tolman

[redacted]

Dr. Bruce R. Johnson

[redacted]

Dr. Gary G. Miller

[redacted]

Dr. Richard Gottwald

TABLE OF CONTENTS

Chapter

LIST OF TABLES

Table

LIST OF TABLES

Table

Table

Table

LIST OF FIGURES

ACKNOWLEDGEMENT

I wish to express my thanks to Dr. Alex Bavelas for his unfailing support, enthusiasm and encouragement.

CHAPTER 1

A PRIORI MEASURES OF SEQUENTIAL COMPLEXITY

Concerted efforts to establish a psychologically relevant measure of sequential complexity began some 30 years ago. Information theory undoubtedly provided a stimulus in this direction, but it was quickly realized that the complexity of psychological sequences depended on more factors than just their binary information content. MacKay (1950) argued for two types of information, structural and metrical. Metrical information was a function of the number of events falling into pre-established categories or codes, and was equivalent to the entropy measure of information theory. Structural information, on the other hand, was a function of the category system itself, which depended on the degree of differentiation or "grain" of the receptive mechanism. MacKay was interested in the epistemology of the scientific experiment, and the relevant mechanisms were therefore scientific recording instruments. However, his analysis did not preclude structuring effects due to the perceptual and logical categories of the observer. Sensory and conceptual systems have a grain also, and this must affect the unitization of sense-data. The point was made by Miller (1956) specifically for psychological processes. Successful responses to a set of stimulus events imply that information has been processed, but the amount of that information depends, not on the raw events themselves, but on the events as encoded by the observer. The information-processing models which appeared subsequent to the general acceptance of this idea were therefore concerned with stimulus coding, and characteristically attempted to simulate how the human processor structures a stimulus sequence (Glanzer & Clark, 1962; Leeuwenberg, 1969; Payne, 1966; Restle, 1967, 1970; Simon & Kotovsky, 1963; Vitz & Todd, 1969).

Identifying the psychological structure is a first step towards quantification of a sequence, since it establishes the appropriate units of a sequence. The second step is to identify what operations should be performed on these units to provide a meaningful overall index of sequence complexity. The most common procedure has been to sum the number of coded groups, to obtain the "code-length" of a sequence (Glanzer & Clark, 1962; Leeuwenberg, 1969; Simon & Kotovsky, 1963). Vitz and Todd (1969) proposed an alternative, information-theoretic measure, H_{code}, which accounted for about 10% more of the variance in performance scores than code-length, but at the cost of a considerable increase in computational complexity. Simon (1972) reported a high correlation betweeh H_{code} and code-length, in the order of .9, and proposed therefore the acceptance of the simpler measure as the basic index of sequence complexity. Simon assumed run-coding in this calculation of code-length, taking runs of adjacent identical elements to be the basic structural units of binary and trinary sequences. The present concern is with binary sequences, and this practice will be followed. Therefore the terms "code-length" and "number-of-runs" will be used synonymously in referring to this particular measure of complexity.

The number-of-runs measure correlates highly with various response measures of pattern complexity, such as the judged complexity of sequential patterns, errors of recall, and number of learning trials. It typically accounts for around 80% of the variance in performance scores. What is the justification then for introducing a new measure of sequence complexity? The major reason is that code-length characteristically generates a highly restricted range of values for a given set of sequences, compared to the variety in performance scores. In typical cases only a few predictor values are correlated with a response variable having many more different-valued scores. Frequently the ratio is in the order of 1:2. This has had two consequences. First, it means that the high correlations reported have given a misleading impression of the actual goodness of fit between code-length and performance. Second, it means that code-length cannot account for the observed range of behaviour; there is more systematic variation in behaviour than there should be, according to the code-length measure of sequences.

Correlation as a Measure of Goodness-of-Fit

It has become standard practice in sequential processing to test a theoretical measure of complexity in terms of its correlation coefficient with performance scores (Leeuwenberg, 1969; Simon, 1972; Vitz & Todd, 1969). This represents a departure from the formal assumptions of correlation, since in these cases one variable is clearly designated the independent variable, and interest lies in prediction in a specified direction, as opposed to the mutual predictability of two variables. However, this departure may not always have serious consequences, and the procedure has the advantage of estimating goodness-of-fit in terms of a single score on a standard scale. This is useful when comparing results across experiments. Nevertheless, seriously misleading impressions may result when the independent variable has a highly restricted range of values compared to the dependent variable. For convenience, the number of different values of scores on a variable will be referred to here as the "variety" of scores. The present criticism is that frequently number-of-runs has a low variety of values, compared to performance scores. For example, in a study by Royer and Garner (1966) considerable variety in performance scores was obtained. Difficulties of course arise in assigning a precise value to variety. Variance depends on the scale of measurement and is not a suitable measure. Variety refers to the number of different scores, but how different is different on a continuous scale of measurement? Adopting the rough procedure of dividing the range by the number of scores to estimate an average "unit," and adopting scores at least one unit apart as having different values, then the variety of performance scores was 12, out of 19 sequences. If in this experiment number-of-runs had been used as a predictor, it would have had the task therefore of predicting 12 different performance values. However, the number-of-runs variable for the sequences itself took on only 3 different values. Nevertheless, the correlation was in the order of .9. This is typically the case. In a study by Galanter and Smith (1958), performance scores took on 4 distinct values, while number-of-runs had a corresponding variety of 2. Again, the correlation was .9. In studies reported by Vitz and Todd (1969) the variety of performance scores for two sets of trinary sequences were 10 and 13. The corresponding code-length scores had a variety of only 4 and 7 respectively. Correlations were again high.

If we think of variety as an index of the sensitivity of a variable, then the code-length measure is considerably less sensitive than the performance measures in use. The ratio of performance to code-length variety ranges here from 4:1 to 1.9:1, and generally appears to be of the order of 2:1. This means that when code-length and performance scores are correlated, the independent variable has typically half as many different scores as the variable which it is being used to predict.

The problem, then, is one of "heterogeneity of variety." A predictor with lower variety than a set of scores cannot account for those scores, and if the performance scores themselves are systematic, then the predictor does not have a high enough degree of differentiation to map onto them. Unfortunately, the correlation coefficient is not particularly sensitive to heterogeneity of variety. So long as scores on two variables fall in approximately the right order, variety on one of the two variables can be reduced considerably without much effect on their correlation. A small simulation was conducted to examine the behaviour of the correlation coefficient under variety conditions typical of the variables used in sequential processing.

The number of sequences employed experimentally may range from 4 to 24, and for purposes of the example a sample size of 10 was selected as fairly representative. Next, a performance variable was constructed having the values 1 to 10. This gave a variety of 10, which is common for performance scores in the area. Thereafter, a set of independent variables was correlated with the performance scores, each independent variable being of lower variety than the preceding one. The first independent variable was identical with the performance variable. Thereafter scores on the independent variable were progressively transformed into the two most extreme values, two at a time. After four such transformations the independent variable consisted of five 1s and five 10s (see Table 1). Thereafter 10s were progressively transformed into 1s, until the variable consisted of nine 1s and one 10. The effects of these progressive reductions in variety on the correlation coefficient are shown in Table 1.

Table 1: Effects on r of Restricting Variety in the Independent Variable

Dependent Variable	The Independent Variable								
	10	8	6	4	2,5	2,4	2,3	2,2	2,1
1	1	1	1	1	1	1	1	1	1
2	2	1	1	1	1	1	1	1	1
3	3	3	1	1	1	1	1	1	1
4	4	4	4	1	1	1	1	1	1
5	5	5	5	5	1	1	1	1	1
6	6	6	6	6	10	1	1	1	1
7	7	7	7	10	10	10	1	1	1
8	8	8	10	10	10	10	10	1	1
9	9	10	10	10	10	10	10	10	1
10	10	10	10	10	10	10	10	10	10
r =	1.00	.99	.97	.94	.87	.85	.80	.72	.55

It can be seen that under the conditions described the variety of scores on the independent variable can be reduced to 2 and yet a respectable correlation of .87 will result.

The preceding column, in which the variety is equal to 4, represents a situation which is highly analogous to sets of sequential processing scores. The ratio of the variety of scores on the dependent and independent variables is

2.5:1, with a correlation of .94. It might be said that the independent variable accounts for 88% of the variance in performance scores. It is clear from the example, however, that this cannot be taken to reflect the true predictability of scores from the independent variable. Drawing a score by chance from the independent variable and guessing the corresponding dependent value the average probability of being correct is .4. The next column to the right represents a more striking case. Here the probability of correct prediction is .2, yet the correlation is still .87. This, of course, is a crude comparison, since the probability of guessing the correct value does not take into account how close the guess is likely to be. But even when this is taken into account, a few-valued variable is likely to be a rather crude analogue of a many-valued variable. For the two-valued column, where the correlation is .87, the best estimate given a value of 1 is that the corresponding dependent value is 3, and, given a 10, that the corresponding value is 8. This represents an extremely crude level of prediction, given the fine range of values on the dependent variable, and this is precisely the problem with the number-of-runs measure; while it tends to place response scores in the right order, it does not discriminate gradations within that order. For this reason number-of-runs can provide only a gross measure of performance.

Simon (1972) proposed that number-of-runs is comparable in power to H_{code} as a measure of sequence complexity. The major evidence was that number-of-runs correlates highly with both H_{code} and performance scores. However, if the variety of the variables is examined, H_{code} is found to be two to three times as sensitive as number-of-runs, and in fact to have about the same variety of scores as the associated performance variables. This is a further example of the misleading impression which may arise if only the correlation between performance and predicted scores is examined.

Because of the restricted range of values typical of the number-of-runs measure, a misleading impression of its goodness-of-fit has arisen. The variable does not have the necessary degree of differentiation to account for sequence performance, and therefore cannot give detailed predictions of behaviour. The deficiencies of the code-length measure are particularly in evidence when it is applied to the results of temporal patterning experiments. The implications of those results for theories and measures of sequential processing are considered briefly below.

Temporal Patterning Experiments and Segmentation Effects

When a sequential pattern is iterated with no cues to indicate where the pattern repeats it may be referred to as a closed-cycle sequence (Glanzer & Dolinsky, 1965). Assuming such a sequence to be already in process, then there are as many different ways to segment the sequence into patterns as there are elements in the sequence. Thus if the sequence consists of n items before repetition, there are n possible starting points from which to generate different patterns of n items in length. Not all logically possible segmentations are equally preferred by subjects, however, and the results of the temporal patterning experiments conducted by Garner and his associates have demonstrated the existence of strong organizational preferences. Summarizing the major findings, it has been shown that, first, when sequences are presented with no systematic biases towards particular segmentations, certain segmentations are nevertheless spontaneously adopted significantly more frequently than others. These preferred patterns are correctly reproduced in fewer trials and with fewer errors than non-preferred organizations (Royer & Garner, 1966, 1970). Second, when subjects are given sequences to reproduce from particular starting points, the preferred type of pattern can be reproduced at a faster rate before error than non-preferred forms (Royer, 1967). Third, when subjects learn sequences by the method of anticipation, sequences started from preferred points tend to be learned in that form, while sequences started on non-preferred points tend to be reorganized into preferred forms (Garner & Gottwald, 1967). Fourth, when familiar patterns are presented, preferred forms are identified more rapidly than non-preferred forms (Handel & Lewis, 1970).

The experiments mentioned above, together with other findings (Garner & Gottwald, 1968; Preusser, Garner & Gottwald, 1970), form a coherent body of empirical results which demonstrate the importance of organizational factors in the processing of sequential patterns. Such organizational effects have been found in a number of different processes, including the acquisition, recognition and reproduction of patterns, and by a number of different experimental approaches, such as tracking, anticipation learning and motor production. There can be no doubt as to the reliability of the phenomenon. The results present a serious challenge to current theories and measures in the field of serial pattern learning. The two areas, temporal patterning and serial pattern learning, frequently employ the same patterns, and have many other similarities. Both types of experiment tend to evoke the same kinds of substructures during the process of acquisition (Garner & Gottwald, 1967; Restle & Burnside, 1972). Theoretical measures of pattern complexity, derived from serial pattern learning models, have been shown to predict performance in temporal patterning experiments (Vitz & Todd, 1969, using data from Royer & Garner, 1966). Nevertheless, there is one important area of non-overlap between them, the segmentation effects described earlier, about which serial pattern learning models have little to say. The Simon and Kotovsky (1963) model assumes that subjects begin processing by establishing a salient anchor-point, and the model will therefore predict the appearance of certain segmentations rather than others. The Vitz and Todd (1969) model assumes that, in the process of encoding, certain organizations are unstable and will reform into stabler encodings while, in other cases, several alternative segmentations may be possible, depending on where encoding begins. Thus the model allows for more than one organization of a sequence. However, in neither of these models do the assumptions appear to be sufficient to account for the empirical findings of temporal patterning. The anchor-point assumption leads to the expectation that subjects will segment sequences by starting on a salient point. In binary sequences this might be with the longest run. This corresponds in part with the results of Royer and Garner (1966, 1970) which indicated a strong tendency for sequences to be organized starting with the longest run. However, there was an equally strong tendency to end sequences with the longest run, which cannot readily be explained by the anchor hypothesis. Taking those sequences from Royer and Garner (1966) which cannot have an equally long run at both the beginning and the end of a pattern, that is, taking sequences where the two tendencies cannot both be satisfied, the proportion of cases where the longest run is placed at the start is 49.3, at the end is 50.7. The tendencies appear to be almost exactly equal. This is consistent also with the results of Royer and Garner (1970), where it was reported that preferred patterns are symmetrical, with the longest runs at the ends of the patterns. This symmetry is not consistent with the anchor hypothesis. The Vitz and Todd (1969) model permits alternative pattern organizations to appear, but again, not in a manner which is entirely consistent with the empirical findings. The model does not permit runs to be broken by segmentation, which means that a sequence cannot start and end with the same type of element, and this is consistent with the experimental results. Subjects appear to avoid this kind of organization (Royer & Garner, 1966), and it is unstable when given to subjects (Garner & Gottwald, 1967). However, as with the Simon and Kotovsky (1963) model, the Vitz and Todd (1969) model does not appear able to predict the end-stacking tendency with long runs.

Other models of serial pattern learning do not address the issue of segmentation at all, and appear to make the implicit assumption that the sequence is presented to a subject from a particular starting point, and that this starting point is retained throughout the acquisition process. Given the finding that certain forms tend to reorganize, this is no longer a plausible assumption (Garner & Gottwald, 1967).

Segmentation effects pose a problem for the code-length measure as well as for the coding assumptions of pattern learning models. The problem is that of restricted range, in a new form.

For any given sequence, some starting points give a pattern which is quickly acquired with few errors, while others yield difficult patterns, still others patterns of intermediate complexity. Thus, by permuting the starting points of a single sequence, a set of patterns ranging in difficulty may result. The code-length index, however, shows no corresponding spectrum of values under permutation of starting point. In fact, there are always only two different values, except when the sequence consists of a single run or alternation, in which case there is only one value. The lower value is obtained by starting the sequence at the beginning of a run. The number of ways this can be done is equal to the number of runs. The higher value, which is always one greater than the lower value, results whenever the starting point breaks up a run. The number of ways that this can occur depends on the number of runs and their lengths. Clearly, the variety in code-length values will always be relatively low when starting-point is permuted, with at most two possible values. This is not sufficient to account for the systematic variation in behaviour due to segmentation effects.

The aim of this chapter was to indicate certain deficiencies in the code-length measure of sequence complexity. Criticisms centred on the limited range of that variable.

In the following chapter a new measure of complexity is introduced, called"embeddedness-of-runs" or "structural embeddedness." The measure is sensitive to more information in a sequential pattern than the number-of-runs, and takes into account in addition the sizes and positions of runs within a pattern. The measure consequently has a greater range of values both for a given set of sequences, and for a set of starting permutations within a given sequence. This makes it a potentially more sensitive measure than code-length, one capable of measuring both sequence complexity and segmentation effects.

CHAPTER 2

EMBEDDEDNESS-OF-RUNS

The aims of the present chapter are to describe the embeddedness-of-runs measure and to test it against available data in the literature.

Embeddedness-of-Runs

The useful convention has arisen of describing binary sequences as ordered sets of run lengths. The sequence XXXOXXOO, for example, can be thus encoded as (3122), where each integer represents the length of a run, and a change of integer signals a change of element. The set of numbers generates either the original sequence, or its complement, OOOXOOXX, and if equivalence between complements is assumed, the notation preserves all relevant information about a sequence. The number of entries in the code gives the code-length of a sequence.

One way of summarizing the results of Royer and Garner's (1966) experiment is to say that subjects behave as if they prefer certain orderings of a run code over others. Longest runs are preferred at the beginning or end of a pattern, which is equivalent to saying that 3122 and 1223 are preferred codes. Royer and Garner (1970) observed an additional principle underlying empirical preferences.

Subjects apparently preferred patterns to have increasing or decreasing run lengths. Good patterns were those with symmetry or progression, or both. Garner concluded that a good pattern is one ". . . that both begins *and* ends with a long run, or that begins *or* ends with a long run either preceded or followed by a run of next shorter length, with regular progression in length beyond that. . . . It's not just a matter of beginning a pattern with a long run that's important--it's the relation of that run length to other run lengths in the pattern that matters" (Garner, 1974, pp. 55-56).

Royer and Garner (1966, 1970) identified certain qualities of sequences which characteristically resulted in good or poor subjective patterns. The new measure of pattern complexity was a result of attempts to give quantitative expression to those empirically derived qualities. Longest runs are preferred at the ends of a sequence, followed by the next longest runs, and so on. Let the distance of a run from either end of a pattern be defined as its "embeddedness," such that for a pattern of n runs, runs 1 and n are least embedded, 2 and n-1 are the next least embedded, and so on. The lowest embeddedness value is defined as one, and thereafter integer values are assigned in increasing order of embeddedness. The embeddedness of each run of the pattern 3122 is shown in Table 2.

Table 2: Embeddedness Values for an Ordered Set of Four Runs

embeddedness (e_i)	1	2	2	1
run size (r_i)	3	1	2	2

Notice that the embeddedness of a run depends on its position relative to the starting-point of the perceived pattern. If the sequence were segmented differently, starting on the singleton, for example, the runs would be assigned different embeddedness values. Embeddedness is therefore a characteristic of sequences which varies with segmentation.

Next, the quantity E is defined as:

$$E = \sum_{i=1}^{n} e_i \cdot r_i$$

where e_i is the embeddedness of the i^{th} run and
r_i is its size.

The variable, E, is referred to as the embeddedness-of-runs of a sequence, or more generally, as its structural embeddedness (for cases where structures other than runs provide the relevant units of a sequence). The embeddedness-of-runs of a given sequence will tend to be minimal when the longest runs are at the ends, the next longest runs next, and so on, and therefore the measure will tend to be minimal when a pattern is organized to correspond with the empirically observed principles of pattern preferences.

Table 3 shows how E changes for all possible starting points of the sequence 3122. Corresponding code-length values are also shown.

Table 3: Structural Embeddedness (E) and Number-of-Runs for Eight Possible Organizations of a Binary Sequence

Starting Point	Run Structure	E	Number-of-Runs
1	3122	11	4
2	21221	15	5
3	11222	15	5
4	1223	12	4
5	2231	13	4
6	12311	17	5
7	2312	12	4
8	13121	15	5

The sequence in Table 3 has eight elements, and therefore eight distinguishably different starting-points. Whenever a starting-point breaks a run an extra run is created, which is the reason why there are two different number-of-run values, 4 and 5. Breaking a run also creates an extra position within a sequence, which means that the set of embeddedness values changes from 1221 to 12321. When the embeddedness values of each position are multiplied by the size of run occupying that position, and summed across all positions, the E-score for that particular organization is obtained. It can be seen from Table 3 that structural embeddedness has a greater variety of different values than code-length, in the ratio of 2:1. This is typical of the different sensitivities of the measures. It will be recalled from Chapter 1 that the ratio of performance score to code-length variety was also about 2:1. This indicates that structural embeddedness and performance measures have a similar sensitivity.

Embeddedness-of-Runs and Segmentation Effects

Embeddedness-of-runs was derived by induction from certain experimental results, and consequently will correlate to some degree with those results. In testing the measure, it is therefore important to distinguish what was put into it in the first place from what are new, emergent properties of the measure itself. The measure was constructed to reflect the observed preferences of subjects to organize sequences with the longest runs at the start or end of a repeating pattern. E will tend to be minimal when a sequence has this property. Consequently, it is to be expected that subjects will prefer organizations which minimize the structural embeddedness of that sequence. This minimizing value will be referred to as E_{min} and positions which yield the minimum score will be referred to as E_{min} positions, or simply as minimum positions.

In Royer and Garner's (1966) study subjects listened to sequences until they felt able to track the sequence on response keys, at which point they began responding. This response point was taken to indicate the phenomenal starting-point of the pattern. It is to be expected, given the derivation of E, that such points will most frequently be points for which structural embeddedness is minimal, and this is the case. Of the 19 sequences, 16 were most frequently organized to yield E_{min} scores, the remaining 3 giving an E-score one point greater in a possible range of five. However, nothing in the empirical results suggested that there would be secondary preferences, and that the next most frequently selected patterns would have particular characteristics. In all cases, when the second most frequently chosen points were examined, however, they were found to be those points which generated the lowest remaining E-scores. This seems to represent some support for E as a continuous variable of pattern goodness, with gradations of preferences corresponding to gradations in the E-values of the different possible organizations.

A further interesting relationship emerges when the frequencies of selection of E_{min} positions are plotted in decreasing order of size. This is shown in Table 4. The table shows the run structure of the 19 sequences, then the minimal E-score for those sequences, then E_{min} divided by the number of positions within a sequence which give that score, and finally the total frequency with which E_{min} positions were selected. It can be seen that low E_{min} scores tend to occur in higher positions, high E_{min} scores

Table 4: Frequency of Response Point Selection as a Function of E_{min} and E_{min}/n

Sequence	Run Structure	E_{min}	E_{min}/n	Frequency of Selection as Response Point
B	11	2	1	100
S	211211	16	2.7	98.5
E	71	8	4	95.3
C	22	4	2	93.8
D	31	4	2	93.8
F	44	8	4	92.2
G	53	8	4	92.2
H	62	8	4	85.9
I	111113	14	7	78.1
L	1115	10	5	77.3
Q	2123	11	5.5	72.7
O	3122	11	5.5	67.1
N	1214	11	5.5	66.4
J	3113	10	10	50.8
K	4211	10	10	47.7
P	2213	11	11	46.1
M	4112	10	10	41.9
R	211112	14	14	31.3
T	212111	15	7.5	28.9

Note: Data from Royer and Garner (1966).

lower in the table. This suggests that while E_{min} organizations are generally the most attractive patterns for a given sequence, the strength of that attraction diminishes as E_{min} increases. There are reversals to the trend, in sequence S for example. Here, however, six of the eight sequence positions generated E_{min}. This suggests that the more ways there are to organize a sequence to obtain a minimal score, the more frequently will such organizations be selected. For this reason the number of E_{min} positions in a sequence was taken into account, by dividing E_{min} by n, the number of E_{min} positions. If the probability of organizing a sequence around an E_{min} position depends on the availability of the position as well as on its absolute value, this function should have a negative correlation with frequency of selection. In fact, the correlation was -.90. The relative variety of scores on the two variables can be compared by inspection. They are roughly equivalent. Defining variety as before, E_{min}/n has 8 "different" values, frequency, 11. The corresponding variety of code-length is only 3.

It may be useful here to stress a point which has been implicitly made up until now. Structural embeddedness is calculated by making assumptions about the coding of a sequence, and the measure can be only as good as those assumptions are correct. Here, runs are taken to be the effective units, but the concept of structural embeddedness will apply equally well to units formed under other coding principles. While it is true that with binary sequences runs are likely to provide the basic structure, depending on the particular sequence, other units may form. With sequence I, for example, where the run structure was 111113, Royer and Garner (1966) have suggested that the first four items may have been coded as two alternations (XO)(XO). This would change the structure to 2213 and E_{min} to 11, instead of 14. Using this value in the correlation calculation, r improves slightly, to -.91. Sequence R is a second example where alternation coding is likely to have occurred, changing the structure from 211112 to 2222 and E_{min} from 14 to 12. This improves the correlation again, to -.92. Occasionally, in cases like these, where there are long strings of single elements, it will be assumed that alternation coding has occurred.

A second study by Royer and Garner (1970) provides further data relevant to the embeddedness-of-runs measure. In identifying subjective starting preferences, the problem arises that the sequence must start somewhere, and this may bias subjects towards a particular starting point. In fact Garner and Gottwald (1968) found that for slow rates of presentation, subjects simply accept the organization implicit in the starting point provided by the experimenter. Royer and Garner (1966) had allowed for this by using exhaustively all possible starting points. In their later study, they used the more economical technique of starting all sequences at a rate of 30 elements per second, too fast for the given starting point to be identified, and then reducing the rate until subjects could report the pattern. The reported form was assumed to reflect subjective organization, and starting preferences were assessed in terms of the frequency of report of the possible forms. Sequences of four and six runs were used, but the six-run sequences tended to have many single-element runs. This raised ambiguity about whether subjects had used run or alternation coding, and so these sequences were ignored. When the 14 remaining sequences were examined, it was found that in all cases E_{min} forms had been the forms most frequently reported by subjects.

If the number of different ways of obtaining a minimal value for a sequence is taken into account, as before, by dividing the E-value by that number, the resulting variable was again found to correlate with the frequency of selection of minimal values. The correlation was $r = -.93$, indicating that the "attractiveness" of E_{min} organizations is a function of the absolute value of its structural embeddedness, and its availability.

The relationship of embeddedness-of-runs to the results of both studies shows the same pattern. In both cases, the organizations most frequently selected by subjects minimized the structural embeddedness of sequences. The strength of this tendency, however, weakened as the absolute value of E_{min} increased. When the availability of E_{min} positions was taken into account, then a strong correlation was found to exist between the absolute value of the minimal E-score positions and the attractiveness of those positions. It appears, then, that structural embeddedness reflects something of the goodness of patterns. The lower the structural embeddedness of a sequence from a particular point, the more likely is the sequence to be segmented into a pattern starting from that point.

Frequency of selection is one measure of pattern goodness. When subjects spontaneously select one type of pattern over another, there is evidence that such a pattern is psychologically preferred. Such patterns are also acquired more quickly, with fewer errors, and these facts together provide a set of operations which define goodness. Further indicants of pattern goodness were established by Garner and Gottwald (1967) in an experiment involving the anticipation learning of sequential patterns.

Subjects were presented with two different lights positioned to their left and right. There were two basic sequences, RRRLL and LLRLR, where R and L indicate right and left lights respectively. Sequences were started from all five positions, giving 10 different patterns in all. Once started, a sequence repeated continuously. Subjects learned two patterns, one from the RRRLL set, one from the LLRLR set. When subjects had learned a sequence they were asked to describe the pattern verbally. Sometimes the starting point of the verbal pattern retained the starting point given at the beginning of the procedure; sometimes it did not. In other words, it appears that subjects sometimes organized the sequence in the given form, at other times reorganized the pattern into an alternative form. Interest here is in whether such shifting occurs systematically, and in particular whether reorganizations occur in the direction of reducing structural embeddedness. If indeed preferred patterns are those with low structural embeddedness, then reorganizations which reduce E should be relatively more frequent than reorganizations which increase it. This can be tested directly. Garner and Gottwald (1967) report the total number of shifts occurring from and to each position in a sequence. A total of 226 occurred. Of these, 25 resulted in higher E-scores, 21 in equivalent E-scores and 180 in lower E-scores. The results support the hypothesis that reorganizations tend to occur in a direction of reducing the structural embeddedness, and indeed of minimizing it, since 170 of the 226 shifts reduced the structural embeddedness of the sequence to its lowest possible value.

There is a second method of examining the relationship between shifts and structural embeddedness. Two sets of probability values can be calculated from Garner and Gottwald's data. One is the probability that a given starting point will be retained. The second is the probability that a given point will usurp the starting position. The first may be regarded as a measure of the stability of a pattern around a given starting point, the second is a measure of the attractiveness of an alternative position. These variables turn out to have a correlation of .96. This indicates that a pattern which is stable when given, is also attractive when not given. Since the measures seem to reflect the same thing, only the first has been shown in Table 5. The table gives the run structure of the patterns, the corresponding E-values and the probability of a pattern retaining its given form.

Table 5: Probability of Retaining a Sequence Organization as a Function of Structural Embeddedness

Run Structure	Structural Embeddedness	Probability of Retaining Organization
32	5	.925
23	5	.825
122	7	.100
221	7	.023
131	8	.075
2111	7	.750
1112	7	.600
1121	8	.250
11211	8	.150
11111	9	.250

Note: Data from Garner and Gottwald (1967).

The upper part of Table 5 shows the probability of retaining a given organization for the simpler set of sequences. The probability values indicate the likelihood that the starting-point given on the first trial was also the starting-point of the verbal description given by the subject at the end of the experiment. It can be seen by inspection that the minimal patterns were the most stable, and there is a trend of decreasing stability with increasing embeddedness-of-runs. A similar trend is apparent in the more difficult set of sequences, in the lower part of the table. Again, E_{min} positions provided the most stable organizations. However, the tendency was less strong than in the case of the simpler sequences, where the absolute value of the minimum position was lower. This corresponds to the previous findings using data from Royer and Garner (1966, 1970), that while E_{min} positions tended to be the most attractive positions, the strength of this tendency varied across sequences as a function of the *absolute* values of E_{min}.

In brief summary, the data from three temporal patterning experiments were examined, and in all cases it was found that the pattern preferences indicated by subjects correlated with the structural embeddedness of the chosen sequences. In particular, the results showed that the lower the structural embeddedness of a pattern, the more likely that pattern was to be selected and retained by subjects. In this respect the structural embeddedness measure is superior to the code-length measure, which has no conceptual relationship to pattern preferences.

Embeddedness-of-Runs and Sequence Complexity

A number of empirical measures of sequence complexity have been reported in both temporal patterning experiments and serial pattern learning experiments. The measures have included, number of trials to a performance criterion (Galanter & Smith, 1958), number of errors to acquisition (Royer & Garner, 1966), rate of motor production (Royer, 1967), errors of recall and complexity ratings (Vitz & Todd, 1969). The correlation of these and other measures with the structural embeddedness of sequences is examined in this section.

Royer and Garner (1966) reported three response measures of sequence complexity, median number of trials to respond, number of errors, and variability in response points selected by subjects. The last variable measured the average uncertainty in bits of pattern choices, and takes on low values when there is one highly preferred starting-point, high values when there are no clear starting preferences. The measure, response point uncertainty (RPU), is taken by Royer and Garner to be a measure of pattern goodness, with better patterns having fewer alternative organizations. Consequently, the measure should correlate with structural embeddedness, which appears to be an a priori measure of pattern goodness. The correlation obtained was .94.* The relationship of E_{min} to the other response measures, errors and trials, was less strong, with coefficients of .72 and .76 respectively. Nevertheless, the values are as high as the intercorrelations of the response measures themselves. The average correlation between the three response measures was .76, while the average correlation between E_{min} and each of the three response measures was .81. Embeddedness-of-runs was therefore as good a

---

*All correlation coefficients reported were significant beyond the .05 level.

predictor as possible, given the reliability of the performance measures.

Royer (1967) measured the complexity of the same sequences by another technique. The sequences were presented on printed cards to subjects and the task was to reproduce the sequences on response keys at gradually increasing rates. The production rate was cued by an auditory pulse which started at a rate of one per second and increased by .2 per second at the end of each cycle. Trials continued until the first error, and the maximum response rate at this point was recorded. Royer compared the maximum rates for 16 8-element sequences with the response scores obtained for those same sequences by Royer and Garner (1966). Royer's measure, response rate, was found to have a correlation of -.77 with errors from Royer and Garner (1966), -.68 with trials and -.91 with RPU. The average correlation between these response variables was -.79. The correlation of E_{min} with Royer's response rate was -.83. Again, structural embeddedness showed as strong a relationship with the response variable as could be expected, given the reliability levels of the response variables.

Preusser (1972) reported a study in which complexity measures were incidentally obtained. The results are of interest here because number-of-runs was constant across sequences, and consequently cannot explain any systematic variation in errors. Subjects listened to repeating sequences of tones which had patterns of from 5 to 12 elements. The task was to describe the pattern as soon as possible, either verbally or by tapping. The lowest number-of-runs for all sequences was four. This was therefore the code-length of a sequence, so long as subjects began the pattern at the start of a run. If, on the other hand, the description began in the middle of a run, code-length increased to five. However, starting in the middle of a run occurred only 16 times in 1320 responses, or about 1% of the time. This meant that almost all sequences had been 4-run sequences in terms of subjects' organizations, and consequently no systematic variation in error scores was to be expected on the basis of code-length. There was nevertheless a range of error scores, having a variety of 10 different values. The correlation of errors with minimum E-scores of the sequences was .81.

The previous results were from the area of temporal patterning. Those which follow are from the area of serial pattern learning. Although the two areas employ similar kinds of stimuli, there are several differences between them. Temporal patterning hypothesizes holistic, perceptual-based processes, serial pattern learning stage-wise, learning-based processes. Temporal patterning is primarily interested in segmentation or starting-point preferences, serial pattern learning in sequence complexity. Temporal patterning typically employs faster presentation rates, and a tracking mode of response, serial learning a slower presentation rate and an anticipation learning mode of response. Nevertheless, certain equivalences have been found between the areas (Garner & Gottwald, 1967, 1968; Restle & Burnside, 1972). The results examined below suggest that structural embeddedness is as good a predictor of the complexity of serial patterns as it was of the goodness of temporal patterns.

Galanter and Smith (1958) used six different sequences in an anticipation learning paradigm. Subjects guessed each element in advance until they could identify the repeating pattern, and the number of trials to this point was recorded. Because of some extreme values, the median number of trials was used. Median values are shown in Table 6 for the six sequences. Both structural embeddedness and code-length values are given for purposes of comparison.

Table 6: Median Number of Learning Trials as a Function of Code-Length and Structural Embeddedness

Run Structure	Structural Embeddedness	Code-Length	Median Learning Trials
11	2	2	5.5
21	3	2	14.0
22	4	2	14.0
31	4	2	14.0
41	5	2	26.5
1121	8	4	49.0

Note: Data from Galanter and Smith (1958).

The correlation between embeddedness-of-runs and median trials was .98, and the variety of scores on the two variables was 5 and 4 respectively. The correlation between number-of-runs and median trials was .90, the variety of scores being 2 and 4 respectively. Although both theoretical variables correlate highly with the response measure, the embeddedness-of-runs has the higher degree of sensitivity. The extremely curtailed range of values of code-length here is a good example of the problems associated with that measure. Although it correlates with behaviour, it does so as a step-function, and is insensitive to performance differences which exist within sequences which have the same code-length.

Vitz and Todd (1969) conducted four experiments to test a coding model and an associated measure of sequence complexity, H_{code}. It was in connection with these results that Simon (1972) proposed number-of-runs as the simplest effective measure of complexity. The purpose here is to compare the performance of number-of-runs and structural embeddedness as predictors in the four experiments. In the first two experiments subjects judged the complexity of repeating binary and trinary sequences respectively. In the third and fourth experiments non-repeating trinary sequences were used. In the former case subjects again gave complexity judgments, in the latter, tachistoscopic presentation was used and errors of recall recorded. Table 7 shows the correlations between structural embeddedness and performance scores, and between code-length and performance scores, for each of the four experiments.

Table 7: Correlations of Structural Embeddedness and Code-Length with Four Sets of Performance Scores

Theoretical Measure	Binary Repeating	Trinary Repeating	Trinary Non-Repeating	Trinary Non-Repeating (Tachistoscope)
Structural embeddedness	.90	.84	.84	.86
Code-length	.91	.79	.82	.90

Note: Data from Vitz and Todd (1969).

In the case of the trinary sequences alternation encoding was assumed for four sequences. The correlations can be improved somewhat by assuming alternation coding in other likely instances, but interest at the moment is in comparing the variables for a given set of assumed encodings, rather than in obtaining the highest possible correlations. On average, both variables performed at the same level. Embeddedness-of-runs had an average correlation of .860 with performance scores, number-of-runs a correlation of .855. However, structural embeddedness was again the more sensitive measure. Estimating variety as the number of "different" scores, the average variety in performance scores was 12.5, in embeddedness-of-runs, 11.5, and in number-of-runs, 5.5. Therefore embeddedness-of-runs was more similar in scale to performance scores than was code-length. It consequently provides the more sensitive theoretical measure of the two. Code-length again showed the same relatively crude, step-wise relationship with performance scores as previously.

The last experiment to be examined showed a similar pattern of results. Scott and Henninger (1933) tested maze-learning in blindfolded human subjects using five different patterns of elevated finger mazes. This is a rather different kind of pattern than those examined until now. However, since the mazes consisted of right and left turns they can be encoded in a sequential code representing a series of left and right turns, such as LLRRLR. If subjects perform by making such sequential encodings then embeddedness-of-runs should again correlate with sequence complexity. There is evidence that many subjects in such experiments form verbal types of code. Woodworth (1938) reports the results of an experiment by Warden, in which over 40% of subjects reported using some kind of verbal mediation in learning stylus mazes. Scott and Henninger (1933) reported results on three dependent variables, errors, learning trials and time to learn. The variables were highly intercorrelated, and both embeddedness-of-runs and number-of-runs correlated highly with all three. The average correlation of embeddedness-of-runs with performance was .99, and of number-of-runs with performance, .95. Again, however, number-of-runs had a lower variety, having three different values, compared to embeddedness-of-runs and performance scores, which both had five different values.

In this chapter a new a priori measure of sequential patterns has been proposed. The measure was designed to reflect certain principles of pattern preferences observed empirically in temporal patterning experiments. The measure was shown to be superior to the code-length measure of patterns, in two respects. First, it was shown to correlate highly with subjective pattern preferences. In this guise it appears as a measure of pattern goodness. Code-length has no theoretical capacity to predict such effects, since permuting starting-points within a sequence may lead to many different patterns all having the same or similar code-lengths. Second, the new measure was found to correlate at least as highly as code length with measures of sequence complexity. The measures were compared on 9 different sets of response measures and in 7 of the 9 cases structural embeddedness correlated more highly than code-length with performance scores. In all cases the new measure was more sensitive, in the sense defined, and showed itself to be of the same scale as performance scores. For this reason, high correlations are more truly reflective of predictive power in the case of the new measure.

Primary Experimental Hypotheses

In reviewing the evidence supporting structural embeddedness as a useful quantification of sequential patterns, some 13 correlation coefficients were reported. The average of the absolute values of these coefficients was .87 and this, together with the less readily summarized relationships of E to pattern preferences, appears to provide considerable preliminary support for the measure. Nevertheless, the inherent weakness of using existing data is that the experiments used to test the measure were not designed for that purpose, and consequently alternative explanations are possible. In the present instance the apparent predictive power of E could be at least partially explained by the intermeshing of two alternative hypotheses. First, with respect to pattern preferences, the anchor hypothesis may be proposed as the true variable underlying E. The explanation might run as follows. In the case of closed-cycle sequences subjects establish a starting-point by identifying a recognizable regularity within a sequence. Single elements are easily confused with other single elements, and so a distinct, unambiguous feature has to be used. The longest run provides such a feature, and consequently sequences tend to be organized starting on the longest run. Such sequences also tend to have low E-scores since the longest run is in the lowest embedded position. According to this explanation, the remaining properties of the structural embeddedness measure are irrelevant. All that matters is that a sequence begins with the longest run.

There are several objections to be raised against the anchor hypothesis as an alternative explanation of preferences. First, it cannot account for the observed symmetry of preferences for long-run starts and long-run ends (Royer & Garner, 1966, 1970). Second, it does not take into account the preferences for orderly progression of run sizes (Royer & Garner, 1970). Nevertheless, in its defence, there was a tendency in Garner and Gottwald's (1967) results for the longest run to be preferred at the start of a sequence rather than the end, which is more consistent with the anchor hypothesis than with the embeddedness hypothesis.

The anchor hypothesis rests on the assumption that the subject establishes a starting-point in a closed-cycle sequence. It follows that if the anchor hypothesis is the real explanation of the apparent relationship between embeddedness and performance, then eliminating the need to establish a starting-point should make that relationship disappear. In other words, there should be no relationship of structural embeddedness with performance to open-cycle sequences. The results of the Vitz and Todd (1969) experiments in the case of non-iterating sequences already suggest that this is not the case. It was found in those cases that E continued to correlate with performance scores under those conditions. However, it is here that number-of-runs may enter as a second alternative explanation. E correlates with number-of-runs, and so could in this case have correlated with performance through mediation of number-of-runs as the true variable. In other words, embeddedness-of-runs may be no more than a combination of the effects of the anchor hypothesis and the number-of-runs measure. For this reason it was proposed to test the measure under conditions where both anchor effects and number-of-runs were controlled for.

The precise relationship of structural embeddedness to organizational preferences is also under some doubt. The studies of Royer and Garner (1966, 1970) showed a symmetry of preferences for having long runs at the start and at the end of a pattern. This is consistent with the embeddedness-of-runs measure, and not with the anchor hypothesis. Garner and Gottwald's (1967) results, however, indicated a preference for the long run to be at the start of a sequence. This is more consistent with the anchor hypothesis than with the embeddedness-of-runs measure. A further experimental test of the relationship of structural embeddedness to pattern reorganizations was therefore required to clarify the nature of this relationship. Two primary experimental hypotheses were tested.

Hypothesis 1: The complexity of a sequence is a function of its structural embeddedness, independently of effects associated with establishing a starting point and independently of number-of-runs.

Hypothesis 2: When iterating sequences are given from a particular starting-point, reorganizations tend to occur in the direction of minimizing structural embeddedness.

CHAPTER 3

THE UNDERLYING PROCESSES OF STRUCTURAL EMBEDDEDNESS

The structural embeddedness measure was derived by induction from empirical findings, not deduced from psychological processes. The question naturally arises of what those underlying processes may be, and the aims of the present chapter are to outline several possible process models. These will constitute supplementary experimental hypotheses to the major hypotheses proposed at the end of the previous chapter.

Structural embeddedness has two parameters, embeddedness and magnitude of runs. The two factors are considered in turn.

Embeddedness as Order-of-Processing

The embeddedness values of sequence items are assigned from the ends of a pattern inwards, and one psychological interpretation is that this represents the order in which items are processed. End items are processed first, and receive the value of one; items 2 and n-1 are processed second and receive the value of two, and so on, until the whole pattern has been processed. "Ends-inward" processing of this kind has been proposed in connection with serial learning (Feigenbaum & Simon, 1962; Glanzer & Peters, 1962; Harcum, 1975) and in connection with other experimental paradigms (Glanzer, 1966; Trabasso & Riley, 1975).

A basic concept of ends-inward explanations has been that of the anchor-point, the assumption that an item (or items) in a series is established first from which other items are learned. Studies by Glanzer and Peters (1962) and Glanzer and Dolinksy (1965) have provided evidence that the longer intertrial interval between successive cycles of a serial list acts as the anchor from which items are processed in order of embeddedness. Such a process would result in a symmetrical serial learning curve, and in order to introduce the degree of asymmetry typical of the serial position effect Feigenbaum and Simon (1962) proposed that the first two items are established first and second respectively, thereafter acting as anchors for a symmetrical ends-inward processing (under certain probabalistic constraints). Farwell and Vitz (1971) report results indicating that subjects are more variable in their preference than would be expected from the fixed order of processing assumed by this kind of model. However, it is possible that subjects may establish more than one anchor, on the basis of personal associations for example, and that order-of-processing models describe only an average tendency. Glanzer and Dolinsky (1965) showed that the anchor could be varied by instruction, and suggested that a variety of cues could be used in the establishment of anchor points. This suggests that idiosyncratic anchors might be used by subjects, thus accounting for departures from order-of-processing models, without necessarily discounting the basic concept of such models.

If order-of-processing is the correct interpretation of embeddedness, the problem arises of what constitutes the anchor. If a space or other cue marked the start of a cycle, then this might be expected to act as an anchor point (Glanzer & Dolinsky, 1965). The embeddedness distribution generated from such an anchor would be symmetrical, like the distributions described in the previous chapter. Typically, however, temporal patterns are presented as closed cycles. This suggests that if an anchor is used, it must be established by using an item within the sequence itself. Such an item would therefore receive the value of 1. Thereafter the two immediately flanking items, 2 and n, would receive the values of 2, items 3 and n-1 the value of 3, and so on, generating a negatively skewed distribution. If this interpretation is correct, then the calculations performed in the last chapter must be approximate only. When recomputed using the above skewed embeddedness distribution, correlations of the same order were obtained, with generally a small improvement. The average correlation for the repeating sequences used by Vitz and Todd (1969), for example, increased from .88 to .90 when a skewed embeddedness distribution was substituted for the previous symmetrical one.

It is therefore proposed that embeddedness depends on order-of-processing, and that the relevant embeddedness distribution depends on the effective anchor point within a sequence. An anchor point between items would generate a symmetrical distribution. An anchor point corresponding to one item would generate a distribution skewed by one degree. An anchor point spanning two items (as in the Feigenbaum and Simon model) would generate a distribution with two degrees of skewness, and so on. A more formal language for embeddedness distributions of this kind is proposed below, following which specific hypotheses are proposed.

Embeddedness Distributions

In any finite sequence all items have two serial neighbours, except for two items, which have one serial neighbour. These are the end items, denoted by r_1 and r_n. It is assumed that in processing a sequence, processing begins with an end item, either r_1 or r_n or both r_1 and r_n. Such items are designated anchors. Items are numbered in order of processing, thus either $r_1 = 1$ or $r_n = 1$ or both $r_1 = 1$ and $r_n = 1$. When $r_1 = 1$, then r_n can take on values from 1 to n. When $r_n = 1$, r_1 can take on any value from 1 to n.

The above assumptions concern the first and last items. The remaining items are given values according to the following rules. Let $r_1 = 1$ and $r_n = b$. Then the value of the i^{th} item is given by $r_i = i$ or $r_i = n - i + b$, whichever value is less. This is equivalent to assuming that two items are processed at a time, one from each end of the list, working inwards symmetrically from $r_1 = 1$ and $r_n = b$. The result is a negatively skewed distribution if $b > 1$. For positively skewed distributions, let $r_1 = a$ and $r_n = 1$, $a > 1$. Then the value of the i^{th} item is given by $r_i = n - i + 1$ or $r_i = a + i - 1$, whichever value is less. Such values are said to represent the embeddedness of items within sequences.

Let e_i be the embeddedness value of the i^{th} item in a sequence. When $e_1 = e_n = 1$ the resulting distribution is symmetrical, with $e_2 = e_{n-1} = 2$, $e_3 = e_{n-2} = 3$, etc. The values of individual items depend on whether or not they occur after the centre item of a sequence. When an item occurs before the centre, then $i \leq (n+1)/2$, and $e_i = i$. When $i > (n+1)/2$, then $e_i = n + 1 - i$.

When $e_1 \neq e_n$ the distribution is said to be skewed. The degree and direction of skewness is given simply by $e_1 - e_n$. When $e_1 > e_n$ the distribution is positively skewed, meaning that the designated first item is processed later than the designated last item. When $e_1 < e_n$, the distribution is negatively skewed, with the first item processed sooner than the last.

Let $e_1 - e_n = \pm k$, so that the distribution has k degrees of skewness, and the direction of skewness is specified. Then the *shape* of the distribution can be denoted by ${}_nE_{\pm k}$ where the symbol E indicates that an embeddedness distribution is being referred to. This means that an embeddedness distribution is uniquely specified by three pieces of information, the number of terms, and the degree and the direction of skewness.

The development of a formal terminology here is not crucial for sequential processing tasks, where the number of items is usually few enough that the distribution can be enumerated. However, it will be apparent from the previous discussion that the results of serial learning experiments may also be described in terms of embeddedness distributions, and it will be shown in the final chapter that different serial learning curves can be described by the formula ${}_nE_{\pm k}$. That is, instead of plotting a serial position effect over 12, 14 or more items depending on the task, the results can be economically described by two numbers only, n and k.

For sequential processing, n rarely exceeds 6 runs, and k typically takes on the value of 0 or -1.

Supplementary Hypotheses

The embeddedness value of a run corresponds to the order in which it is processed, with items valued one processed first, items valued two processed second, etc. When an anchor point is given, an E_0 distribution is appropriate. When the sequence is closed cycle, E_{-1} is the relevant distribution.

The Interpretation of Magnitude of Runs

Structural embeddedness was defined in terms of two components as:

$$E = \sum_{i=1}^{n} e_i \cdot r_i$$

where e_i is the embeddedness of the i^{th} run and
r_i is the magnitude of the i^{th} run.

It was proposed above that embeddedness is a quantification of the order in which runs are processed, with more embedded runs being processed at a later point than less embedded runs. The problem remains of interpreting r_i, the run magnitude factor. There seem to be two plausible alternatives.

Embeddedness is a difficulty factor, in the sense that the more embedded an item the later will it be processed and consequently the more errors will occur in that position, under an anticipation learning paradigm. Run magnitude may be a difficulty factor also, in that the larger a run, the longer it takes to process or construct. Having long runs in the most embedded positions has a multiplicative effect, increasing the trials required to complete processing. This interpretation implies that the embeddedness-of-runs measure applies from the beginning of processing to the point at which all runs have been formed.

A second alternative proposes that larger runs are more salient or outstanding than shorter runs, and for this reason tend to be processed sooner. This would mean that both embeddedness and run magnitude are factors which can affect the date-of-processing of a run. When the largest runs are in the least embedded positions the two order-of-processing factors are compatible, and the subject will tend to process a sequence in an orderly fashion. When largest runs are in the most embedded positions the two factors are incompatible, and a sequence may be processed as a series of isolated clumps. This might be expected to delay the process of acquisition. This interpretation gives embeddedness and run magnitude opposite signs. Early processing results from small embeddedness values and large run magnitudes. The variables in the structural embeddedness calculation might then be interpreted to be of different signs, resulting in a negative value in the overall measure. A further implication of this interpretation is that embeddedness-of-runs applies *after* runs have been formed. A larger run can be more salient than a smaller run only after it has been formed into a run. This in turn implies that processing takes place after run formation, and that runs, once formed, are processed in order of embeddedness.

The two interpretations of the run magnitude factor will be referred to as the "effort" and "salience" hypotheses respectively. There were no firm a priori grounds for preferring one interpretation over the other, and they were forwarded as empirical questions. This resulted in a total of three hypotheses and one research question to be tested experimentally. These were:

1. The complexity of a sequence is a function of its structural embeddedness, independently of effects associated with establishing a starting-point and with number-of-runs.

Experimental support of this hypothesis is necessary to prove the effects of structural embeddedness in a situation where the alternative anchor hypothesis is untenable.

2. When iterating sequences are given from a particular starting-point, reorganizations tend to occur in the direction of minimizing structural embeddedness.

3. The embeddedness value of a run corresponds to the order in which it is processed, with items valued one processed first, items valued two processed second, and so on. When an anchor point is given, a symmetrical embeddedness distribution applies. When the sequence is closed cycle, a skewed embeddedness distribution applies.

4. The magnitude of a run reflects either the effort of forming that run, *or* the salience of that run once formed.

A new experimental method was introduced in testing the hypotheses. To ease the burden of describing a new method and particular experiments employing that method simultaneously, the following chapter provides a general description of method, together with methods for the analysis and categorization of responses. Following this the sequential processing experiments are reported.

CHAPTER 4

A METHOD FOR TRACKING SERIAL ANTICIPATIONS

Method

Apparatus. A display mechanism presented sequence elements one at a time. An essential feature of this mechanism was the linkage which existed between the display and response mechanisms, so that the display advanced one position each time a response was made. This was done by modifying a standard typewriter.

Stimulus sequences were typed on display cards which could be mounted to a bracket fixed to the typewriter carriage. This meant that whenever a key was pressed the carriage advanced one position and carried the display card with it. A screen concealed all of a sequence except for one element which appeared in a small display aperture.

In typing the stimulus sequences a single space was left between elements. This meant that after a response to a letter the display advanced to show a blank in the aperture. At this point a neutral response was required. This involved pressing the space bar at a central point, which advanced the mechanism once again, bringing the next letter into view.

The letters used were D and K. These were equidistant from the neutral response key, lying to the left and right respectively. In the triangular shaped area lying within these three points, a barrier was placed. This had the effect that in the event of a movement in the wrong direction, a corrective movement had to be made back through the neutral point and then out to the appropriate key. Figure 1 shows a schematic diagram of this arrangement.

Two types of measurement were taken. A video camera suspended over the keyboard recorded all finger movements, while an event-recorder wired to the response keys tracked the time intervals between responses.

Procedure

Subjects were asked to type two baseline sequences and two experimental sequences. The first baseline required the sequence DDKK to be typed repeatedly with a space between letters. Only the index finger of the preferred hand was used. Instructions were given to type as fast as was comfortable without making errors. This familiarized subjects with the letter-space combination of movements. In addition it provided a baseline measure of what are referred to here as "coded responses." These are defined as responses requiring no stimulus pick-up from the display.

The second baseline condition required subjects to copy a 35-element sequence from the display. The sequence had been generated randomly, and familiarized subjects with the process of copying a sequence from the display. Finally,

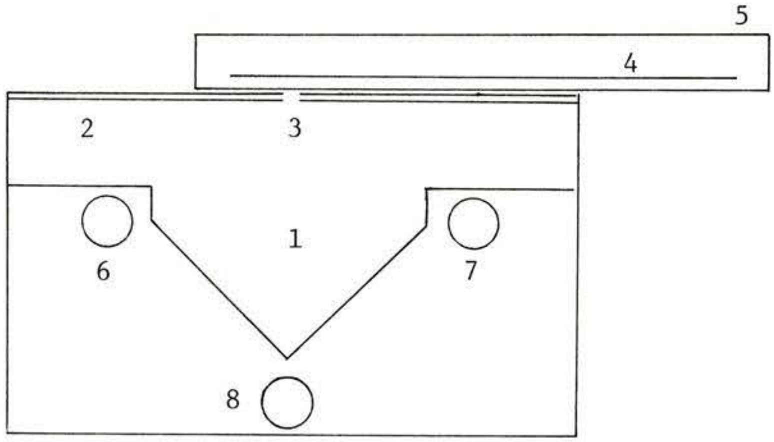

Figure 1: Schematic diagram of response/display mechanism

Key:
1. Barrier
2. Screen
3. Aperture
4. Display Card
5. Moveable Carriage
6. "D" Response Key
7. "K" Response Key
8. Neutral Response Key

subjects reproduced two patterned sequences. This part of the procedure will be described in detail in the next chapter.

Response Classification

Following the neutral response subjects frequently made anticipatory movements towards one or other of the response keys. Movements of this kind were recorded by video camera and identified later, using the following criteria. The initial displacements of all movements were measured by noting positions on a transparent co-ordinate system mounted to the video playback screen. Positions were recorded to the nearest centimetre, using the centre of the subject's fingernail as a reference point. Any immediate, directed movement of 1 cm or more following the neutral response was classified as an anticipatory movement. Movements towards the correct and wrong keys were designated as correct and wrong anticipations respectively. Hesitations over the neutral key prior to a directed movement were designated as copy responses.

The reaction time data were not sufficiently reliable under these conditions to provide a precise classification of responses. Nevertheless, it provided evidence of the validity of the visual response classification.

The time interval between the offset of the neutral response and the onset of the next response was measured, for each response. (This is the period of time during which the stimulus element is in view.) The accuracy of these reaction time measurements was probably limited by the error of the distance measurements made between the relevant spike components. This was done by hand, measuring to the nearest millimetre or, in time measure, to the nearest 50 msecs. Consequently errors as great as 25 msecs may have been made. The rise time of the recorder pens was estimated to be less than this, at about 20 msecs. Error of measurement was therefore taken to be in the order of 25 msecs. This represented 3.5% of the mean reaction time.

Figure 2 shows the distribution of reaction times taken from the two baseline conditions and from the patterned sequences. The responses were produced by 26 subjects and represent about 7,000 reaction times. It can be seen that there are four peaks. These correspond to the four different types of response, wrong anticipations, copy responses, correct anticipations, and coded responses.

Figure 3 is a breakdown of the total distribution into three component distributions. The distributions have been trimmed by 2% of scores from both extremes, but otherwise no smoothing has been carried out. The wrong anticipation and copy response distributions are unimodal, and the skewness reflects a fairly sharp cutoff at the lower bound. This was to be expected since there must be a fixed lower time limit for carrying out the component operations of

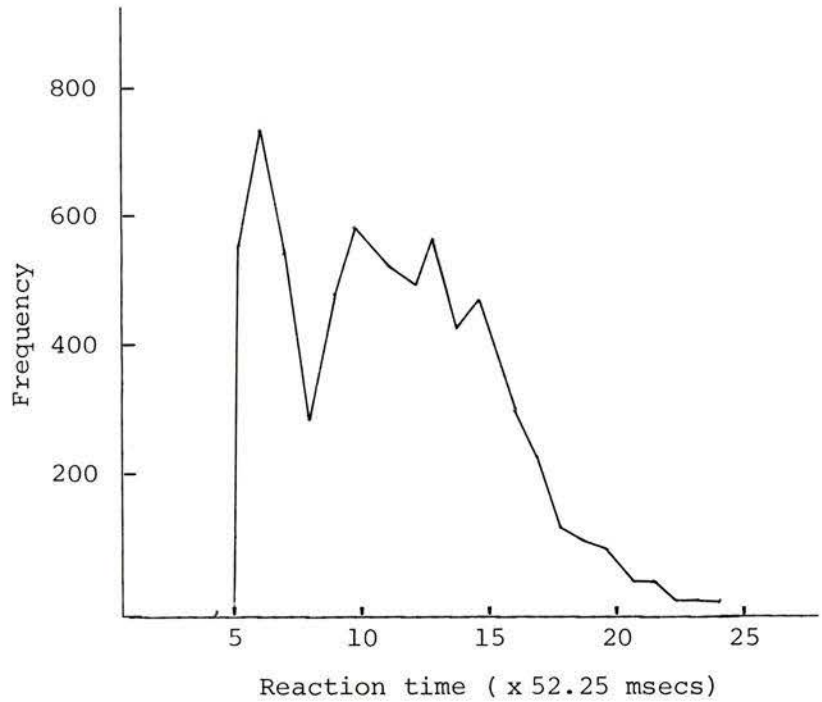

Figure 2: Frequency distribution of all reaction times.

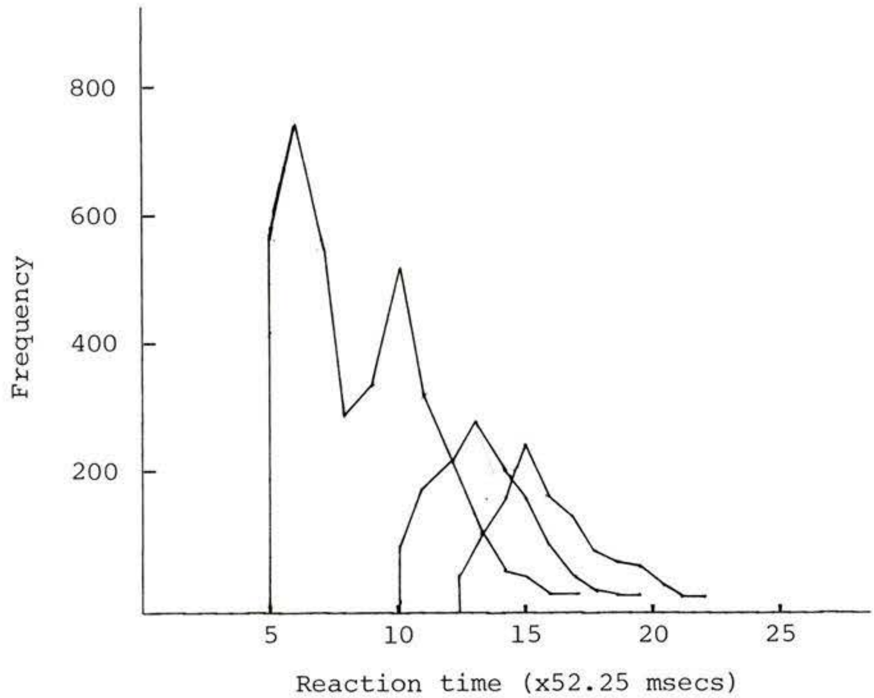

Figure 3: Frequency of different reaction times for responses visually classified as correct anticipations, copy responses and wrong anticipations, from left to right respectively.

types of response. In a copy response, for example, the subject has to wait to process the stimulus, then selects and executes a response. This requires a visual discriminative reaction time followed, probably, by a visually guided movement involving a second visual reaction (motor errors were so rare--0.2%--that most subjects must have looked down). Finally, time is required for the movement itself. It is unlikely that this sequence of operations could be performed in much less than 500 msecs. There is no corresponding upper limit, which accounts for the direction of skew.

Wrong anticipations involve an initial movement in the wrong direction, followed by stimulus registration, response selection and execution, which involves an additional movement back through the neutral response zone. The wrong anticipation response is likely to require more time since it includes all the component operations of a copy response, plus a return movement, and in addition a possible delay due to the "surprise" factor. The mean reaction times for copy responses and wrong anticipations were in fact 710 msecs and 845 msecs respectively. The difference was about twice as great as could be expected from the additional return time alone, and suggests that the surprise factor may indeed lead to additional processing time.

The third distribution in Figure 3 represents reaction times for responses visually classified as correct anticipations. This method of classification does not differentiate between coded responses and other correct anticipations, which may account for the bimodal nature of the distribution. Coded responses were defined as occurring without reference to the display, and the coded response rate was estimated independently from both the first baseline condition, and from performance measures taken after learning. This rate was less than the time normally required for a visual discrimination, and clearly such responses must occur in the absence of any prior stimulus processing. However, not all correct anticipations are of this type. The more conservative correct anticipation was slower, and was characterized by an initial directed movement, frequently interrupted by a pause or change of direction, followed by completion of the response. The frequently observed pauses may have represented a delay while subjects registered the stimulus. Alterations in the direction of the anticipatory movement probably mean that the subject had looked down from the display and had brought the movement under visual guidance.

The estimated coded response distribution was subtracted from the bimodal correct anticipation distribution to give an estimate of the mean latency of the pure correct anticipation. The mean reaction times for the four types of response are shown in Table 8.

Table 8: Mean Reaction Times for Four Different Types of Sequential Response

Response Description	Mean Reaction Time (Est.)
Wrong anticipations	845 msecs
Copy responses	710 msecs
Correct anticipations	525 msecs
Coded responses	275 msecs

Discussion

Responses were classified into three categories on the basis of visually observed anticipatory states. The category headings were: wrong anticipation, copy response, and correct anticipation. A further distinction was made within the last category between correct anticipations and coded responses, on the basis of reaction time.

A possible alternative explanation of the observed anticipatory states is that they were random movements unconnected with expectations regarding the next response. However, two arguments may be raised against this hypothesis.

First, the distributions of correct and wrong anticipations changed systematically with the learning of a sequence. In the case of the random sequence, and for the early cycles of patterned sequences, *both* correct anticipations *and* wrong anticipations frequently occurred. In the case of patterned sequences, however, correct anticipations became relatively more frequent, until eventually *all* responses were correct anticipations. This systematic change in the probabilities of types of response with learning is not consistent with the random movement hypothesis.

Second, when responses were classified visually, and their reaction times examined, systematic distributions of reaction time were observed which were consistent with previous findings.

Serial choice reaction time experiments have been conducted in which subjects were required to indicate expectations prior to each stimulus (Hale, 1967; Keele, 1969; Williams, 1966). A consistent finding of this type of study has been that the mean latency of correctly anticipated responses is lower than those of wrongly anticipated responses. The present results are in agreement with those findings. This provides evidence of the correctness of the visual classification. Systematic differences in reaction time of this kind are not consistent with the random movement hypothesis. If indeed the movements immediately following the neutral response were unconnected with a subject's cognitive state--which is what the random movement hypothesis implies--then only small differences in reaction time between correct and wrong anticipations would be expected. The differences would occur because of the greater distance required of responses initially in the wrong direction. The time taken to traverse the extra distance was estimated from the average return time of responses back to the neutral key to be less than 100 msecs. The observed difference between correct and wrong anticipations, however, was greater than 300 msecs. This is three times as large a difference as would be expected under the random movement hypothesis. It was concluded, therefore, that movements reflect anticipatory states, and that responses were appropriately classified into correct anticipations, copy responses and wrong anticipations on the basis of the visual data.

The distribution of reaction times for visually identified correct anticipations was bimodal. The interpretation was made that the distribution contained two types of response, correct anticipations, where the stimulus was monitored, and coded responses, where the response was executed without reference to the stimulus display. The latter type of response would clearly be consistent with a higher degree of confidence in the subject and, if always correct, with a higher degree of knowledge about the sequence. It would therefore provide a criterion of learning in addition to correct anticipations. However, the question arises of the precision of this type of classification. The differentiation between coded responses and correct anticipations on the basis of reaction time is likely to be conservative, in that while very slow coded responses may be assigned to the correct anticipation category, it is unlikely that any correct anticipations occur fast enough to be misclassified in the opposite direction. Correct anticipations were interpreted as involving stimulus monitoring; the stimulus was understood to be anticipated but checked before completion. The two-choice visual reaction time required in checking occupies about 300 msecs (Woodworth, 1938). This means that very few correct anticipations could occur at the mean latency for coded responses, which was about 250 msecs. Even if it were possible, little time would remain for a corrective movement, which means that the response would occur anyway, making it conceptually equivalent to a coded response. Such responses occur too quickly to be both checked and corrected. Consequently, if responses are always correct, they must reflect a high degree of knowledge about a sequence, and provide a useful criterion of a high degree of learning. This was one criterion used in the sequential processing experiments described next.

CHAPTER 5

EXPERIMENTAL TESTS OF THE STRUCTURAL EMBEDDEDNESS MEASURE

Experimental Hypotheses

At the end of Chapter 3, four hypotheses were proposed. They were that:

(1) The complexity of a sequence is a function of its structural embeddedness, independently of effects associated with establishing a starting point.

(2) When a sequence is given from a particular starting-point, reorganization tends to occur in the direction of minimizing structural embeddedness.

(3) The embeddedness value of a run corresponds to the order in which it is processed.

(4) The magnitude of a run reflects *either* the effort of forming that run, *or* the salience of that run once formed.

The first two hypotheses were primary. These were concerned with testing the relationship of structural embeddedness to performance under conditions where alternative explanations were controlled for. The first experiment was designed to test Hypothesis 1. Four sequences of different structural embeddedness were generated from a single sequence by varying the starting-point. This tended to keep sequence properties other than structural embeddedness constant. The sequence structure used was XXXOXXOO, the underlined elements indicating the four starting-points used. The resulting four sequences are shown in Table 9, with their embeddedness-of-runs and number-of-runs scores.

Table 9: Run Structure, Number-of-Runs and Embeddedness-of-Runs for the Sequences Used in Experiment 1

Run Structure	Number-of-Runs	Embeddedness-of-Runs
1. 3122	4	11
2. 2231	4	13
3. 11222	5	15
4. 12311	5	17

The embeddedness-of-runs score was calculated from a symmetrical distribution. (A negatively skewed distribution of one degree gives the embeddedness values of 1232 and 12332 for the 4 and 5 run sequences respectively, and results in embeddedness-of-runs scores of 15, 17, 19 and 19 for sequences 1, 2, 3 and 4 respectively.)

It can be seen from Table 9 that number-of-run values and embeddedness-of-run values both vary across sequences without being completely confounded. This allowed an opportunity to observe the possible contributions of both factors.

In order to eliminate the possibility of effects associated with establishing a starting-point, the sequences were given well-defined starting-points. The first element of a sequence was always underlined, and the space immediately following the last element always contained an asterisk. These cues appeared to have been used by all subjects, since no sequence reorganizations took place in Experiment 1.

Experiment 2 was designed to examine reorganizations and consequently no starting cues were provided. Again four sequences were used. They are shown in Table 10, with number-of-runs and *two* embeddedness-of-runs values. The higher set was computed using a negatively skewed embeddedness distribution. It was thought that this might be the relevant distribution due to circumstances associated with establishing a starting-point. The starting-point only would be learned first, followed by the second and last items, and so on, following the order described by E_{-1}.

Table 10: Run Structure, Number-of-Runs, and Two Sets of Embeddedness-of-Runs Values for the Four Sequences Used in Experiment 2

Run Structure	Number-of-Runs	Structural Embeddedness	
		E_0	E_{-1}
1. 2112	4	8	11
2. 2211	4	9	11
3. 1221	4	10	13
4. 11121	5	11	14

Experiment 1

Method

Subjects. Forty students enrolled at the University of Victoria participated as voluntary subjects. Their ages ranged from 18 to 45 years. Ten subjects were randomly assigned to each of four experimental conditions.

Procedure. The same apparatus and general procedure were used as described in the previous chapter. Subjects were required to type two baseline sequences. First the sequence DDKK was typed repeatedly, with a space between each letter. This familiarized subjects with the letter-space movement combination. Following this, subjects typed a 35-element random sequence, made up of the letters D and K. The letters were presented one letter at a time in the display window, contingent upon completion of the preceding neutral response. Subjects were instructed to work as quickly as possible, with accuracy being more important than speed. This task familiarized subjects with the procedure of copying a sequence from the display.

Finally, the patterned sequence was introduced. The sequences were of the structure described previously, and consisted of the elements D and K. For each experimental condition five of the sequences began with the letter D, five with the letter K. That is, five of the subjects in an experimental condition received one form of a sequence, the remaining five the complement of that sequence.

Subjects were instructed to copy a sequence at first but that since it repeated they would gradually learn it, and be able to reproduce it without reference to the display. The goal of the task was to continue until it was possible to type the sequence as quickly as in the first baseline condition. Subjects were told that a sequence repeated "several times" in the space of one carriage length, and that on reaching the end of the carriage they were simply to reset it, and continue. It was explained that to help them see where the sequence repeated, two cues had been provided. The arrangement of these cues was carefully described. The procedure began when it was clear that a subject understood the task.

While subjects worked on a sequence the experimenter observed the event-recorder output. When the latencies between all responses were of the order of the first baseline condition the experiment was terminated. On completion, subjects were asked to describe the sequence verbally, as a check on pattern organization. Finally, the display card was offset and subjects were asked to type the sequence from memory as quickly as possible for one whole length of the carriage. This provided a measure of performance rate at the coded response level. This was included in order to examine whether rate of performance also varied as a function of sequence complexity.

Results

Interest lay primarily in the relationship between overall response measures of sequence complexity and the structural embeddedness of sequences. Secondary interest lay in the process of pattern acquisition. It was therefore convenient to discuss results in terms of (a) complexity measures and (b) process measures. A further dichotomy emerged on the basis of the criteria of learning which were adopted. A first criterion was provided by the point at which all responses became correct anticipations; a second by the point at which all responses occurred at the coded response rate. The first criterion was referred to as "replication," since it indicated an ability to reproduce each element of the sequence from memory, given the preceding element. The second criterion was referred to as "integrated performance."

Replication appeared to be a non-arbitrary criterion in that once subjects had correctly anticipated all elements of the pattern for one whole cycle they tended to continue to do so. The overall rate of wrong anticipations prior to replication was 19%, dropping to 4% after replication. Most of this figure was contributed by the two subjects who represented the only departure from the tendency to maintain cycles of correct anticipations once attained. The correct anticipation rate prior to replication was 61%, and rose to 93% following replication. One subject from group 3 failed to reach replication in 200 trials, and was assigned the arbitrary score of 200.

Complexity Measures

The first analysis examined the number of trials to replication. No significant differences were found between groups (F = 1.67; df = 3,36; p < .190). The second analysis examined the rate of correct anticipations to replication. Responses during the first cycle were treated as guesses and ignored. Thereafter the correct anticipation rate was computed as the number of correct anticipations divided by the number of trials, for each subject, and averaged across subjects. Table 11 shows the mean correct anticipation rate for each of the four experimental conditions.

Table 11: Mean Percentage Rate of Correct Anticipations to Replication for Each Experimental Condition

Condition			
1	2	3	4
73.5	70.2	46.3	54.4

Analysis of variance indicated that there were significant differences among the means (F = 7.93; df = 3,36; p < .01). Multiple comparisons were not conducted, It was clear from inspection that the pattern of differences did not correlate with the increases in structural embeddedness across the four conditions. Differences appeared to correspond more closely with the differences in the number-of-runs between the sequences. Sequences 1 and 2 had four runs each, sequences 3 and 4, five runs.

Integrated performance was defined in terms of coded response rate, and ideally would be defined as a complete cycle of responses each at the coded rate of 250 msecs or less. However, two departures were made from this ideal criterion. Frequently a pause occurred on the first element of a pattern. The first element apparently has a special phenomenal character which causes it to be marked off in this way by a slight hesitation. Consequently reaction time to the first element was ignored. A second departure was made to allow for a certain amount of variability in behaviour. Inexplicable delays occasionally occur in serial reaction time tasks. Subjects were permitted one such delay in the present case, so long as the response had been correctly anticipated. Consequently integrated performance was operationally defined as the first cycle on which all 8 items were correctly anticipated and on which 6 of the 7 items following the first were executed at a rate of 250 msecs or less.

Two subjects from condition 4 failed to reach this criterion and were given scores based on the number of trials reached by the end of the experiment.

The mean number of trials from replication to integrated performance are shown in Table 12.

Table 12: Mean Number of Trials from Replication to Integrated Performance for Each Experimental Condition

Condition			
1	2	3	4
9.6	33.6	58.7	64.0

An analysis of variance indicated significant differences among the means. Table 13 gives a summary of the analysis.

Table 13: Summary Table of Analysis of Variance of Number of Trials from Replication to Integrated Performance

Source	*SS*	*df*	*MS*	*F*	*p*
Treatment	18,506	3	6,169	10.70	<.001
Error	20,173	35	576		
Total	38,679	38			

Differences due to treatment accounted for 48% of the variance in scores. (The convention has been adopted of reporting variance accounted for when greater than 10%. When it is not reported, the reader may assume a "weak" experimental effect.)

A Newman-Keuls multiple comparisons procedure was used to test contrasts between the means. The harmonic mean of the number of subjects within groups was used, to allow for unequal cell sizes. All differences between means except for conditions 3 and 4 were significant beyond the .05 level. Table 14 gives a summary of the comparisons.

Table 14: Differences Between Mean Number of Trials from Replication to Integrated Performance

	Condition			
	1	2	3	4
Mean	9.6	33.6	58.7	64.0
9.6	-	24.0*	49.1**	54.4**
33.6		-	25.1*	30.4*
58.7			-	5.3 ns
64.0				-

*p < .05
**p < .01

A test of trends indicated a highly significant linear trend (F = 29.66, df = 1,35, p < .001), accounting for 92% of the variation due to treatment. There was no significant quadratic component (F = 0.73, df = 1,35).

Figure 4 shows a plot of the means with the best fitting straight line, least squares solution.

There was no significant variation in the rate of correct anticipations between replication and integrated performance. This was because of the uniformly high rate of correct anticipations following replication. Wrong anticipation rate showed somewhat greater variability, but a problem arose in analyzing wrong anticipations because of the atypical rates for two subjects from condition 3. The mean rate ignoring these subjects was 3.9%. The rates for

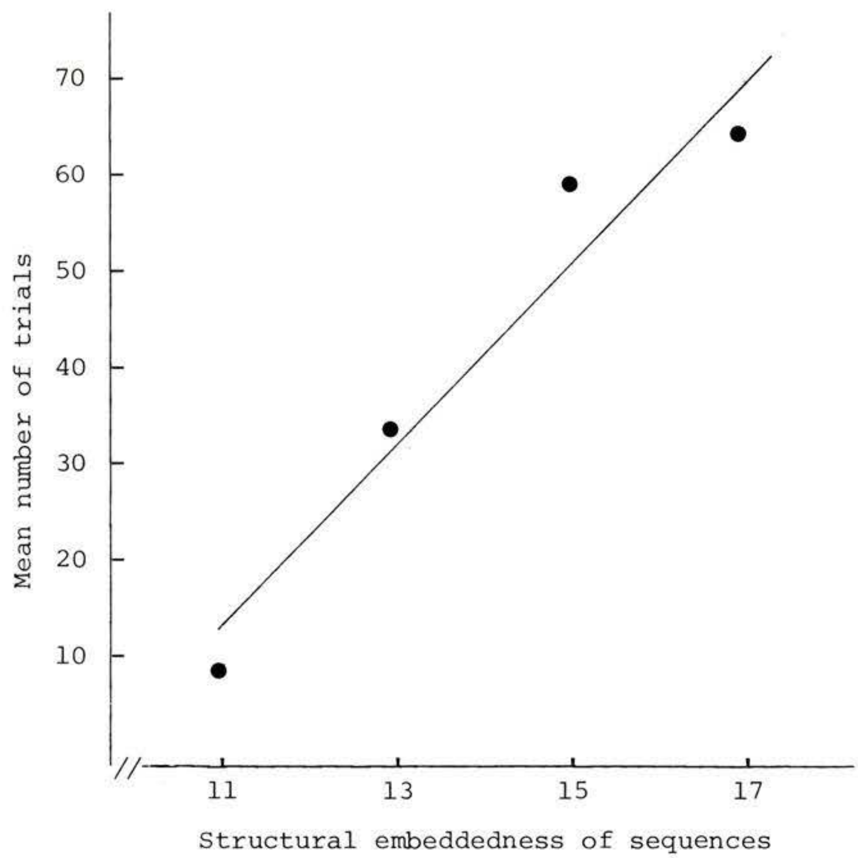

Figure 4: Mean number of trials from replication to integration of a sequence as a function of structural embeddedness.

those subjects were 17.8% and 27.8%, which were closer to the mean rate before replication (19%) than to the mean rate following replication. It was consequently assumed that replication in these two cases was not stable, and that the subjects had "forgotten" the transitions after replication. Their scores were omitted from the analysis of wrong anticipations.

Table 15 shows the mean wrong anticipation rate from replication to integration for the four groups.

Table 15: Mean Percentage Rate of Wrong Anticipations from Replication to Integrated Performance for Each Experimental Condition

Condition			
1	2	3	4
0	2.30	3.62	6.90

An analysis of variance showed significant differences among the treatment means. Table 16 is a summary of the analysis.

Table 16: Summary Table of Analysis of Variance in Wrong Anticipation Rate from Replication to Integrated Performance

Source	*SS*	*df*	*MS*	*F*	*p*
Treatment	171.9	3	57.3	5.34	<.01
Error	278.7	24	10.7		
Total	450.6	27			

(The total degrees of freedom were reduced here since, in addition to the results of the two subjects removed from the analysis, six subjects from group 1 and two subjects from group 2 first replicated the sequence at the integrated performance rate. The rate of wrong anticipations in these cases was therefore 0 ÷ 0, which is undefined. Wrong anticipation information was unavailable for one further subject because of a fault in the recording apparatus.)

Treatment effects accounted for 38% of the variance in wrong anticipation rate. Multiple comparisons between means indicated significant differences between the lowest and highest means only ($p < .01$), but a test of trends indicated no significant departures from linearity ($f = .97$ $df = 2,24$). A significant linear trend accounted for 88.4% of the variation due to treatment effects ($F = 14.1$ $df = 1,26$ $p < .001$). A plot of the means is shown in Figure 5 with the best fitting, straight line (least squares solution).

A final measure of sequence complexity was provided by the mean response time to elements from replication to integration. Table 17 shows mean reaction time per element for the four experimental conditions.

Table 17: Mean Reaction Time per Element for Each Experimental Condition (x 52.25 msecs)

Condition			
1	2	3	4
6.6	7.8	8.6	8.7

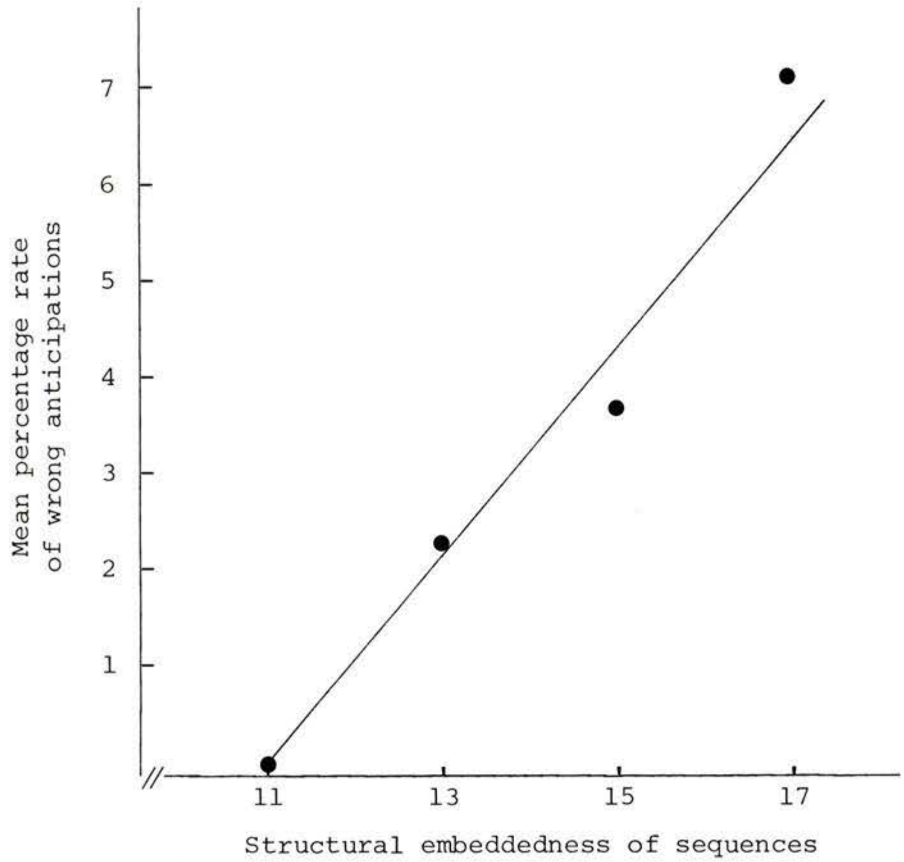

Figure 5: Mean percentage rate of wrong anticipations from replication to integration as a function of structural embeddedness.

(Reaction time is shown in distance units. Multiplying by 52.25 converts the scores to msecs.)

An analysis of variance showed significant differences among the means. Table 18 is a summary table of the analysis.

Table 18: Summary Table of Analysis of Variance of Reaction Time per Element

Source	*SS*	*df*	*MS*	*F*	*p*
Treatment	731	3	243.8	12.4	<.001
Error	29,679	1508	19.7		
Total	30,410	1511			

Multiple comparisons between means were tested using the Newman-Keuls procedure. Significant differences were found between the lowest mean and all other means ($p < .01$). A test of trends showed a significant linear component only ($F = 35.6$, $df = 1,1508$, $p < .001$). The linear trend accounted for 96% of the variation due to treatment. Figure 6 shows a plot of the means as a function of the structural embeddedness of sequences, with the best fitting straight line.

Process Measures

The analyses of process measures to be reported all followed a single basic pattern, which it will be convenient to describe in advance.

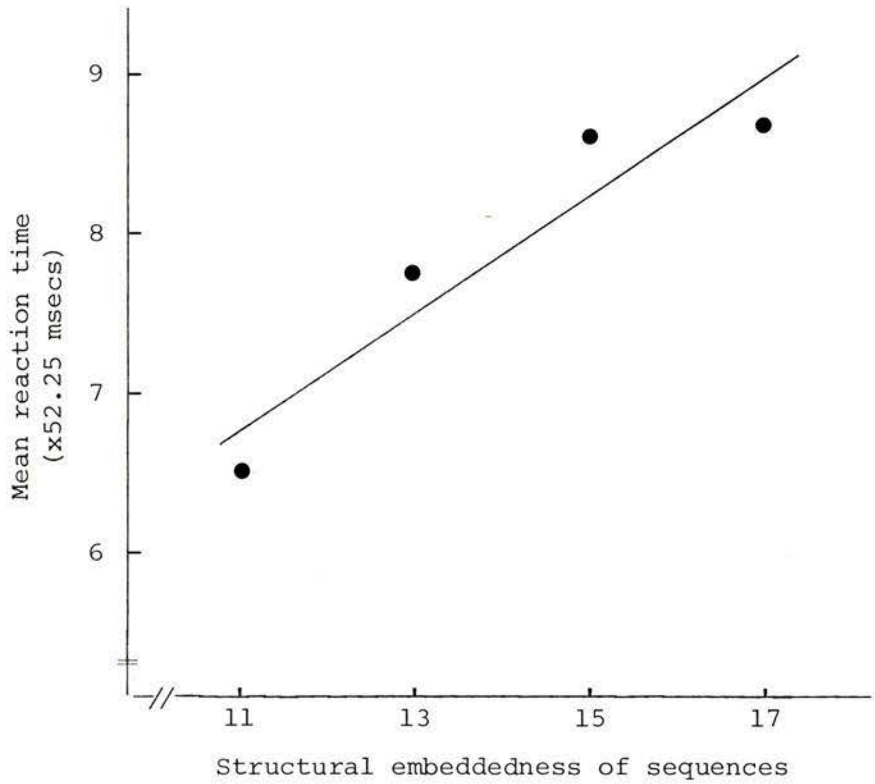

Figure 6: Mean reaction time per element (x52.25 msecs) for each experimental sequence from replication to integration as a function of structural embeddedness.

The aim of the process analyses was to track the order in which runs were processed and to examine the effects of run magnitude on processing. The nature of responses to runs necessitated, however, that internal run members and starting members be analyzed separately. Table 19 indicates why this was so. The table shows the mean number of trials required for the appearance of the first correct anticipation in each position, for each of the sequences. Parentheses have been used to mark run boundaries, the four sequences having the structures 3122, 2231, 11222, and 12311. It can be seen by inspection that values were generally lower within runs than they were for the item at the beginning of a run, indicating that within-run elements are learned more readily than are the transitions between runs.

Table 19: Number of Trials to the First Correct Anticipation in Each Position

Sequence	Sequence Position							
	1	2	3	4	5	6	7	8
3122	(3.1	2.2	2.0)	(2.4)	(3.8	2.6)	(3.1	2.6)
2231	(4.5	2.2)	(2.9	2.2)	(3.0	2.1	2.5)	(2.6)
11222	(5.0)	(4.8)	(4.7	3.7)	(3.6	2.5)	(3.5	3.1)
12311	(6.8)	(5.8	4.6)	(3.8	3.1	3.0)	(5.1)	(4.7)

The same pattern appeared in all other measures of position difficulty, and is familiar from the results of previous research, which has shown that errors tend to be lower within runs than they are at the transitions between runs (Restle & Brown, 1975; Jones & Zamostny, 1975). It seems that the first element of a run has a different character from within-run members, standing at a position of relative uncertainty compared to the relative certainty of the within-run items. Averaging across the two types of position therefore increases variability, and to avoid this, within-run members only were analyzed. Thus, triplets provided two scores, doublets a single score and singleton elements dropped out entirely. This meant that for purposes of analysis the first two sequences consisted of a triplet and two doublets (each reduced by one element); the third sequence consisted of two doublets, and one additional doublet formed by breaking the triplet; the fourth sequence consisted of the triplet and one of the two original doublets, the second having been lost due to interruption by the starting-point. The new doublet of sequence 3 was ignored since it occurred only once, which meant that for present purposes the sequence structures were the triplet, and two doublets. These were designated XXX, OO, and XX.

Next, the embeddedness indices of these structures were determined for each of the four sequences. Their values are shown in Table 20. It can be seen that while the runs XXX and XX were represented once in each embeddedness position the OO doublet appeared once in an $e_i = 1$ position and three times in an $e_i = 2$ position. This led to unequal

Table 20: Embeddedness Values of Selected Runs from the Four Experimental Sequences

Sequence	Run		
	XXX	OO	XX
3122	1	1	2
2231	2	2	1
11222	-	2	3
12311	3	2	-

cell frequencies when two-way analyses of variance were conducted. The general format of the analysis is shown in Table 21; the table also shows the maximum number of subjects represented in each cell.

Table 21: Format of Two-Way Analysis of Variance for Examining the Effects of Run Embeddedness and Run Magnitude. Cell Entries Show the Maximum Number of Subjects in Each Cell

Run Magnitudes	Embeddedness		
	1	2	3
XXX	10	10	10
OO	10	30	-
XX	10	10	10

Although runs OO and XX had the same magnitude, scores were not collapsed across them. The possibility existed that qualitative factors and not simply run sizes were important, and the distinction between the doublets was therefore retained. Consequently 3x3 analyses of variance were performed. The method of fitting constants was employed. This takes into account the unequal cell frequencies and is unaffected by the completely empty cell. The only consequence of the empty cell is that degrees of freedom for interaction had to be reduced by one (Snedecor, 1950). This analysis has been used throughout the following section, except in one instance, where heterogeneity of cell variances became extremely large, and a Kruskal-Wallis non-parametric analysis was applied. Finally, some comments are necessary concerning the effects of departures from homogeneity of variance in the present case.

Although F is reportedly robust to departures from homogeneity of variance when cell frequencies are equal, this is not so in cases of unequal cell frequencies (Lindman, 1974). Mean square error is a weighted average of the sample variances, and therefore greater weight is given to the variance estimates supplied by cells with larger n. The pooled variance is therefore biased in the direction of the larger frequency cells. Thus when the largest n group has comparatively low variance, mean square error is too low, and F tends to be overestimated. When the largest n group has comparatively high variance, mean square error is relatively too large, and F tends to be underestimated. This latter condition is the one which applied in the present analyses. The central cell, with n = 30, always had the greatest variance. This means that the significance figures reported here have a conservative bias. Tables published by Hsu (1938) permitted an estimate of the extent of the bias, and indicated that the true alpha level may have been smaller than the alpha levels reported by a factor of 10.

The first process analysis examined the effects of run embeddedness and run magnitude on the number of cycles to the first correct anticipation of a run (in the case of the triplet, both members had to be correctly anticipated). Table 22 shows the mean number of cycles to the first correct anticipation of a run, as a function of the type of run and run embeddedness.

Table 22: Mean Number of Cycles to First Correct Anticipation of a Run as a Function of Type of Run and Run Embeddedness

Type of Run	Embeddedness		
	1	2	3
XXX	2.2	2.5	3.4
OO	2.7	3.1	-
XX	2.2	2.6	3.7

An analysis of variance indicated no significant differences among the means, but this was a case where extreme cell heterogeneity obtained ($F_{max} = 25.06$, $p < .01$). The cell means showed a consistent increase in values across levels of embeddedness, though there were no strong indications of an interaction or run magnitude effect. Consequently scores were collapsed across type of run and then the number of cycles to correct anticipation ranked. Table 23 shows the mean rank order of correct anticipations within runs as a function of embeddedness only.

Table 23: Mean Rank Order of Occurrence of Correct Anticipations of Runs, as a Function of Run Embeddedness

Run Embeddedness		
1	2	3
41.3	49.12	69.0

A Kruskal-Wallis analysis of variance showed significant differences among the means ($\chi^2 = 18.32$, $p < .001$). A method of testing contrasts using rank sums was applied (Dunn, 1964), and indicated a significant difference between the second and third means ($p < .05$). The results indicated that the most embedded runs required significantly more cycles in order to be correctly anticipated. In general, the trend of the results indicated that the more embedded a run, the more trials are required to learn it.

A second analysis used the rate of wrong anticipation as dependent variable. Table 24 shows wrong anticipation rate (up to replication) as a function of embeddedness and

Table 24: Mean Percentage Rate of Wrong Anticipations Within Runs, as a Function of Run Embeddedness and Type of Run

Type of Run	Run Embeddedness		
	1	2	3
XXX	3.1	15.8	21.3
OO	17.0	14.8	-
XX	10.8	22.4	39.1

run size. It can be seen that rate of wrong anticipation increased with embeddedness level in two of the three cases. In addition, wrong anticipation rate was generally higher for doublets than for the triplet. Table 25 is a summary of the analysis of variance.

Table 25: Summary Table of Analysis of Variance of Wrong Anticipation Rate as a Function of Run Embeddedness and Type of Run

Source	*SS*	*df*	*MS*	*F*	*p*
Embeddedness (E)	45.7	2	22.9	3.14	<.05
Run (R)	17.3	2	8.7	1.19	ns
E x R	12.6	4-1	4.2	.57	ns
Error	662.8	91	7.3		

Only the embeddedness factor was significant. However, bearing in mind the conservative nature of the analysis, the runs factor may have had a significant effect. A parallel analysis of the correct anticipation rate showed both factors to be significant. (It should be noted that correct and wrong anticipation rates are not merely complementary because of the degree of freedom due to the copy response alternative.) Table 26 shows mean percentage correct anticipation rate prior to replication as a function of embeddedness and type of run. The rate of correct anticipation decreased with

Table 26: Mean Percentage Rate of Correct Anticipations as a Function of Run Embeddedness and Type of Run

Type of Run	Run Embeddedness		
	1	2	3
XXX	94.2	82.5	76.6
OO	73.9	70.2	-
XX	78.4	59.9	44.9

increasing embeddedness, and was lower in doublets than in the triplet. Analysis of variance showed both factors to be significant. Table 27 gives a summary of the analysis.

Table 27: Summary of Analysis of Variance of Percentage Rates of Correct Anticipations as a function of Run Embeddedness and Type of Run

Source	*SS*	*df*	*MS*	*F*	*p*
Embeddedness (E)	61.5	2	30.8	4.1	<.05
Run (R)	86.7	2	43.4	5.7	<.01
E x R	11.5	4-1	3.9	0.5	ns
Error	687.4	91	7.6		

Figure 7 shows both wrong anticipation rate and correct anticipation rate across the three embeddedness levels, for each of the three types of run. The correct anticipation scale in the lower part of the figure has been inverted to give the scales the same direction. The two figures are almost perfectly congruent and show that the more embedded a run, the more difficult it is to process. In addition, doublets tended to be more difficult than the triplet. There was only one reversal to these trends in each case.

The preceding analyses focussed on processes of sequence construction prior to first replication of a sequence. The final two analyses examined processes subsequent to replication. Following replication there was little variation in type of response, the majority of responses being correct anticipations. For this reason, reaction time was used as the most informative measure of performance. Two dependent variables were examined: first, the number of cycles required after replication for the first appearance of coded response rates within runs; second, mean reaction time for responses within runs, from replication to integrated performance.

Table 28 shows the mean number of cycles from replication for the first appearance of coded responses within runs.

It can be seen by inspection that the number of cycles required increased across levels of embeddedness, except at one point. This was the second embeddedness value of the

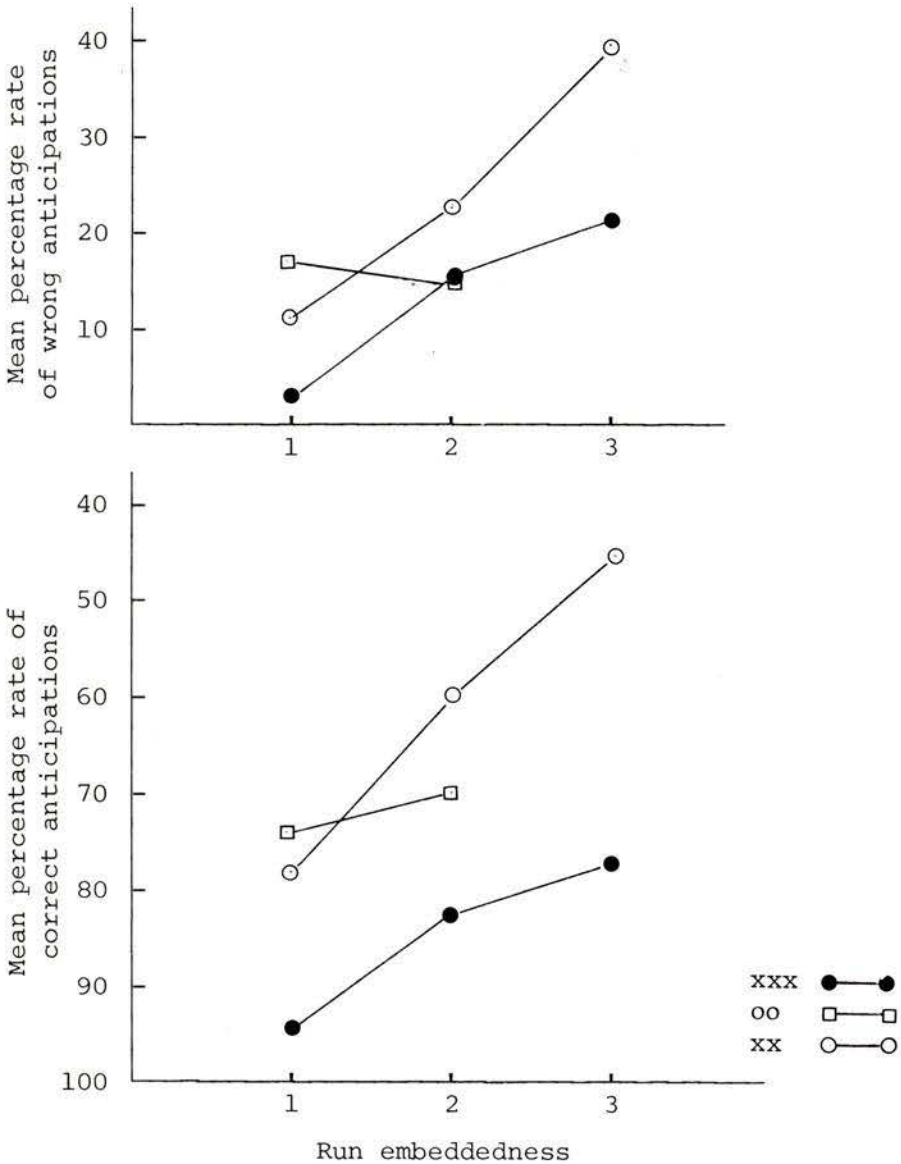

Figure 7: Mean percentage wrong anticipation and correct anticipation rates for three types of run as a function of run embeddedness.

Table 28: Mean Number of Cycles from Replication for First Appearance of Coded Responses as a Function of Run Embeddedness and Type of Run

Type of Run	Run Embeddedness		
	1	2	3
XXX	.50	2.10	3.50
OO	.70	2.64	-
XX	2.56	1.22	3.75

XX run. The extremely low value at this point was probably due to the fact that the XX run took on this value in the first sequence. Sequence 1 was so simple that six subjects reproduced it at the integrated rate on replication. This reduced mean values for all runs from that sequence.

Analysis of variance showed that only the embeddedness factor was significant. Table 29 gives a summary of the analysis.

Table 29: Summary Table of Analysis of Variance of Number of Cycles to First Appearance of Coded Responses Within Runs

Source	*SS*	*df*	*MS*	*F*	*p*
Embeddedness (E)	62.5	2	31.2	4.3	<.05
Run (R)	4.6	2	2.3	0.3	ns
E x R	36.6	4-1	12.2	1.7	ns
Error	589.9	81	7.3		

The final dependent measure was the mean reaction time to run members for responses from replication to integrated performance. The less difficult positions should show lower mean values during this phase. These values are shown in Table 30.

Table 30: Mean Reaction Time (x 52.25 msecs) to Run Members from Replication to Integrated Performance as a Function of Run Embeddedness and Type of Run

Type of Run	Run Embeddedness		
	1	2	3
XXX	4.97	5.84	6.50
OO	5.77	7.79	-
XX	7.43	5.48	8.11

Table 30 shows the same pattern of results as Table 28. The results of both indicated that performances were poorer in the case of more embedded runs, and that doublets were generally more difficult to integrate than the triplet. Again, the only reversal to these trends occurred with the XX run in the second embeddedness position.

Analysis of variance showed significant main effects for both factors and a significant interaction. Table 32 is a summary of the analysis.

Figure 8 shows a plot of the means from the two preceding analyses. The two sets of curves are almost perfectly

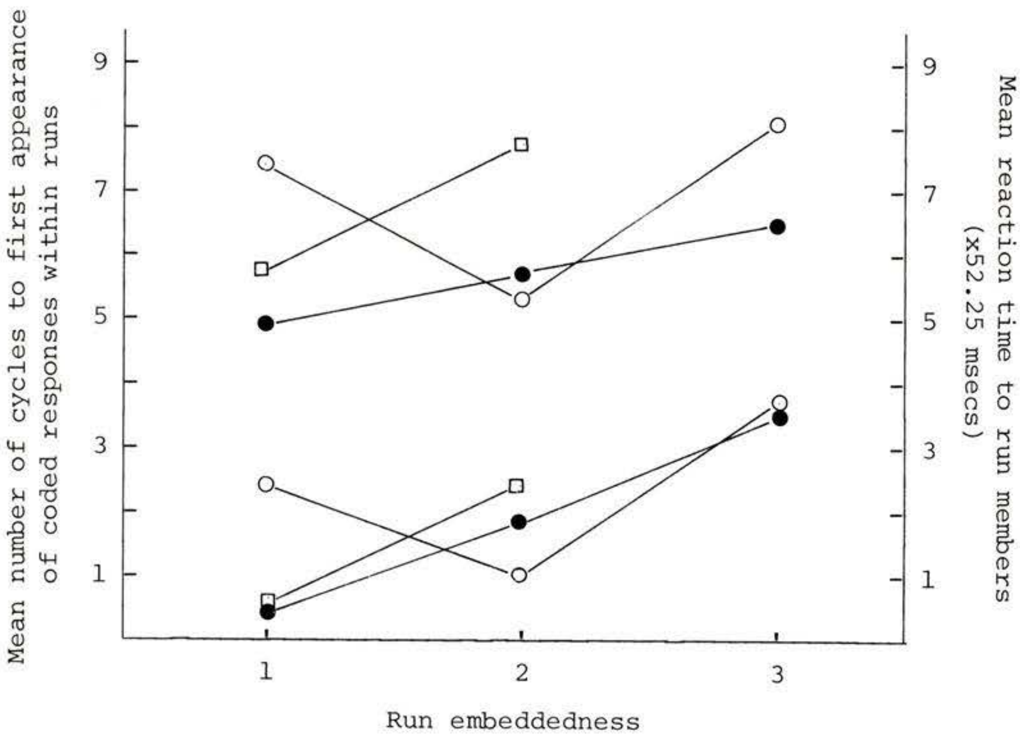

Figure 8: Mean number of cycles from replication to first coded response of a run, and mean reaction time to run members, for three types of run, as a function of run embeddedness.

Table 31: Summary Table of Analysis of Variance of Reaction Times to Run Members

Source	*SS*	*df*	*MS*	*F*	*p*
Embeddedness (E)	126.8	2	63.4	5.84	$<.005$
Run (R)	383.8	2	191.9	17.69	$<.001$
E x R	135.6	4-1	45.2	4.17	$<.01$
Error	6,464.8	596	10.9		

congruent, and in general indicate increasing complexity across levels of embeddedness, with doublets being more difficult to process than the triplet. The one reversal to these trends may have been due to the fact that the anomalous point represented performances from sequence 1, where most subjects required no trials beyond replication to reach integrated performance. Consequently mean scores from replication to smooth performance tended to be extremely low for all runs from that sequence.

Discussion

The hypothesis tested was that the complexity of a sequence varies as a direct function of its structural embeddedness, independently of starting effects or of code-length. The results of Experiment 1 supported this hypothesis. *Sequence complexity, as measured by three separate performance variables, was found to vary systematically* across levels of structural embeddedness, under conditions where the possible alternative explanations could not apply.

The further control, that all experimental sequences had been generated from a single parent sequence by varying starting-point, meant that all pair-wise transitions between elements were kept constant across conditions. This further strengthened the evidence supporting embeddedness-of-runs as the explanatory variable.

However, an important qualification must be added concerning the stage of processing to which the embeddedness measure applies. Two such stages were identified, learning to replication and learning to performance integration. The results suggested that the complexity of the first of these stages depends on number-of-runs, of the second, on embeddedness-of-runs. A simple interpretation of this finding is that the first stage of processing was concerned with the formation of runs, and so depended largely on the number of runs to be formed. The second stage involved the further processing of runs into an integrated higher-order unit, and depended largely on their structural embeddedness. This interpretation implies that runs are formed during the first stage and are stable structures by the end of that stage. A number of findings suggested this to have been the case.

If runs had not formed during the first stage of processing, then there is no reason why performance to run members should have been different from performance to other elements. However, the results indicate that within-run elements were processed more readily prior to replication.

Table 19 showed that in all cases within-run members were correctly anticipated sooner than elements at the start of the runs. All other dependent measures exhibited the same pattern.

The results indicate that runs were forming prior to replication. There is also evidence that they had formed and were stable structural entities, either with replication or shortly thereafter. Following replication, the rate of wrong anticipations dropped to 4%, while the correct anticipation rate increased to 93%. (The remaining 3% were copy responses.) This means that runs were almost perfectly predicted following replication, which is all that is meant here by "stable."

The evidence suggests that runs were stably formed with replication. It seems that the subjective representation of a sequence at this stage may have been that of an ordered assembly of runs, and that replication itself represented the first correct anticipation of the transition between runs.

One reason for describing the cognitive representation of a sequence as an "assembly of runs" at this point is that the links within runs were apparently still stronger immediately after replication than the links between them. This is shown by the fact that within-run members were reduced to a coded response rate in significantly fewer cycles than were between run members. (The means were 1.7 and 3.2 cycles respectively, $F = 13.6$, $df = 1,233$, $p < .001$). Also the mean reaction time to within-run members was lower over this stage of processing than the mean reaction time to elements transitional between runs (340 and 510 msecs respectively). Both results suggested that, following replication, elements are integrated into coded response units and that this integration takes place first within runs and then between runs. It is possible that further stages exist within this stage, and that runs are formed into progressively larger units leading up to integration of the whole pattern. However, this kind of fine observation lay beyond the limits of the present analysis. Nevertheless, it seems clear that integration takes place and that what was little more than a set of pair-wise associations between runs at the time of replication was a single unified structure at the time of integrated performance. This process appears similar in many features to the processing model proposed by Vitz and Todd (1969).

If sequences are processed through several stages of unit-formation, as the previous analysis suggests, then at different stages there may exist what are, effectively, different sequences. At the start of processing the sequence is a set of elements. At a later point it is a set of runs. At a still later point it is a single pattern. Given this, there is no reason to suppose that "sequence complexity" remains constant throughout processing, and different measures of complexity may apply at different levels of processing. This idea has already been proposed on theoretical grounds. The Vitz and Todd measure, H_{code}, is the sum of the different complexity measures of different levels of processing. Simon (1972) and Greeno and Simon (1974) have also proposed that different types of processing may require different measures of sequence complexity. So far as I am aware, the present results provide the first experimental verification of these ideas in which different levels of processing of a sequence were observed to be described by different theoretical measures of complexity. The complexity of processing elements into runs appeared to depend on the number-of-runs measure, the complexity of then integrating them into a single unit, on embeddedness-of-runs.

If embeddedness-of-runs and number-of-runs are measures of the complexity of two different stages of processing, then presumably total complexity will be some function of the weighted mean of the two. This raises the question of why, in Chapter 2, sequence complexity was found to be well described by structural embeddedness only. A possible reason is that in the temporal patterning experiments examined, stimulus elements were presented at relatively fast rates, with no response required until after processing. Under such conditions runs may form very rapidly. In the remaining experiments, sequences were frequently presented spatially on printed cards (Royer, 1967; Vitz & Todd, 1969).

The experience of looking at a card of this kind is that of the runs standing out immediately, as blocks of identical items. If run formation is indeed almost immediate under such circumstances, the number-of-runs factor would contribute little to the overall complexity of sequences. In such cases, structural embeddedness alone would provide a close fit to the data. This interpretation suggests the following prediction. Subjects pretrained on runs would still have to integrate those runs. If integration requires more processing time than run formation under the conditions described, then pretraining should have little effect. No differences would be expected between naive and pretrained individuals.

The experiment clarified the roles of two a priori complexity measures. In addition, light was shed on the psychological processes underlying pattern acquisition. In particular, psychological interpretations of the two structural embeddedness factors, run embeddedness and run size, were suggested.

With respect to embeddedness, the results consistently supported an order-of-processing interpretation. Six analyses were conducted, using different dependent variables and examining two stages of processing, and in every case a significant embeddedness effect was found. The direction of the effect indicated that more embedded runs were processed later than less embedded runs. This was evidenced by the fact that less embedded runs were first correctly anticipated earlier, had lower rates of wrong anticipations, higher rates of correct anticipations, were reduced to a coded rate in fewer trials and had lower mean reaction times than more embedded runs. All of these findings are consistent with the interpretation that the less embedded a run, the earlier it is processed.

The ends-inwards order of processing indicated by the results has important implications for present models of sequential processing. All current models assume a linear order of processing, and elements are assumed to be encoded by the operation of coding rules applied in a left-to-right direction (Restle, 1970; Simon & Kotovsky, 1963; Vitz & Todd, 1969). The possibility of a serial position effect in such sequences has not been given theoretical consideration. However, a previous experimental finding suggested that different positions within sequences might be processed differently (Jones & Zamostny, 1975). The experiment was conducted to test between alternative types of sequential processing models, but the major finding was inconsistent with all such models. The finding was that errors within a sequence arose in large part from the relative positions of coding rules within sequences.

> As it turns out, rules occurring near the beginning and end of a series are easier to learn than those in the middle. (p. 306)

The results anticipated the present embeddedness interpretation of processing, and translate directly into a statement that the more embedded a coding rule, the more difficult it is to code a section of a sequence described by that rule. This previously isolated experimental finding makes perfect sense in the context of an embeddedness explanation. More embedded rules will be applied later, with fewer opportunities for success and consequently with more errors than less embedded rules. How this order of processing of items interacts with types of coding rule remains an open question. Are rules applied in reverse direction at the end of sequences, or are they only applied sooner to the ends of sequences than has been previously supposed? This is one question which will have to be answered if bi-directional processing is to be applied by coding models.

There was some evidence in the results that the ends-inward order of processing did not proceed symmetrically from both ends of a sequence. Assuming symmetrical processing, the structural embeddedness values of the four sequences were 11, 13, 15, and 17. Assuming the first run to have been processed first, giving the skewed distribution, E_{-1}, the values would have been 15, 17, 19, and 19 respectively. In some cases the latter distribution would have provided the better fit, and the possibility exists that processing proceeded inwards in both directions from the first run, rather than from the space between the last and first runs. However, the present concern was to test the general applicability of structural embeddedness rather than to make decisions concerning specific parameter values. Nevertheless, the results made it more likely that a skewed distribution would be relevant in Experiment 2, where no anchor cues were to be provided.

The second parameter of the structural embeddedness measure was run magnitude. Two possible interpretations of run magnitude were entertained. Either it reflects the effort of forming elements into runs, or it reflects the salience of those runs, once formed. The effort hypothesis predicts that structural embeddedness will apply *prior* to run formation. It further predicts that larger runs will be processed *less* readily than smaller runs. The salience hypothesis predicts that structural embeddedness will apply *after* the processing of runs, and further predicts that larger runs will be processed *more* readily than smaller ones. The results supported the salience hypothesis in both types of prediction. Embeddedness-of-runs was found to apply *after* the formation of runs, which is consistent with the salience hypothesis, not the effort hypothesis. In addition, the process analyses suggested that the larger run was in general *more* readily processed than the smaller runs. This again was consistent with the salience hypothesis and inconsistent with the effort hypothesis.

In the process analyses, the runs factor was significant in only two instances. However, the conservative bias of the analyses may have obscured other significant cases. Certainly the trends in all cases were consistent with the salience hypothesis. The triplet was consistently associated with better performance scores than the doublets. In the majority of cases it had been correctly anticipated sooner, had incurred more correct anticipations and fewer wrong anticipations, had been integrated sooner into a coded response and had a lower overall mean reaction time. All of these results suggest that the triplet was more readily processed than the doublets.

It is of interest that the pattern of differences between doublets and the triplet coincided almost exactly with the differences between high and low embedded items. This suggests the interesting hypothesis that both run size and run embeddedness affect order of processing. The results suggest that larger runs may be processed sooner than smaller runs, holding embeddedness constant, and that less embedded runs tend to be processed sooner than more embedded runs, holding run size constant. If this is true, then structural embeddedness is a measure reflecting the compatibility of two factors which independently affect order of processing. Salient features and less embedded features would both tend to be processed readily. When those factors are compatible, with salient features in less embedded positions, the result is a good pattern. When they are in conflict, with non-salient features in less embedded positions, and vice versa, the result is a poor pattern. The poet Alexander Pope defined good style as proper words in proper places. It seems now that something similar may be true of good patterns. The essence of good patterns, by this theory, consists in placing the most readily processed features in the most readily processed positions. The question arises immediately of why this should be so? Since in learning a sequence all items have to be processed eventually, why should it matter what elements are processed first? To answer this question, some further theoretical assumptions will be required.

The present theory treats a sequence as a set of runs, or items, and as a set of positions. The process analyses demonstrated that it is meaningful to consider these factors independently. A position--in terms of embeddedness--has an effect on processing which can be assessed independently of the item occupying that position. Similarly, the effects of an item may be assessed independently of position. In this respect the theory adopts a stance opposed to the associationist theory of a sequence as nothing but a set of items (Wickekgren, 1965). If item properties were held constant, processing would proceed in the direction of increasing embeddedness. This in itself is a statement of a theory of serial learning, which will be taken up at a later point.

One consequence of processing spreading from lower to higher embeddedness values is that a sequence consists always of one already processed region and one "gap," still to be processed. With processing, the former region grows and the latter shrinks until the whole sequence has been acquired. The introduction of item differences may or may not change this simple order of processing. Some items appear to be readily processed, others less readily so. The former may be thought of as facilitating processing, the latter as blocking it. When readily processed items are in the less embedded positions, processing will proceed rapidly throughout those regions. Both embeddedness and item characteristics will combine to define a particular order of processing, which reduces the gap to smaller and smaller dimensions. The gap contains the core of recalcitrant items, those most difficult to process. By the time they are reached, a great deal will have been already acquired of the structure of a sequence, and few alternatives will remain for how the final items are to be combined within that structure. However, when the positional and item factors are incompatible, processing is likely to be less orderly. A low embeddedness item may be acquired in one location, a highly salient item in another, so that the emerging sequence consists of a number of known items and a number of gaps. It is possible that the more gaps exist in a sequence under process of construction, the more sequencing errors are likely to arise. A mechanism which could produce sequencing errors under such conditions has already been proposed by MacKay (1970) in another context.

MacKay (1970) wished to explain Spoonerisms, those sequencing errors in speech where sounds from different positions may be transposed. The explanation assumes the preactivation of certain units of speech, an idea anticipated by Lashley (1951). The basic idea is that speech, and presumably other motor sequences, is preassembled into a "queue" of waiting response units which are triggered at the appropriate instant by a mechanism which scans the queue. This scanner raises each unit in turn to the level of excitation necessary to "fire." However, if for some reason a waiting unit is at a higher than normal level of excitation initially, then it may be fired prematurely. One may suppose here that the edge of the scanning beam is a gradient rather than a sharp boundary, and that the leading edge of excitatory potential may be sufficient to raise the energy level of certain elements to suprathreshold levels. In the present case, salient items may represent just the kind of element likely to be affected in this manner. Salient items in the middle of a series would therefore represent the condition most likely to lead to sequencing errors. The effect would be magnified if the additional assumption were made that gaps in a sequence have a lower resistance than items to spreading potential from the scanner. Gaps in a sequence followed by salient items would therefore represent a combination particularly susceptible to sequencing errors. This is also the combination which would arise in conditions of high structural embeddedness, according to the previous remarks. Central salient items would be processed prior to less salient preceding items, and consequently at certain stages of processing a sequence would consist of a gap followed by a salient item. The result would be the premature anticipation of the salient item. The resulting confusion might be expected to further delay processing.

In conclusion, the first experiment supported the hypothesis that sequence complexity is a function of structural embeddedness. The experiment further clarified the psychological effects of embeddedness and run size. Both factors were interpreted as factors which affect the order in which sequential items are processed. It was proposed that structural embeddedness measures the degree of compatibility of the two factors. Good patterns, with low structural embeddedness, have the most salient items in the most accessible positions. As an explanation of why this arrangement should lead to better processing, a scanning model was proposed.

Experiment 2

Method

Subjects. The 40 subjects from Experiment 1 participated in Experiment 2. A 20-minute distractor task was interpolated between the two experiments. Subjects were assigned to one of four experimental conditions on the basis of their previous experimental conditions. Those who had received complex sequences in Experiment 1 were given simple sequences in Experiment 2, and vice versa. This departure from random assignment was made to reduce the amount of time and effort required of any one subject. (Time and effort costs were considerable in the more complex conditions. Frequently 15 minutes or more were required, involving perhaps 300 trials, with 600 key presses.)

Procedure

The experiment was identical in general procedure to the previous experiment. The run structures of the sequences employed were 2112, 2211, 1221, and 11121, generated from a single sequence by varying the starting-point. (Taking the basic sequence to be XXOXOO, the four starting-points were the underlined elements.) The sequences were typed on display cards, with single spaces between all elements. Six repetitions of a sequence were typed on a display card as a single series, with no cues indicating the segmentation. Subjects were instructed to copy a sequence until able to reproduce it at the fast baseline rate. It was explained that the sequence repeated several times in one carriage length, but that no cues would indicate repetition, and part of the task was to identify this. It was then explained that on reaching the end of the carriage it might seem that the sequence had stopped somewhere in the middle. "This may happen," it was explained, "because the point where *you* see the sequence as starting is not the same as the point given at the beginning of the card. This often happens, and if it does, try to go on seeing the sequence in your own way." Finally, subjects were warned that the sequence might be longer, shorter or the same length as the sequence in the previous experiment.

Trials continued until all responses were at a rate approximating coded responses. At this point the experiment was terminated, and the subject was asked to describe the pattern verbally.

Results

Responses were classified as correct anticipations, wrong anticipations and copy responses on the basis of the visual record, and as coded responses on the basis of reaction time data. "Replication" was defined as the first six consecutive correct anticipations. "Integrated performance" was defined as the first six consecutive correct anticipations in which the last five had reaction times of

260 msecs or less. The number of trials from replication to integrated performance was found for each subject.

The verbal descriptions given by subjects at the end of processing were taken to represent the subjective final form of a pattern. The element starting a sequence on the display card defined the initial form. The number and direction of shifts from initial to final forms are shown in Table 32.

Table 32: Frequency and Direction of Shifts During Processing

Run Structure of Initial Form	Run Structure of Final Form			
	2112	2211	1221	11121
2112	7	0	0	0
2211	0	10	0	0
1221	1	4	1	0
11121	2	2	1	3

The table represents performances based on 31 subjects, not 40. Eight subjects failed to replicate, and were unable to give a verbal description of the sequence. There were two failures from group 1, four from group 3, and two from group 4. In addition, the data for one subject from group 1 were lost due to an equipment fault.

The first thing to be noted in Table 33 is that sequence 2 was at least as stable as sequence 1, and as effective in attracting responses. This was consistent with both sequences being of the same structural embeddedness, which would have been the case assuming the skewed distribution, E_{-1}. The structural embeddedness values generated from this distribution were 11, 11, 13, and 14 for sequences 1 to 4 respectively. The first analysis compared the number of trials from replication to integration for the first two sequences only. No significant differences were found (F = .073, df = 1,24, p > .5). Table 33 shows the mean and standard deviation of scores from the two groups. It can be seen that they were almost identical.

Table 33: Means and Standard Deviations of Number of Trials from Replication to Integration

	Mean	Standard Deviation
Sequence 1	40	40
Sequence 2	36	36

On the basis of this result, the assumption was made that an E_{-1} distribution was appropriate in the present case. The scores from groups 1 and 2 were therefore pooled. Groups were defined on the basis of the final form of a sequence. This was done because reorganizations were likely to have taken place prior to replication. Thus, in examining scores after replication the effective experimental conditions depended on the reorganized form of a sequence, not on its initial form.

The mean number of trials from replication to integration for the three redefined conditions are shown in Table 34.

Table 34: Mean Number of Trials from Replication to Integration for the Three Redefined Experimental Conditions

Experimental Conditions		
1	2	3
37.4	103.0	157.3

The low number of subjects remaining in the last two groups (2 and 3 respectively) raised doubts about the reliability of the results. Nevertheless, analysis of variance indicated significant differences among the means. Table 35 is a summary of the analysis. Treatment effects

Table 35: Summary Table of Analysis of Variance of Number of Trials

Source	*SS*	*df*	*MS*	*F*	*p*
Treatment	43,963	2	21,981.5	6.6	<.005
Error	93,251	28	3,330.4		
Total	137,214	30			

accounted for 32 % of the variance in performance scores.

A plot of mean trials as a function of structural embeddedness is shown in Figure 9, with the best-fitting straight line. It can be seen from Figure 9 that sequence complexity increased linearally across conditions as a

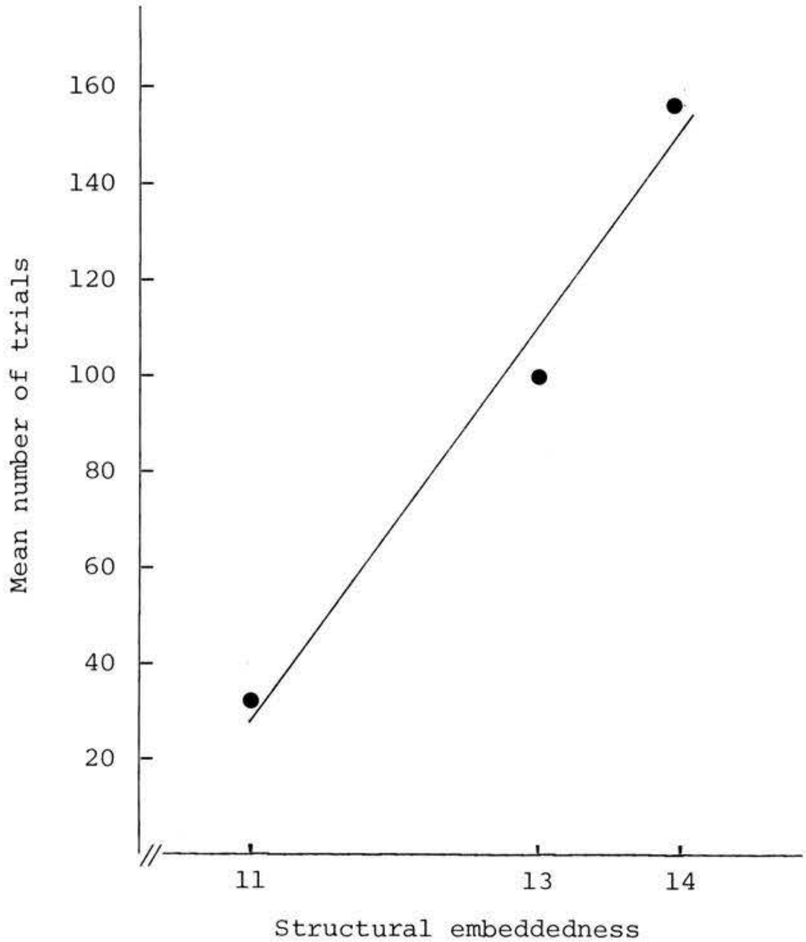

Figure 9: Mean number of trials from replication to integration as a function of structural embeddedness.

function of structural embeddedness. This replicated the finding of the first experiment. It also provided additional support for the choice of the skewed embeddedness values in the present case, since by using those values the linearity between complexity and embeddedness-of-runs was best preserved.

Organizational shifts were re-examined on the basis of the new embeddedness values. Table 36 shows the occurrence of shifts as a function of the structural embeddedness of sequences.

Table 36: Sequence Reorganizations as a Function of Structural Embeddedness

Structural Embeddedness of Initial Form	Structural Embeddedness of Final Form		
	11	13	14
11	17	0	0
13	5	1	0
14	4	1	3

Entries in the first row show that when a sequence was given in the minimum form, no reorganizations occurred. In the second row, 5 of the 6 subjects who began with a sequence valued 13, and who could describe the sequence after processing, described it in a reorganized form. In all cases the reorganizations were to the minimum form. In the third row, 5 of the 8 subjects who began with the sequence valued 14 and who could later describe the sequence, described it in a reorganized form. In this case four of the reorganizations were to the minimum form, and one to a form with a lower value than initially.

In order to test the results statistically an expected distribution must be specified under a null hypothesis. The nature of the null hypothesis depends on how many shifts were expected. Given a certain number of shifts, it is possible to specify a null distribution. In the present case, 10 shifts occurred. One null hypothesis is that shifts occur randomly, and that as many shifts are to be expected in the direction of increasing embeddedness as in the direction of decreasing embeddedness. This permitted evaluation of the experimental hypothesis, that shifts occurred in the *direction* of minimizing structural embeddedness. All 10 shifts occurred in the predicted direction. The probability of this occurring by chance, given that 10 shifts occurred, was less than .001. (Calculated by the probability of all 10 events falling below rather than above the main diagonal, $p = 1/2^{10}$. This calculation did not take into account the number of shifts within particular rows. It is to be noted that the proportion of shifts was greater in the second than in the third row of the table. This goes counter to the prediction, since the third row represented a theoretically more difficult sequence and should have resulted in more shifts.)

Discussion

The experiment was designed to test the hypothesis that when shifts occur in sequence organization they do so in a direction of minimizing structural embeddedness.

To evaluate the hypothesis, the appropriate measure of structural embeddedness had to be determined. The present situation differed from that of the first experiment in that no cues were provided this time for the segmentation of a sequence. Segmentation was something which subjects themselves had to do. It seemed probable that they would do so by establishing a run as an initial anchor point, following which the remainder of the sequence would be learned in the usual ends-inward order. This would give the first, anchoring, run a value of 1, the second and last runs a value of 2, the third and second last runs a value of 3, and so on. These are the values of the skewed embeddedness distribution, E_{-1}. Embeddedness-of-runs values computed from this distribution were found to provide a better fit to the data than values from the symmetrical distribution.

A linear relationship was again found between the structural embeddedness of sequences and the number of trials required to process a sequence from replication to integrated performance. The result generalized the major finding of the first experiment to sequences for which no starting cues were provided. Again, the experimental findings indicated that structural embeddedness is a measure of the complexity of integrating a sequence into a single pattern, once the runs have been formed.

The results also supported the main hypothesis of the present experiment. In all cases where reorganizations of a sequence occurred, they occurred in a direction of minimizing structural embeddedness. It was apparently very difficult to learn the more highly embedded sequences without first reorganizing them. Of the 20 subjects who received the two most difficult initial forms, 10 reorganized to a simpler form, 6 were unable to describe any pattern, and only 4 learned the patterns to integration in the given form.

At first sight it seems obvious why subjects reorganize sequences into simpler forms. They do so because it makes the task simpler. On closer examination, however, it can be seen that this explanation will not suffice. It is teleological and implies that subjects are motivated towards a particular end state--simplifying the task--and choose patterns on that basis. This in turn implies prior knowledge of a pattern. In effect, it assumes that a subject knows which form of a sequence is simplest to learn before having learned it.

A simpler explanation is that reorganizations are caused by wrong anticipations, which precipitate attempted reorganizations. If the attempted reorganizations also lead to many wrong anticipations, they too will tend to be reorganized. If, on the other hand, the new organizations result in a reduction in wrong anticipations, then they are more likely to be retained. Experiment 1 established that learning a sequence which is highly structurally embedded results in high rates of wrong anticipation. Consequently, it is to be expected that wrong anticipations occurred more frequently when subjects attempted to learn the present sequences in more highly embedded forms. If the assumption is made that wrong anticipations lead to a loss of material in memory storage, then the initial starting-point of a sequence may be lost and the subject obliged to pick-up a new starting-point. Such new points are likely to be salient, readily processed features. Having such features at the beginning of a pattern is likely, in turn, to lead to a reduction in structural embeddedness. As a consequence, wrong anticipations will become less frequent, loss of material from storage rarer, and the sequence will tend to stabilize in the new form. If, on the other hand, the new starting-points do not reduce structural embeddedness, wrong anticipations will continue at a high rate, and the process is likely to repeat until further reorganizations lead to a simpler form. In effect, there is a natural selection of simpler forms from the various "mutations" which arise in the course of processing.

The scanning model described in the previous chapter readily accounts for the higher rate of wrong anticipations in more structurally embedded sequences. Such sequences tend to have salient items towards the centre, being therefore of a form in which premature firing is likely to result, and so bring about sequencing errors.

One consequence of the model outlined above is that if a subject persists in trying to learn a sequence in a difficult form, the same sequencing errors are likely to recur. A preliminary test of this prediction was carried out by comparing the performances of subjects who had shifted, with those who had failed to shift, in the case of the two more difficult sequences. Persistence of wrong anticipation was measured by counting the number of times a wrong anticipation on one cycle led to a wrong anticipation to the same element on the next cycle. The mean number of such perseverations in cases of shift was 4.6, in cases of no shift, 19.0. The difference was highly significant ($F = 16.9$, $df = 1,18$, $p < .001$). The result suggests that sequencing errors tend to perseverate in highly embedded sequences, and decrease when shifts occur to a simpler form. This is consistent with the scanning model, and with the more general model of sequential processing proposed here.

In summary, the results of the second experiment supported the hypothesis that organizational shifts occur in the direction of minimizing structural embeddedness. In all cases the reorganizations which occurred were consistent with the hypothesis. In addition, the results of the second experiment provided an additional confirmation of the results of the first. The complexity of sequences from replication to integration was again found to be a direct function of structural embeddedness. The only difference was that requiring a subject to establish a starting-point leads to a slight change in the relevant embeddedness distribution. The nature of this change was in itself consistent with the order-of-processing explanation of embeddedness.

CHAPTER 6

GENERAL DISCUSSION AND CONCLUSIONS

The present investigations began with the proposal of a new quantification of sequential patterns, called embeddedness-of-runs or structural embeddedness. The new measure was derived by induction from previous empirical findings, and tests using existing data provided considerable preliminary support. The measure correlated highly with performance scores from a wide variety of experimental settings and in contrast to the most widely accepted current measure, code-length, it readily accounted both for differences in complexity across sequences, and for organizational preferences within a single sequence.

However, alternative explanations of the new measure were possible, and the major goal of the present research was to test the measure under conditions where the alternative explanations were ruled out. Two such experiments were reported, in which embeddedness-of-runs was shown still to bear a systematic relationship with behaviour under these controlled conditions. In the first experiment, the behavioural complexity of sequences was found to vary as a direct function of its structural embeddedness. In the second experiment, the preferred organizations of an iterating sequence were found to vary with its structural embeddedness.

In addition to confirming the major hypotheses, the experimental findings provided information concerning the psychological processes which might account for the measure. In effect, the measure was a mathematical model of sequential processing, with two parameters. The first of these parameters, run embeddedness, was consistently found to be interpretable as the order-of-processing of items. Prior to replication, the results indicated that runs had been formed by an ends-inward direction of processing. Following replication, it appeared that runs had been integrated into a single pattern in the same, ends-inward order. Previous research had indicated the importance of order-of-processing in serial tasks (Feigenbaum & Simon, 1962; Trabasso & Riley, 1975). It seems that when faced with a sequential task, in which limitations of capacity prevent an all-at-once kind of processing, subjects allocate processing to parts of the sequence in an orderly, systematic manner. However, in the present case, the results signified that the second parameter of the model, run magnitude, could influence this order of processing. In almost all cases the longest run was processed more readily than the shorter runs, indicating that it had been processed first.

Both parameters of the model were therefore interpretable in terms of the readiness with which parts of a sequence were processed. The embeddedness factor described the readiness with which particular *positions* were acquired. The run magnitude factor described the readiness with which particular kinds of *items* were acquired. This led to a simple conceptual statement of the emerging psychological theory. When the two factors affecting processing are compatible, with the most readily processed items in the most readily acquired positions, then structural embeddedness will be minimal, and the sequence will have a good pattern as determined by subjects' performances. When the two factors are incompatible, the result will be higher structural embeddedness, and patterns which are less good in terms of performance measures. A possible explanation of this theory was proposed, in the form of a modified version of a scanning model described by MacKay (1970). The model was designed to account for certain kinds of sequencing errors, by proposing that they are due to the premature activation of elements in a serial queue. It was proposed here that high structurally embedded patterns are likely to lead to this kind of sequencing error, and that consequently such patterns will be difficult to acquire. The same explanation served to account for the observed organizational shifts in the direction of patterns having lower structural embeddedness. The explanation took the form that sequencing errors tend to result in a loss of material from storage. In cases where subjects are attempting to establish a starting-point, this loss of material leads to the selection of new possible starting-points. Continued errors lead to continued shifts, until a starting-point is selected which leads to a drop in errors. Such points are more likely to be retained. For two reasons, such points are also likely to be those for which embeddedness-of-runs is low. First, salient starting-points are more likely to be selected. This in turn is likely to lead to low structural embeddedness. Second, the rate of sequencing errors is likely to drop when a pattern of low structural embeddedness is formed. Consequently, there will be something analogous to a "natural selection" of the simpler, less embedded forms.

It was possible, then, to propose psychological interpretations of the experimental findings. The result of this was a unified conceptual theory which was able to subtend the mathematical model, and to provide psychological explanations of pattern complexity and pattern preferences. The attempt will be made in the remaining pages to extend these ideas into two neighbouring areas, serial learning and verbal processing. Following this, some final questions will be raised about the precise psychological locus of the embeddedness measure. Does it apply only to acquisition, or also to storage and retrieval?

Embeddedness and Serial Learning

The major difference between serial pattern learning and serial learning is, appropriately enough, the absence of patterns in the latter case. Serial items are carefully selected to inhibit any tendencies to combine them into sub-patterns. Arbitrary material is selected initially, then equated for association value and finally combined into a list in a way that eliminates alliteration, rhyme or any other obvious cue for the grouping of adjacent items. It is unlikely that these elaborate precautions actually eliminate grouping effects at the individual subject level. Even Ebbinghaus was unable to prevent himself from using a trochaic rhythm for the grouping of arbitrary items (Harcum, 1975). However, when performances are averaged across subjects it is unlikely that any systematic grouping effects will remain. In other words, it is reasonable to suppose that the salience of items in a serial list is more or less constant across the items. Applying this to the structural embeddedness model, the proposition may be made that the difficulty of a serial list will be given by

$$\begin{aligned} E &= \Sigma e_i \cdot r_i \\ &= \Sigma e_i \cdot k \end{aligned}$$

where e_i is the embeddedness of the i^{th} item,
r_i is the salience of the i^{th} item, and
k is a constant.

Let the difficulty of the i^{th} position within a serial list be defined as f_i, where f is some measure of performance such as number of failures, or number of trials to criterion. Then the total complexity or difficulty of a serial list, E, is given by

$$E = \Sigma f_i \quad .$$

It follows that

$$\Sigma f_i = \Sigma e_i \cdot k,$$

$$f_i = e_i \cdot k$$

and $$f_i \propto e_i$$

Thus the number of failures in a serial position will be directly proportional to the embeddedness of that position. It has become customary to report the results of serial learning experiments in terms of the *relative difficulty* of items, given here by $f_i/\Sigma f_i$. This makes allowance for differences in item difficulty in comparing across lists, and provides equivalent serial position effects from different serial lists (McCrary & Hunter, 1953).

Since $$f_i = e_i \cdot k,$$

$$f_i/\Sigma f_i = e_i \cdot k/\Sigma e_i \cdot k$$

$$= e_i/\Sigma e_i$$

This states that the relative difficulty of a serial position is equal to the relative embeddedness of that position. This provides a very simple model of serial learning. However, nothing has been said yet about the type of embeddedness distribution appropriate in serial learning. In fact, this appears to be variable, depending on certain characteristics of the experimental situation, but by and large the classical serial position effect is quite well described by the distribution ${}_nE_{-2}$, where n is the number of items in the serial list. One example of the goodness-of-fit of the model is shown in Figure 10. Many other examples could have been selected from the literature. The data in Figure 10 were taken from Hovland (in Stevens, 1951). The figure shows the mean number of errors in each position as a percentage of the total mean number of errors. The predicted curve was obtained by calculating the embeddedness of each position as a percentage of the total of the embeddedness values. As it turns out, the model makes predictions which are very similar to those of the Feigenbaum and Simon (1962) model. (The two models diverge as list length decreases. For very short lists of eight or fewer items the embeddedness model continues to give a serial position effect, while the Feigenbaum and Simon model predicts a fairly flat distribution, except for the first two items.) Like the embeddedness model, the Feigenbaum and Simon model also assumes an ends-inward order of processing, which is the main reason why the two models give similar predictions. The advantages of the embeddedness model lie in its computational simplicity, and its greater generality. The serial model proposed here is simply a special case of the general sequential model, $E = \Sigma e_i \cdot r_i$, and it seems that this single mathematical model may underlie the areas of temporal patterning, serial pattern learning and classical serial learning.

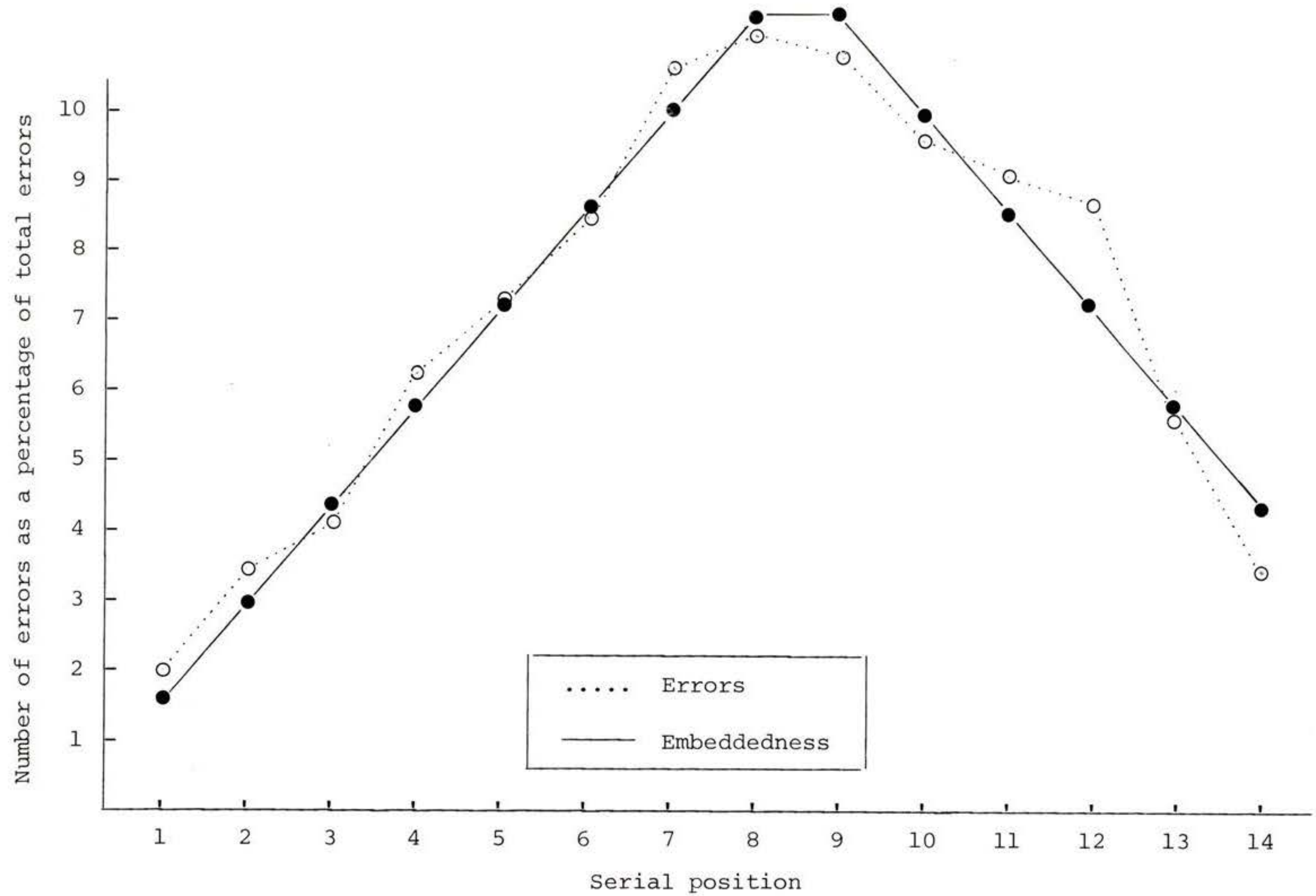

Figure 10: Relative number of errors and relative embeddedness at each position of a serial list. (Data from Hovland, in Stevens, 1951.)

A further advantage of the embeddedness model is that it provides a very simple notation for specifying the shapes of different serial learning curves. It may be recalled that the notation ${}_{n}E_{\pm k}$ specifies a particular embeddedness distribution, for given values of n and k. One example may serve to suggest the possible usefulness of such a notation.

Glanzer (1966) reports results where a serial list of stimuli were presented all at once, on printed cards, at low exposure times. Eight-item series were presented under four exposure conditions of 200, 400, 800, and 1,600 msecs. Harcum (1968) transformed Glanzer's data by expressing errors in each serial position as a percentage of total errors, the same transformation as the one used in preparing Figure 10. The resulting curves were quite similar to each other, and Harcum concluded that the serial position effects resulting from the four different exposure conditions had been identical. However, it is in cases like this where the more specific notation proposed above can provide a more precise classification of the curves.

Results from the two longest exposure conditions in Glanzer (1966) were found to correspond best to a ${}_{8}E_{-5}$ distribution. This suggests that processing proceeded in almost perfect left-to-right order, except that the last item was processed at almost the same time as the sixth item, the seventh item being processed last. Goodness-of-fit was estimated by linearizing the obtained and predicted distributions and computing a correlation coefficient between them. In both cases the correlation was .98. The next lowest exposure condition, 400 msecs, resulted in a serial position effect which was best described by the distribution ${}_8E_{-4}$. Here the correlation was .99. The shape of the distribution is consistent with a less strict left-to-right order of processing, with items 1 to 4 having been processed in order, followed by 5 and 8, processed about the same time, followed finally by items 6 and 7. The results from the lowest exposure condition, 200 msecs, show a further continuation of this trend away from left-to-right processing and towards greater ends-inward order. The results in this case were best described by the distribution ${}_8E_{-3}$, with a correlation of .98. This distribution suggests left-to-right processing up to and including item 3, following which items 4 and 8 were processed at about the same time, followed by 5 and 7, and then finally by item 6.

The results suggest the trend that, as exposure time decreases, a series of items presented all at once is processed less in a left-to-right fashion and more in an increasingly symmetrical ends-inward direction of processing. Rather than being identical, there appear to be systematic differences between the sets of curves. However, the differences only become apparent when there is a precise enough way of describing them. Does the observed trend continue, with further decreases in exposure time? The data on this are ambiguous, but the results of at least one study suggest that it does. Haslerud and Clark (1957) presented subjects with 9-letter words exposed for 40 msecs. If the responses with no grouping effects are examined (the authors' "fragmentary response" category) then the results are almost perfectly described by the distribution ${}_9E_0$, with a correlation of .99. This means that the best-fitting distribution was consistent with a completely symmetrical ends-inward order of processing in this case.

From the results examined here it appears that the embeddedness model may have some predictive power when applied to serial learning and related topics. The model accounts quite well for the shapes of serial position effects, and provides a notation for describing a family of such curves. The particular embeddedness distributions appropriate to specific cases presumably depend on particular characteristics of the situations. One factor which might be important in the "perceptual" serial position effect appeared to be that of exposure time, discussed above. This is not the place to consider psychological models for explaining the relationship of embeddedness to serial learning in specific cases. However, the growing generality of the embeddedness model seems to indicate that it may be worthwhile to pursue questions of this kind.

Salience of Items and Verbal Processing

It was proposed, as part of the theory of sequential processing developed here, that iterating sequences tend to be segmented so that salient, rapidly processed elements are in first position of a perceived pattern. One reason for this is that salient items have the best chance of being established as potential starting-points. A second reason is that, once established, they tend to lead to simpler patterns, which have a high probability of stabilizing without serious disruption from sequencing errors.

Up until now the theory has been limited to binary sequences, where the salient items appeared to be the longest runs. If, instead of binary sequences, the sequences had been composed of phonemes, there is some reason to suppose that the most salient items would have been consonant sounds as opposed to vowel sounds. Consonants appear to be more informative than vowels and written languages have existed in which vowels were omitted entirely, with apparently no undue loss of information. Arabic and Hebrew are examples, and also forms of Ogam, an ancient Celtic alphabet. It has also been demonstrated experimentally that consonants carry more information than vowels (Miller & Freidman, 1957).

Information load is one possible interpretation of salience, but in Experiment 1 salience had more to do with the speed of processing of an item than with its information content. The longest run was consistently found to have been processed more readily than the shorter runs. If speed of processing is used as the operational definition of salience, then consonants again are likely to be more salient than vowels. In an experiment where subjects were required to respond to target phonemes embedded in sequences of syllables, Savin and Bever (1970) found that consonant targets were responded to more quickly than vowel targets. This suggests that consonants may be processed more rapidly than vowels.

If it is true that consonants are more salient linguistic features than vowels, and if the present theory is correct, then iterating sequences of speech sounds will tend to be segmented into units beginning with consonants rather than into units beginning with vowels. This in turn could have implications for the segmentation of normal speech. Two unpublished experiments have been conducted which tested this hypothesis (MacGregor, 1978).

In the first experiment 20 subjects listened to 20 iterating syllables which had ambiguous starting-points. Examples of the stimulus sequences used were "... ocksocksocks ..." which could be heard as "ox" or "sock," "... endendend ..." which could be heard as "end" or "den" and "... ilsilsils ..." which could be heard as "ills" or "zill." There were 10 pairs of such ambiguous syllables, one member of each pair beginning with a vowel, the other with a consonant. The pairs were roughly equated for frequency of occurrence, with an overall bias in favour of the vowel-start forms, which had a mean frequency of 58 per million words compared with 30 per million for the consonant-start forms. The 20 stimulus words were read as iterating sequences onto magnetic tape. Each syllable was spoken at an average rate of 3 repetitions per second, and repeated an average of 21 times. During recording of the stimulus sequences the volume control was increased from zero after reading began and turned back to zero before the reading ended. This fading-in and fading-out of the sequence was done to obscure the true starting-point.

Subjects were asked to listen to each sequence and write down what seemed to be the basic repeating sound. Interest lay in whether subjects showed any systematic bias towards preferring consonant-start forms over vowel-start forms. As it turned out, there was an extremely strong bias in this direction. Ninety percent of the 400 syllables repeated began with consonants, a highly significant porportion ($Z = 15.85$, $p < .00003$). In fact the effect was significant beyond the .01 level in 19 of the 20 individual sets of responses.

It is true that consonant-start syllables of this type are more common in the spoken language than vowel-start forms, and the argument can be made that this prior exposure to a particular form was what influenced subjects' responses in the present case. This may be true, but it would seem to indicate that the relative frequency of the reported forms should at least roughly mirror the relative frequencies of the two forms in the language. This would lead to an expected ratio of about 4:1, compared with the obtained ratio of 9:1. It seems that the experimental effect was about twice as powerful as would be expected under a relative frequency explanation. There are, in fact, other serious problems with the relative frequency hypothesis. Subjects frequently reported nonsense words which began with consonants, in preference to common English words beginning with vowels. As an example, the word "ills" was reported only twice, while its counterpart, "zill" or "ziddle," was reported a total of 15 times. This is inconsistent with any explanation which supposes that subjects draw only on their previous experience of words in forming responses.

A more plausible alternative explanation was that the reader, who was naive to the purpose of the experiment, nevertheless biased the stimuli in reading them. To eliminate this possibility a second experiment was carried out, which required the subjects themselves to read the sequences. Ambiguous words were this time written in cursive script on strips. The strips were then formed into rings so that a word iterated in closed cycle around a ring. Subjects were handed rings one at a time in random order and asked to read the repeating word.

The results again displayed strong preferences for consonant-start words. Of the 108 responses, 81% began with consonants, 19% with vowels. Once again, nonsense words beginning with consonants were frequently preferred to English words beginning with vowels. This never occurred in the opposite direction. The word "ember" was most frequently read as "berem" or "rember," the word "owlet" as "letow" or "towlet." Thus, even though the proportion of responses was not inconsistent this time with the frequency explanation, the nature of the individual results certainly ruled it out.

The results of both experiments supported the experimental hypothesis, that iterating sequences tend to be segmented with the more salient, consonant sounds at the start of perceived patterns. However, one would wish stronger independent evidence that consonants are in fact more salient than vowels, before the results could be regarded as powerful extensions of the sequential processing model. Nevertheless, the results are certainly consistent with the model, and suggest that it may have interesting applications beyond the simple perceptual and learning situations in which it was developed.

The Production of Known Sequences

The discussion so far has been concerned with possible extensions of the model into related areas. There remains, however, a central question concerning the model itself, and it will be appropriate to conclude with a brief consideration of that question.

Learning is a compound process, and a distinction must be drawn between the sub-processes of acquisition, storage and retrieval (Greeno & Simon, 1974). In the course of developing the present theory, little direct reference was made to this distinction. The term "acquisition" has been the most frequently used of the three, and the model referred to the "integration" of patterns, which further implies an acquisition process. Yet nothing in the performances of subjects can be used to so identify the true locus of the experimental results. Performance required acquisition, storage, and retrieval, and it is impossible to isolate which combination of these processes was responsible for the observed differences in performance.

From a theoretical standpoint, the model appealed to both acquisition and retrieval processes. Pattern integration implied an acquisition process, while the proposed scanning mechanism was based on retrieval considerations. Indeed, the assumption was made that retrieval and acquisition take place concurrently, and that subjects make anticipations based on the retrieval of events from storage at the same time as they were in the process of integrating the pattern.

One direction of future research on the model lies in investigating whether the effects of structural embeddedness are limited to acquisition or whether they also affect retrieval alone, following acquisition. It seems that the question could be pursued initially by testing for structural embeddedness effects in the reproduction of already known patterns. If *known* patterns with high structural embeddedness involve more errors or delays than known patterns with low structural embeddedness, then it would indicate that the measure applies beyond the stage of pattern acquisition. A result of this kind could have important theoretical, and also practical, consequences.

An attempt was made in the present experiments to obtain some preliminary information on whether structural embeddedness effects continue after acquisition. When subjects had finally integrated a sequence, and could reproduce it at the coded rate, they were asked to type it as fast as possible from memory. The aim of this was to discover if differences continued to exist between the patterns, after acquisition. Unfortunately the results were inconclusive because performances were so fast in all cases that error of measurement became sufficiently high to obscure possible differences. Subjects were able to produce patterns at a rate of between 150 and 200 msecs, while measurements were made in step units of about 50 msecs. No differences could be observed between sequences, but clearly this could have been due to the relative crudeness of measurement rather than to a true absence of differences.

Previous data are unable to answer the question, either. Royer's (1967) study had required subjects to reproduce sequences from printed cards, in which case it may seem that he was measuring performance rather than acquisition. However, subjects presumably still had to integrate a sequence in the course of reproducing it, and so we must assume that acquisition was taking place, and not simply retrieval of an already known sequence.

It seems that the reproduction of simple *rhythmic* patterns might provide a useful paradigm for testing the relationship of embeddedness to the retrieval of sequences. Such patterns could be readily acquired and repeated, so that on testing the rate or accuracy of reproduction, it would be relatively certain that any differences could not be due to acquisition or storage differences. It is true that problems might arise in computing embeddedness values, but in rhythms involving stressed and unstressed beats it seems natural to refer to the stressed beats as salient. Consequently, sequences having the stressed beats at the end could, in a rough sense, be considered less structurally embedded than sequences with the stressed beats in the middle.

The results of one experiment give some preliminary information on this point. Weaver (1939) used subjects who were trained musicians to reproduce different types of rhythms. The forms were simple, and there can be no question that subjects had to acquire them in any sense. They merely had to produce them in a given time from memory. The simplest involved a basic 4-beat rhythm. By convention, such a rhythm has a major stress on the first beat and a minor stress on the third beat, with the second and last beats unstressed. Assuming that stressed elements are salient, a less embedded rhythm would be obtained by moving the minor stress from the third to the final beat. Was there any evidence that Weaver's musicians simplified patterns in this way? A rough test of this is to examine in how many cases the final beat received a greater stress than the third beat. This happened in three out of nine subjects, but interestingly enough, when the tempo was doubled, *all* subjects made the final beat stronger than the third beat. Assuming that the increase in tempo made the sequence more complex to reproduce, this is what might be expected. In order to compensate for the increase in complexity, subjects reduced the embeddedness of the patterns by relocating the stressed beat.

The result is suggestive only, but it seems to indicate that studies of rhythm might be useful in isolating the true locus of structural embeddedness effects. The question is theoretically important, for the explanation of structural embeddedness proposed earlier involved the use of a scanning model. The scanning model is not limited in its application to the process of acquisition. Indeed, in a limited sense, it *is* a model of a retrieval process. It follows that if the scanning mechanism is an important component in the explanation of structural embeddedness, then such effects should continue to influence retrieval only. Conversely, if no effects of embeddedness can be found when performance is limited to retrieval, then the scanning model is unlikely to be central to an explanation of structural embeddedenss, and another explanation must be sought.

Pinpointing the location of structural embeddedness effects could have practical as well as theoretical consequences, and it may be worthwhile mentioning in conclusion one such possible application.

The use of number codes is currently undergoing a rapid expansion in our society. Almost everyone has a social insurance number, a telephone number, a postal code, a vehicle license number, a chequing account number and a savings account number, a passport number, and so on. Many other codes exist which are more specific to particular occupational roles, such as student registration numbers, military personnel numbers and government department codes. Almost all mechanical devices have serial numbers, and frequently their parts have different numbers, some or all of which may have to be recorded whenever the device is serviced, sold or officially consigned to the breaker.

The still increasing use of computers is likely to lead to further expansion in this use of numbers in place of names. It seems improbable that *everything* humankind has so far named will now be renamed in number, but just how far the process will go remains to be seen.

In the final analysis all names may be arbitrary, but there is an important difference between the arbitrary names conferred by our ancestors and the kind of number-naming which we do today. Spoken and written names have some redundancy, sufficient to tolerate a certain degree of inaccuracy. Inaccuracies, in their turn, are the mutations of language, to be selected or discarded by the evolutionary pressures of common usage. This is presumably the process by which foreign-sounding innovations first become transformed into more familiar combinations of sounds, by which Livorno became Leghorn, and by which Baal Zevuv became Beelzebub. Thus, previously formidable names may, by a process of misspelling and mispronunciation, be recast into a form that no longer mystifies, embarrasses or causes other psychological discomfort to the daily user.

Number-names, by contrast, have no redundancy and brook no mistakes. They are immune to this kind of evolutionary process which transforms other names into convenient and comfortable forms. Even though at any given moment some millions of number-codes may be in process of being dialed, typed, punched, spoken or transcribed by human operators, their still more frequent use by machine ensures their sanctity. Once coined, we tend to be stuck with them. Would it not be wise, then, to select them with more care initially, keeping the human user in mind? The total savings in human effort might be considerable, not to mention the economic consequences of reducing errors and increasing speed in communication, clerical and other data entry systems. It is doubtful if such considerations would be worthwhile if only the acquisition of codes was in question. Most codes require only transcription and relay, and long term retention is not an issue. However, if it were found that these simpler skills, too, were affected by predictable properties of codes, there might be practical consequences to avoiding certain kinds of difficult codes and allocating the simpler codes on the basis of frequency of usage, consequences of error, and other such considerations.

Structural embeddedness is a theoretical measure which predicts the relative complexity of arbitrary codes, and suggests a likelihood of certain kinds of sequencing errors. If such errors continued to be in evidence during simple transcription and retrieval as well as during acquisition, then certain types of code might best be avoided. Once the properties which lead to difficulties were specified, it would be a simple matter to generate by computer lists of available codes in order of "goodness," and to allocate them on this basis. Given the relatively small effort that would be required, compared to the lack of adaptation of codes once created, this humanizing of number-names might not be an entirely quixotic venture.

REFERENCES

Dunn, O. J. Multiple comparisons using rank sums. *Technometrics*, 1964, 6, 241-252.

Farwell, B. A. & Vitz, P. C. A test of the Feigenbaum and Simon model of serial learning. *Journal of Experimental Psychology*, 1971, 91, 240-244.

Feigenbaum, E. A. & Simon, H. A. A theory of the serial position effect. *British Journal of Psychology*, 1962, 53, 307-320.

Galanter, E. H. & Smith, W. A. S. Some experiments on a simple thought problem. *American Journal of Psychology*, 1958, 71, 359-366.

Garner, W. R. & Gottwald, R. L. Some perceptual factors in the learning of sequential patterns of binary events. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 582-589.

Garner, W. R. & Gottwald, R. L. The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology*, 1968, 20, 97-109.

Glanzer, M. Encoding and the perceptual (visual) serial position effect. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 92-97.

Glanzer, M. S. & Clark, H. H. Accuracy of perceptual recall: An analysis of organization. *Journal of Verbal Learning and Verbal Behavior*, 1962, 1, 289-299.

Glanzer, M. & Dolinsky, R. The anchor for the serial position effect. *Journal of Verbal Learning and Verbal Behavior*, 1965, 4, 267-273.

Glanzer, M. & Peters, S. C. Re-examination of the serial position effect. *Journal of Experimental Psychology*, 1962, 64, 258-266.

Greeno, J. G. & Simon, H. A. Processes for sequence production. *Psycholgical Review*, 1974, 81, 187-198.

Hale, D. J. Sequential effects in a two-choice serial reaction task. *Quarterly Journal of Experimental Psychology*, 1967, 19, 133-141.

Handel, S. & Lewis, W. E. The effect of practice on the perception of temporal patterns. *Quarterly Journal of Experimental Psychology*, 1970, 22, 97-108.

Harcum, E. R. *Serial learning and paralearning*. New York: Wiley, 1975.

Haslerud, G. M. & Clark, R. E. On the redintegrative perception of words. *American Journal of Psychology*, 1957, 70, 97-101.

Hsu, P. L. Contributions to the theory of student's t-test as applied to the problem of two samples. *Statistical Research Memoirs*, 1938, 2, 1-24.

Jones, M. R. & Zamostny, K. P. Memory and rule structure in the prediction of serial patterns. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 104, 295-306.

Keele, S. W. Repetition effect: A memory-dependent process. *Journal of Experimental Psychology*, 1969, 80, 243-248.

Lashley, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley, 1951.

Leeuwenberg, E. L. L. Quantitative specification of information in sequential patterns. *Psychological Review*, 1969, 76, 216-220.

Lindman, H. R. *Analysis of variance in complex experimental designs*. San Francisco: W. H. Freeman, 1974.

MacGregor, J. Consonants as anchor points in iterating sequences of ambiguous syllables. Unpublished manuscript, University of Victoria, 1978.

MacKay, D. G. Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, 1970, 3, 210-227.

MacKay, D. M. Quantal aspects of scientific information. *Philosophical Magazine*, 1950, 41, 289-301.

McCrary, J. W. & Hunter, W. S. Serial position curves in verbal learning. *Science*, 1953, 117, 131-134.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81-97.

Miller, G. A. & Friedman, E. A. The reconstruction of mutilated English texts. *Information and Control*, 1957, 1, 38-55.

O'Connor, J. D. *Phonetics*. Middlesex: Penguin Books, 1973.

Payne, B. The relationship between a measure of organization for visual patters and their judged complexity. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 338-343.

Preusser, D. The effect of structure and rate on the recognition and description of auditory temporal patterns. *Perception and Psychophysics*, 1972, 11, 233-240.

Preusser, D., Garner, W. R., & Gottwald, R. L. Perceptual organization of two-element temporal patterns as a function of their component one-element patterns. *American Journal of Psychology*, 1970, 83, 151-170.

Restle, F. Grammatical analysis of the prediction of binary events. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 17-25.

Restle, F. Theory of serial pattern learning: Structural trees. *Psychological Review*, 1970, 77, 481-495.

Restle, F. & Brown, E. Serial pattern learning. *Journal of Experimental Psychology*, 1970, 83, 120-125.

Restle, F. & Burnside, B. L. Tracking of serial patterns. *Journal of Experimental Psychology*, 1972, 95, 299-307.

Royer, F. L. Sequential complexity and motor response rates. *Journal of Experimental Psychology*, 1967, 74, 199-202.

Royer, F. L. & Garner, W. R. Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception and Psychophysics*, 1966, 1, 41-47.

Royer, F. L. & Garner, W. R. Perceptual organization of nine-element auditory temporal patterns. *Perception and Psychophysics*, 1970, 7, 115-120.

Savin, H. B. & Bever, T. G. The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 295-302.

Scott, T. C. & Henninger, L. L. The relation between length and difficulty in motor learning: A comparison with verbal learning. *Journal of Experimental Psychology*, 1933, 16, 657-678.

Simon, H. A. Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 1972, 79, 369-382.

Simon, H. A. & Kotovsky, K. Human acquisition of concepts for sequential patterns. *Psychological Review*, 1963, 70, 534-546.

Snedecor, G. W. *Statistical methods*. Ames: Iowa State College Press, 1950.

Trabasso, T. & Riley, C. A. The construction and use of representations involving linear order. In R. L. Solso (Ed.), *Information processing and cognition*. New Jersey: Lawrence Erlbaum, 1975.

Vitz, P. C. & Todd, T. C. A coded-element model of the perceptual processing of sequential stimuli. *Psychological Review*, 1969, 76, 433-449.

Weaver, H. E. Syncopation: A study of musical rhythms. *Journal of General Psychology*, 1933, 20, 409-429.

Wickelgren, W. A. Acoustic similarity and retroactive interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1965, 4, 53-62.

Williams, J. A. Sequential effects in disjunctive reaction time: Implication for decision models. *Journal of Experimental Psychology*, 1966, 71, 665-672.

Woodworth, R. S. *Experimental psychology*. New York: Holt, Rinehart & Winston, 1938.

VITA

Surname: MacGREGOR Given Names: JAMES NORMAN

Place of Birth: GLASGOW, SCOTLAND

Date of Birth: May 2, 1947

Educational Institutions Attended, with Dates of Entering and Leaving:

GLASGOW UNIVERSITY, GLASGOW, SCOTLAND	1965 to 1970
UNIVERSITY OF VICTORIA, VICTORIA, B.C.	1972 to 1979
	to
	to

Degrees, Diplomas, Etc., Awarded, with Dates and Names of Institutions:

M.A. (Hons.)	1970	Glasgow University, Glasgow, Scot.
M.Sc.	1975	University of Victoria, Victoria

Honours and Awards:

University of Victoria Scholarship, 1972-73

University of Victoria Fellowship, 1973-74

NRC Scholarship, 1975-79

Publications:

PARTIAL COPYRIGHT LICENSE

Title of Thesis/Dissertation

A MODEL FOR THE PROCESSING OF SEQUENTIAL INFORMATION

Author: [redacted]
Signature

James Normal MacGregor
Name

24th July 1979
Date