

Chapter 5: The Start-End Model

A new, positional model of serial recall

The previous chapters presented evidence for positional information in serial recall. This chapter describes a computational model in which this positional information is made explicit as the basis for serial recall. This *Start-End Model* (SEM) provides good quantitative fits to data in Chapters 2, 3 and 4, and makes predictions that are tested in Chapters 6 and 7.

The Core Assumptions of SEM

In brief, SEM assumes that position in a sequence is coded relative to the start and end of that sequence. This positional information is encoded during each presentation (and rehearsal) of an item, creating a episodic token in short-term memory. The order of items is recalled by cuing with positional codes for each position in sequence and selecting the best matching token for that position. Each of these three assumptions is now examined in turn (a more precise formalisation of SEM is given in Appendix 3).

Coding of Position

The initiation and termination of a temporal sequence are the most psychologically salient events in the processing of that sequence. As such, they provide potential reference points, or anchors, with which the sequence can be ordered. With this idea in mind, SEM's coding of position presumes a start marker and an end marker (Houghton, 1990). The start marker is strongest at start of a sequence, and decreases in strength towards end of the sequence. Conversely, the end marker is weakest at start of the sequence, and grows in strength towards the end of the sequence. The relative strengths of the start and end marker therefore provide an approximate two-dimensional code for a position in a sequence.

Such markers may also apply for the coding of spatial position (e.g., Nelson & Chaiklin, 1980). For example, the relative distance from the two ends of a horizontal array might provide an approximate code for an item's position within that array. Within the temporal domain, one might wonder how an item's temporal position can be coded with

respect to an end marker at its time of presentation, if the end of the sequence has not yet occurred. One possibility is that the strength of the end marker corresponds to expectation for the end of the sequence. This possibility, together with other interpretations of the start and end marker, is discussed in Chapter 6. For the moment, the start and end markers can be regarded as a simple means with which to formalise positional information.

The strength of the start and end markers for position $i = 1, 2, \dots, N$ in a sequence of N items, $x(i)$ and $y(i)$ respectively, can be parameterised as:

$$x(i) = S_0 S^{i-1} \quad y(i) = E_0 E^{N-i} \quad \text{Equation 5-1}$$

where $S_0, E_0 > 0$ are the maximum strengths of the start and end markers, and $0 < S, E < 1$ are the rates of exponential change of these strengths. Figure 5-1 shows example strengths of the start and end marker for each position in a sequence of five items.

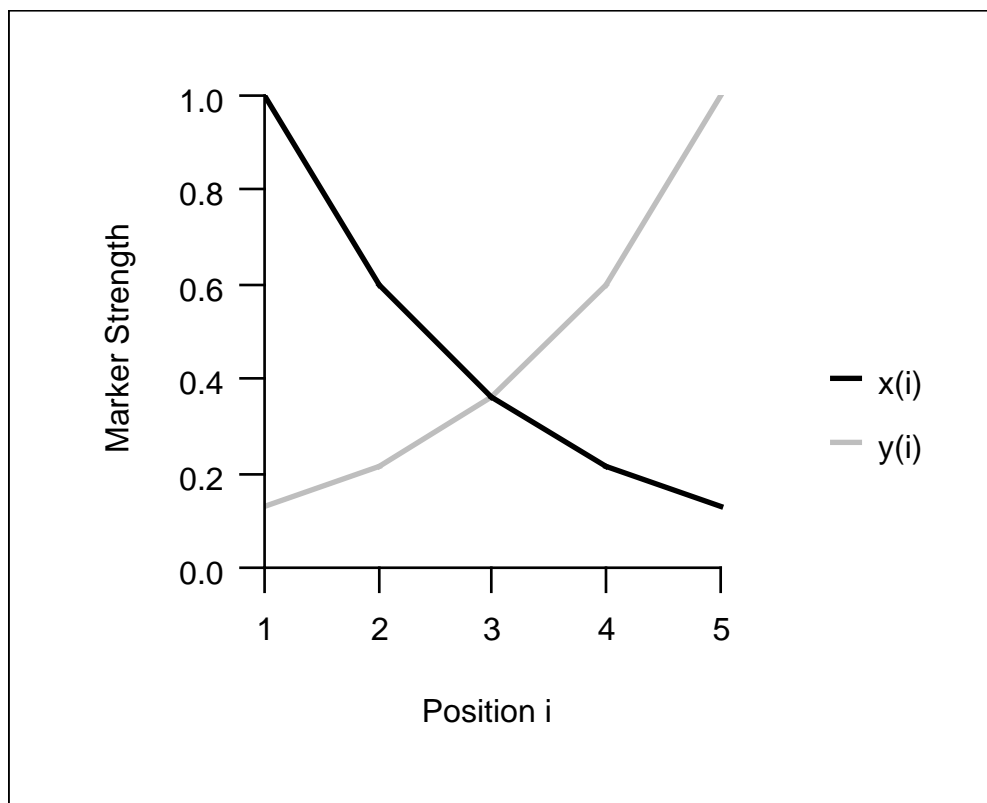


Figure 5-1: Start and end marker strengths, $x(i)$ and $y(i)$, on Positions $i=1..5$ of a five-item list. ($N=5, S_0=E_0=1.00, S=E=0.60$).

The positional code for Position i can be represented by the vector $\mathbf{p}(i)=\langle x(i) y(i)\rangle$. For example, the first position in Figure 5-1 has the code $\langle 1.00 0.13\rangle$, whereas the middle position has the code $\langle 0.36 0.36\rangle$. These codes are assumed to be approximate, in the sense that they share some similarity with one another. This similarity is defined by the overlap, $o(\mathbf{p},\mathbf{q})$, between positional codes $\mathbf{p}(i)$ and $\mathbf{q}(j)$ for positions i and j :

$$o(\mathbf{p},\mathbf{q}) = \sqrt{\mathbf{p} \cdot \mathbf{q}} \times \exp\left(-\sqrt{\sum_k (p_k - q_k)^2}\right) \quad \text{Equation 5-2}$$

where k indexes the (two) components of each vector. The upper panel of Figure 5-2 shows the overlap between positional codes for all Positions $i,j=1..5$, using the same start and end marker parameters as in Figure 5-1. Each curve shows the *positional uncertainty function* for a position, resembling those found in position-probed item recall (Fuchs, 1969) and item-probed position recall (McNicol, 1975). They also resemble the intrusion gradients in Figure 3-6.

The second term in Equation 5-2 is a Euclidean metric of the similarity between two vectors (McNicol & Heathcote, 1986; Nosofsky, 1986), sharpened by an exponential function (Houghton, 1994). This measure of similarity is maximal when $i=j$, and decreases as $|i-j|$ increases. This produces the basic triangular-shape of the positional uncertainty functions.

The prior term in Equation 5-2 is the square-rooted, inner product of the two vectors, representing the combined strength of the start and end markers at the two positions. The effect of this premultiplier is to modify the height and sharpness of the positional uncertainty functions. For example, it lowers and widens the functions for middle positions relative to terminal positions. In general, the height of the functions is increased by increasing the maximum values of the marker strengths (increasing S_0, E_0), while the sharpness of the positional uncertainty functions is increased by increasing the rate of change of marker strengths (decreasing S, E).

The positional uncertainty functions in the upper panel of Figure 5-2 are symmetrical. In subsequent fits, the end marker is generally weaker ($E_0 < S_0$) and changes faster ($E < S$) than the start marker. The effect of these changes is to make the positional uncertainty functions asymmetrical, being skewed towards earlier positions. This asymmetry sometimes appears in position-probed item recall, particularly when allowing for response bias (Murdock, 1968).

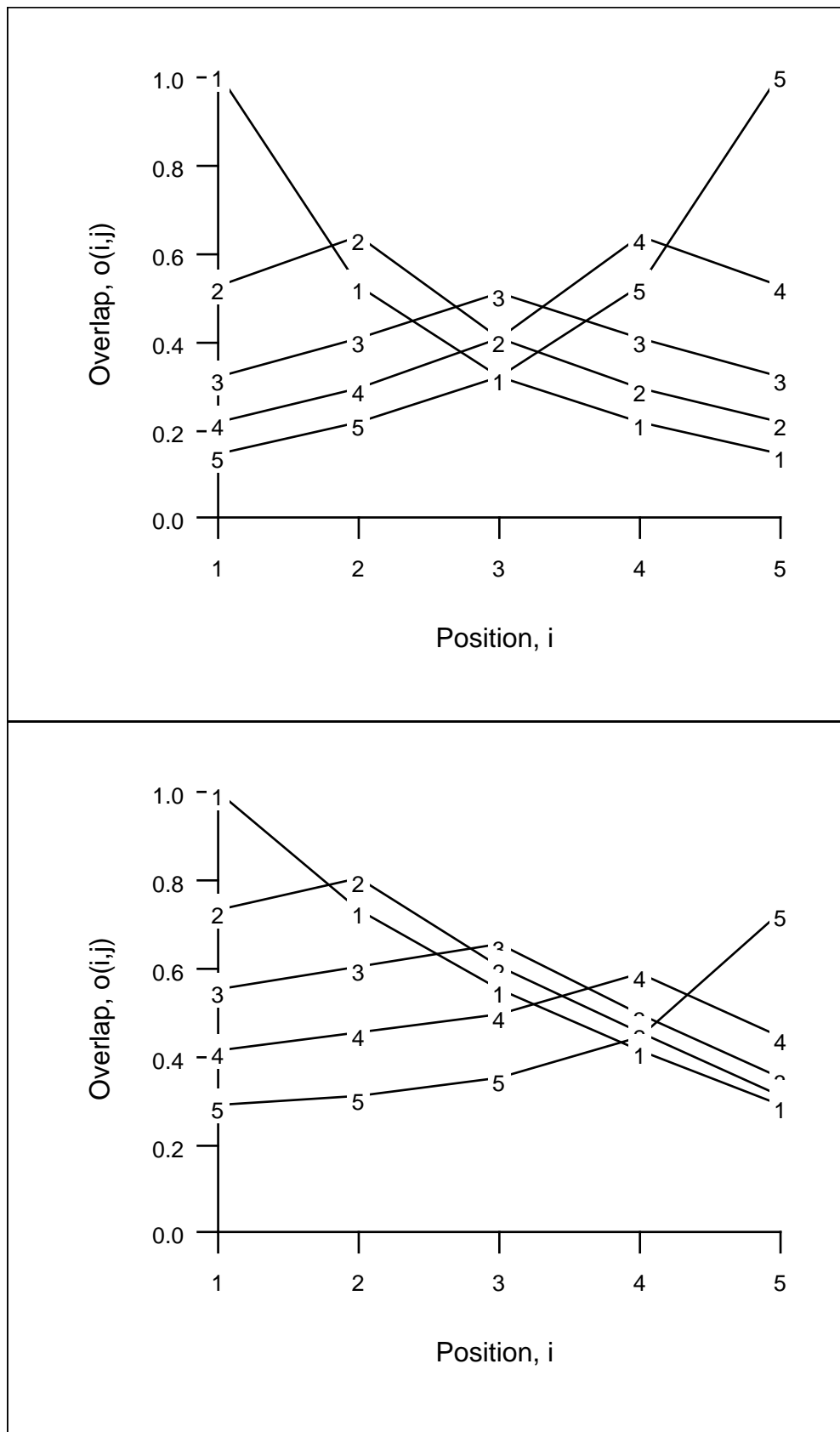


Figure 5-2: Positional uncertainty functions for each Position $j=1..5$ of a five-item list. (Upper panel, $S_0=E_0=1.00$, $S=E=0.60$; lower panel, $S_0=1.00$, $E_0=0.60$, $S=0.80$, $E=0.48$.)

The lower panel of Figure 5-2 shows positional uncertainty functions for such a case, using the same parameter values as in Fits 1 to 4 (below). One important consequence of this asymmetry is that it allows SEM to produce the correct level of fill-in (Chapter 4): The stronger, longer-lasting influence of the start marker biases errors towards earlier items (Fit 1).

In summary, the start and end markers defined in Equation 5-1, together with the positional overlap defined in Equation 5-2, produce positional uncertainty functions resembling those in the data. They allow the N^2 values of positional uncertainty functions for a list of N items to be condensed into 4 parameters. However, before making contact with data, further assumptions are required about the storage and retrieval of items in a sequence.

Storage of Positional Tokens

Each presentation and rehearsal of an item is assumed to create a new token in short-term memory. These tokens are episodic records that a particular item occurred in a particular spatiotemporal context. In other words, positional information is encoded together with items, such that memory for an item is “coloured” by the context in which it was perceived.¹ The representation of an item at the start of a sequence is therefore quite different from the representation of the same item at the end of a sequence. Thus, short-term memory is not viewed simply as a subset of active LTM type representations (Cowan, 1993), but an unordered set of new, episodic tokens. This assumption is important in modelling sequences with repeated items (Chapter 7).

In SEM, tokens contain several components. Some components represent item information, while others represent the positional codes described above. For example, after encoding of the three-item list *RMQ*, short-term memory would contain three tokens like those depicted below (using the same start and end marker parameters as in Figure 5-1):

<	{R}	1.00	0.36	>
<	{M}	0.60	0.60	>
<	{Q}	0.36	1.00	>

1. Unlike the context-sensitive tokens of Wickelgren (1969) however, this context is an abstract positional code, rather than the surrounding items, and unlike the time tags of Yntema and Trask (1963), this code is only defined relative to the start and end of a sequence; it does not refer to absolute time.

The first component $\{X\}$ codes the identity of Item X , the second component is the strength of the start marker during the encoding of Item X , and the third component is the strength of the end marker during the encoding of Item X .

It is assumed that $\{X\}$ represents a central code for Item X . This reflects the fact that the tokens, though positional, are not necessarily superficial representations of items. Tokens are assumed to be created after several stages of stimulus processing. Similar tokens are therefore created for both auditory and visual material (though this is not to deny other differences between the two modalities, as discussed later.) The central representations are assumed to be unitised (lexical) rather than phonological, concordant with people's ability to recall lists of phonologically identical items (Crowder, 1978), and with latencies in item recognition tasks (Clifton & Tash, 1973). There is no phonological similarity between tokens; the effects of phonological similarity arise in a second stage of response selection (Fit 4).

The assumption that tokens are created during recoding of the stimulus means that start and end marker strengths do not need to change in real time. In fact, these strengths are assumed to change only with position. SEM models real-time effects, such as presentation and rehearsal rate, with an additional contextual component to tokens and an assumption of phonological decay (Fit 6).

Retrieval of Items in Order

Tokens in short-term memory are stored unordered; their ordering occurs during recall. To recall a sequence, SEM cues each response by reinstating the positional code corresponding to the position being recalled. These positional codes are based on the same start and end markers assumed above. For example, the cue for the second response in the previous example can be depicted as:

$$< \quad \{?\} \quad 0.60 \quad 0.60 \quad >$$

This cue is matched against all tokens in parallel, with the overlap between the positional code in the cue and the positional code in the tokens defined in Equation 5-2. These overlaps determine the strengths with which each item competes for output in competition space (Chapter 4). More specifically, competition is held over LTM type representations,

activated in proportion to the maximum overlap between the cue and tokens of each type (Appendix 3). Access to these LTM type representations is assumed necessary in order to give a categorical response.

With short sequences such as in Figure 5-2, and in the absence of any other factors, the strongest item is always the correct item (because the peaks of the positional uncertainty functions correspond the correct position). Thus, in a perfect system, recall of such sequences will always be correct.² To model the vagaries of human short-term memory, noise is added to SEM. One form of noise is a random value added to the strength with which each item competes for output, introducing potential errors in recall. Hence positional uncertainty functions indicate only the probability of correct recall.

The final assumption of SEM's recall process is that once an item has been recalled, its LTM type representation is temporarily suppressed. This reduces the probability of recalling that item again (at least within the same trial), which is necessary to explain why repetitions are rare (Chapter 4). Any model that has fill-in will produce far too many repetitions without suppression. Indeed, all other successful models of serial recall presume this process (e.g., Burgess & Hitch, 1992; Houghton, 1990; Page & Norris, 1996b). Response suppression has independent justification from the fact that people often fail to recall the second occurrence of a repeated item (Chapter 7). Indeed, suppression of previous actions is assumed to be a general consequence of sequential behaviour (Houghton & Tipper, 1996).

In summary, for the simple case of a list of N unique items, the strength with which Item i competes for Response j , $c^{(i)}(j)$, is given by:

$$c^{(i)}(j) = o(i, j) (1 - s^{(i)}(j)) + n \quad \text{Equation 5-3}$$

where $o(i, j)$ is the overlap between positional codes for Positions i and j (equivalent to $o(p, q)$ in Equation 5-2), $0 < s^{(i)}(j) < 1$ is the suppression of Item i come Response j , and n is a random variable drawn from a Gaussian distribution for each item and each response. The Gaussian

2. In principle, perfect serial recall can be obtained from Equation 5-2 without the need for an end marker (i.e., with $E_0=0$). Indeed, SEM can even produce a primacy-gradient of cued-strengths in such cases, as in the Primacy Model (Page & Norris, 1996b), providing the start marker changes rapidly (i.e., S is small). The end marker is necessary however to obtain the complete pattern of errors in the data (below; Chapter 6).

distribution has a mean of zero and a standard deviation given by the parameter G_C . In the simplest form of SEM, $s^{(i)}(j)=0$ if Item i has not been recalled up to Position j , and $s^{(i)}(j)=1$ if it has. (In later versions of SEM, suppression is refractory, wearing off during recall; Fit 3).

Fitting SEM to Data

Equations 5-1 to 5-3 represent the most basic form of the model. Given the probabilistic nature of Equation 5-3, together with the fact recall of one item depends on recall of previous items (via their suppression), analytical solutions of SEM's behaviour are hard to obtain. Consequently, these equations are implemented in a computer program that simulates recall of lists. Indeed, the program can be run on the same lists given to subjects, producing reports which can be compared directly. In subsequent sections, SEM was fitted to a range of data, during which the basic model was refined and extended, capturing an increasing number of the important characteristics of short-term serial recall. In particular, Fits 1 to 5 came from a single-trial version of SEM, which does not model intertrial effects, while Fit 6 (together with Fits 7-8 in Chapter 7, and Fits 9-12 in Appendix 3) used a multiple-trial version. The present chapter describes these versions verbally, while more formal specifications are given in Appendix 3, together with the full range of parameter values used in each Fit.

Given the random nature of the noise in SEM, simulation of many trials is necessary to ensure accurate numerical solutions. In fact, for all subsequent fits, the model was run for 100,000 trials (i.e., the model "recalled" 100,000 lists). With this many trials, the variance in SEM's outputs is very small. For example, running SEM 12 times with different random seeds produced variances less than 0.03% for each point in the serial position curve in Fit 1. As a consequence, variances for SEM's results are not given in subsequent fits (they are assumed negligible) and the simulation results are treated as exact predictions.

Quality of Fit

The quality of SEM's fits to data is judged in several ways. One index is the Root-Mean Square Error (RMSE) between SEM's predictions and the means of a set of data points.³ The smaller the RMSE, the better the fit. However, the RMSE does not take into account the variance or covariance amongst in data points. A large RMSE may not be a

3. In all present fits, the data is untransformed.

problem if there is considerable random error in the data. Given the variance and covariance amongst data points, a second index is to test whether the model predictions differ significantly from the data. The test used here is Hotelling's T^2 -test, which assumes neither homogeneity of variance nor independence between data points. A good fit should have a low value of T^2 and a F -ratio that does not test as significant, given the number of data points and the sample size (Appendix 1).

Finally, it remains possible for good quantitative fits as judged by RMSE or Hotelling's T^2 -test, without a model exhibiting an important aspect of the data. For example, a model could produce a reasonable fit to nine data points of a serial position curve, without actually showing any recency (the model might show a monotonic function for example). In such a case, a small RMSE for the first eight points might mask the larger error for the last position, and there may not be enough power for the model to test significantly different from the data (though the problem might be apparent in a nonrandom pattern of residuals). Thus a final important criterion for a good fit is that individual effects identified as significant in the data (such as recency) should also be shown by the model. In subsequent fits, all three criteria are employed, to ensure that the model meets the empirical constraints identified in Chapter 4.

Optimising Fits

Finding optimal values for the free parameters of a model, in order to give the best fit to the data, is a difficult problem. Though automatic procedures exist to optimise a model with respect to an error measurement (e.g., gradient descent methods to minimise RMSE), most of these procedures become very slow as the number of parameters grows beyond three or four. With no intuitive understanding of the parameter space (e.g., what effect increasing a particular parameter will tend to have), automatic procedures can sometimes be more of a hindrance than a help. Thus in all subsequent fits, the model was fitted by hand. Though time-consuming, such an approach has the advantage that it engenders a good understanding of how a model behaves. Also, parameter values were only fitted to the first or second significant figure. Though it remains possible that smaller RMSE's could result with even finer tuning, fits that passed Hotelling's T^2 -test, and which gave the same patterns that were significant in the data, were deemed sufficient.

It is important to minimise the number of parameters that are free to be optimised to the data. In the limit, the number of free parameters should not, of course, exceed the degrees of freedom in the data. The fewer the number of free parameters required for a satisfactory fit, the more powerful the model. As described so far, the basic model has five parameters (S_0 , E_0 , S , E and G_C). To reduce the number that were free to vary, the start marker parameters were fixed at $S_0=1.00$ and $S=0.80$ for all subsequent fits. The end marker parameters were redefined in relation to these values, replacing the four parameters with two free parameters, F_0 and F :

$$F_0 = \frac{E_0}{S_0} \quad F = \frac{E}{S} \quad \text{Equation 5-4}$$

In other words, F_0 represents the maximum strength of the end marker relative to that of the start marker, and F represents the rate of change of the end marker relative to that of the start marker. In subsequent fits, the end marker was generally weaker than the start marker, and changed faster (i.e., $F_0 < 1$ and $F < 1$).

Fit 1. Primacy, Recency, Locality and Fill-in

The most basic form of SEM was fit to the error position curve from the Long condition of Experiment 3. This fit had 3 free parameters: F_0 , F , parameterising the end marker relative to the start marker, and G_C , the amount of noise in competition for output. With $F_0=F=0.60$ (the same values used in the lower panel of Figure 5-2) and $G_C=0.14$, the RMSE to the 5 data points was 4.05%. SEM produced the correct pattern of prolonged primacy and last-item recency (upper panel of Figure 5-3). Indeed, Hotelling's T^2 -test showed that the model did not differ significantly from the data, $T^2=0.85$, $F(5,13)=0.13$, $p=.98$.

The transposition gradients produced by SEM are shown in the bottom panel of Figure 5-3. SEM clearly met the locality constraint, with transpositions decreasing with increasing transposition distance (the RMSE to the 25 data points was 4.95%). The sharpness of these gradients derives from the elongated tail of the Gaussian distribution of noise, without needing to attribute separate sources to correct responses and errors (Drewnowski, 1980a).

More detailed analysis of transpositions showed that SEM also produced the correct level of fill-in (Table 5-1). Both model and data showed that, if Item $i+1$ was recalled too early

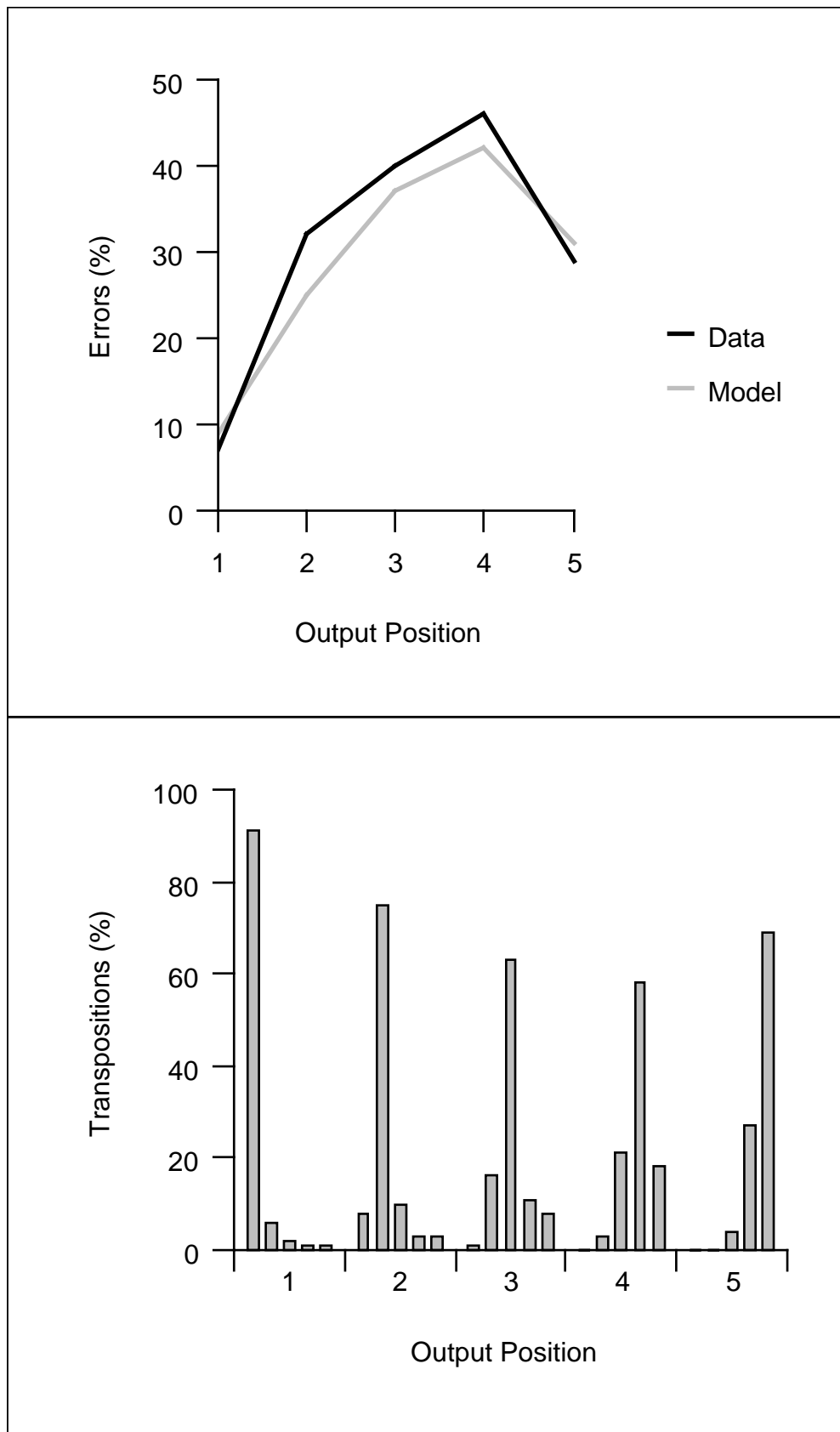


Figure 5-3: Errors by position from data and from SEM (upper panel) and transposition gradients from SEM (lower panel) in Fit 1.

on Position i , the most likely next response was Item i (weak fill-in). However, if neither Item i nor Item $i+1$ was recalled on Position i , the most likely next response remained the correct Item $i+1$, and not Item i (no strong fill-in). The latter is the defining characteristic of positional models like SEM (Chapter 4), and contrary to ordinal models like the Primacy Model (Page & Norris, 1996b). The main discrepancy between SEM and the data was a greater percentage of *Other* errors in the data, which probably reflected less systematic errors such as guesses.

	Fill-in (Item i)	Correct (Item $i+1$)	Associate (Item $i+2$)	Other
First Error of Item $i+1$				
Data	.49	.00	.25	.25
Model	.65	.00	.32	.03
First Error of Item $j>i+1$				
Data	.34	.51	.05	.10
Model	.35	.63	.01	.01

Table 5-1: Proportion of responses following a first error on Position i from SEM in Fit 1.

SEM produced weak fill-in because the start marker was stronger and slower changing than the end marker. This makes positional uncertainty functions asymmetrical, biased towards earlier items (at least for the first few positions; lower panel of Figure 5-2). This can be illustrated in competition space in Figure 5-4: Item 1 remains the most likely response following erroneous recall of Items 2 and 3 (cf. symmetrical positional cuing in Figure 4-1).

In summary, SEM demonstrated a good quantitative fit to the data, and met the primacy, recency, locality and fill-in constraints. These constraints follow naturally from SEM's positional coding and its recall process. Nonetheless, fitting 5 data points with 3 free parameters is not particularly impressive. Subsequent fits extend SEM's coverage, fitting many more data points, while only adding a few new parameters.

Fit 2. Omissions

The main problem with the basic form of SEM used in Fit 1 is that it produces only order errors. Yet item errors comprised 30% of errors in the data in Figure 5-3 (a cautionary illustration of how a good quantitative fit to serial position curves can be achieved without

SEM (Positional)



Figure 5-4: Competition space within SEM for the first three responses to a list *12345* recalled as *23...*, illustrating weak fill-in.

necessarily respecting the nature of the underlying errors). However, the addition of one new parameter allows SEM to fit both item and order errors. This parameter is an omission threshold, T_O . The strongest competitor is suppressed as before, but if its strength does not exceed T_O , then the corresponding item is not output and an omission is indicated instead.

With this simple addition, the model was fitted to transpositions and omissions from the Long condition of Experiment 3 (intrusions in the data were pooled with omissions for the purpose of this fit). In fact, values for the 3 basic parameters, $F_O=F=0.60$ and $G_C=0.14$, remained the same as in Fit 1. Setting the one free parameter $T_O=0.48$ gave an RMSE of 3.95% to 10 data points for each error type at each output position (lower panel of Figure 5-5). SEM produced the correct pattern of recency in transpositions, but not omissions, consistent with the omission constraint (Chapters 3 and 4; cf. lower panel of Figure 4-3).

The monotonic increase in omissions with output position did not mean that the last item was omitted more than any other. This is shown in the upper panel of Figure 5-5 (cf. upper panel of Figure 4-3). When scored against input position, omissions did show a recency effect, meaning that the last item was more often recalled somewhere than the penultimate item. Indeed, Hotelling's T^2 -test showed the model did not differ significantly from this data, $T^2=7.07$, $F(10,8)=0.33$, $p=.95$.

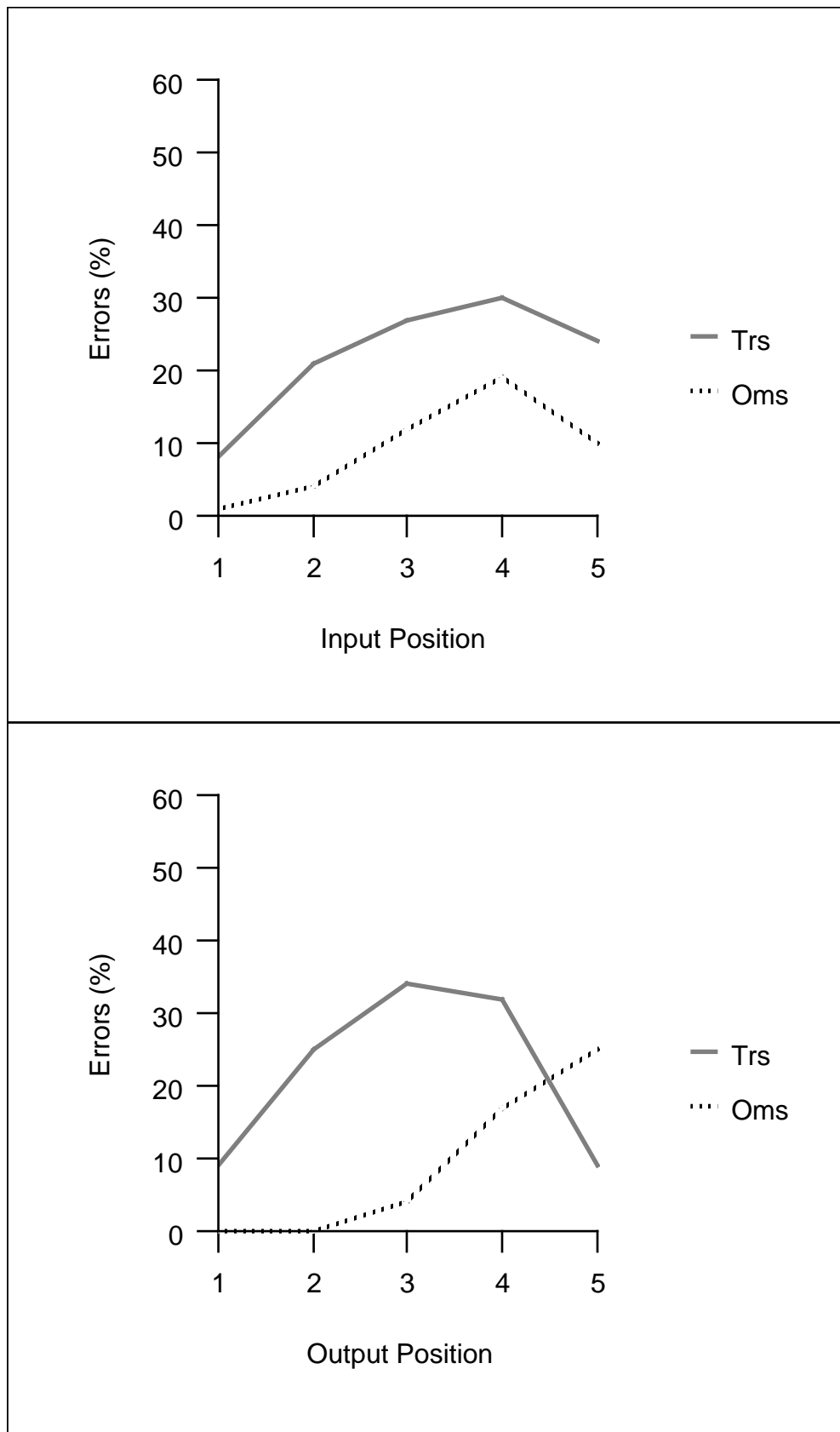


Figure 5-5: Omissions and transpositions by input position (upper panel) and output position (lower panel) from SEM in Fit 2.
(Trs=Transpositions; Oms=Omissions.)

How does SEM produce this pattern of omissions? The short answer is that, when the last item is recalled too early, it is likely to be followed by omissions. This can be illustrated in competition space in Figure 5-6, where the horizontal line indicates the omission threshold. Come the fourth response, random noise in the strengths of Item 4 and Item 5 can cause the last item to be recalled too early. The reason this is usually followed by an omission is that the positional uncertainty function for the last position is relatively sharp (lower panel of Figure 5-2). In other words, only the last item is cued strongly at the last position, and if that item has already been recalled and suppressed, it is less likely that other items, such as Item 4, will be cued above the omission threshold. The fact that Item 5 is more likely to be recalled in Position 4 than Item 4 is to be recalled in Position 5 leads to recency when omissions are scored against input position, but not when scored against output position.

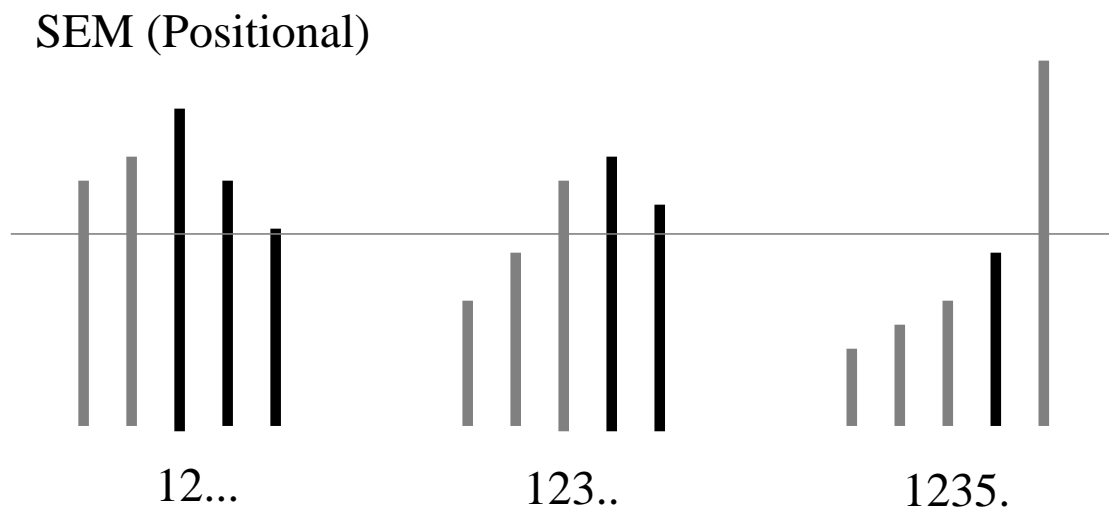


Figure 5-6: Competition space within SEM for last three responses to a list *12345* recalled as *1235-*, illustrating possibility of omissions.

No other model appears able to explain this pattern of omissions. The Perturbation Model (Lee & Estes, 1977, 1981) assumes that omissions are only ever flat or monotonic across input position, which may be true of short lists (e.g., Healy, 1974), but is not true of longer lists (Experiment 2). The Primacy Model (Page & Norris, 1996b) produces omissions that increase towards the end of recall, but only through more omissions of the last item than

any other, which is not always the case (Chapter 4). The Articulatory Loop Model (Burgess & Hitch, 1992) fails for the same reason. In SEM, the complete pattern of item errors falls out of the dynamics of the recall process, together with the simple assumption of weak yet sharply-tuned end marker. This behaviour was an unexpected emergent property of the model.

Fit 3. Repetitions

As it stands, SEM does not produce enough repetitions. The suppression process means an item is extremely unlikely to be recalled more than once within the same trial. To capture repetitions of the sort described in Chapter 4, suppression is assumed to wear off during recall, by letting:

$$s^{(i)}(j+1) = s^{(i)}(j) \exp(-R_S) \quad \text{Equation 5-5}$$

where $R_S > 0$ is a new parameter reflecting the rate with which suppression decays. An example suppression profile for an item recalled on Position 1 is shown in Figure 5-7. Suppression is maximal during the immediately following response ($s^{(i)}=1$), but decreases during subsequent responses, eventually returning to the baseline level ($s^{(i)}=0$) between trials.

This version of SEM was fitted to the PN condition of Experiment 1 (which had more repetitions than the Long condition of Experiment 3). Again, the end marker parameters were unchanged from previous fits. The noise and threshold were changed, to allow for differences in the materials and procedure of Experiment 1. The 3 free parameters were $G_C=0.08$, $T_O=0.32$, $R_S=0.50$, giving an RMSE of 5.79% to the 6 data points of the error position curve; a fit that did not differ significantly from the data, $T^2=2.50$, $F(6,42)=0.37$, $p=.89$.⁴

Figure 5-8 shows the frequency of repetitions at each input and output position (cf. Figure 4-4). In both model and data, most repetitions occurred towards the end of recall and were repetitions of the first few items in the list. Repetitions were generally far apart in a report, being 3.65 positions apart on average in the model, and 3.34 positions apart in the data. These figures reflect the time it took for suppression to wear off significantly.

The model did not produce as many repetitions as found in the data. In the model,

4. Note that allowing decay of suppression in previous fits did not compromise the goodness of those fits.

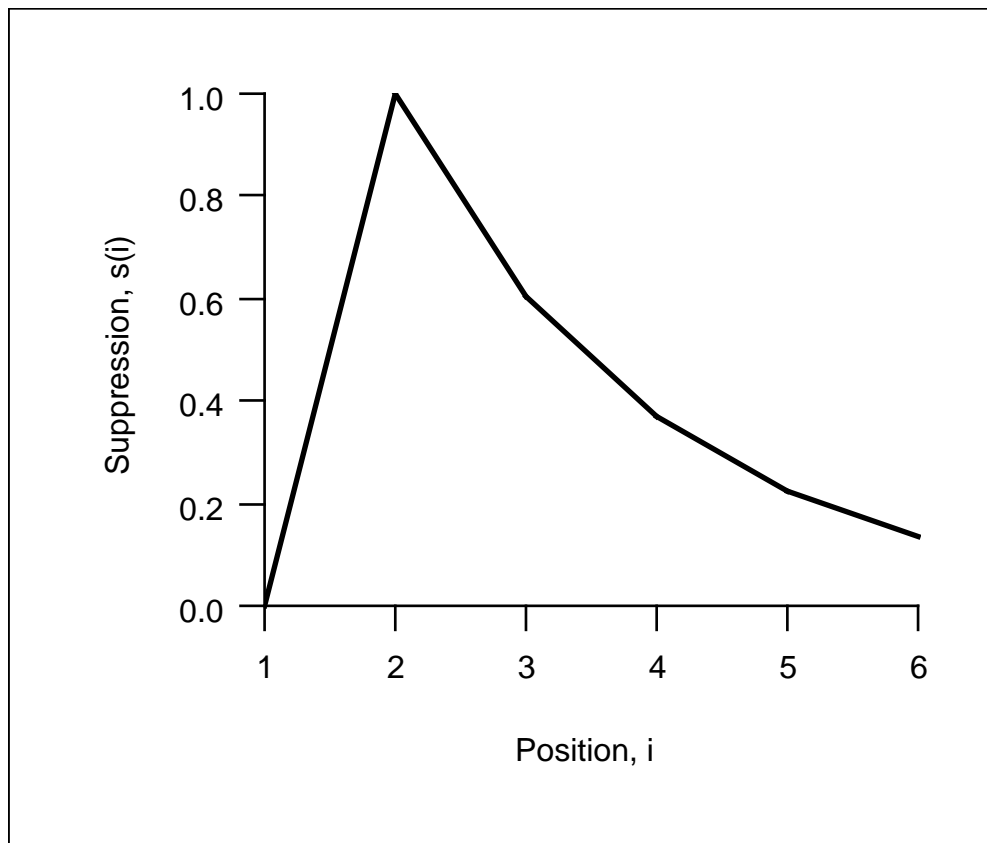


Figure 5-7: Suppression profile for an item recalled at Position 1.

repetitions comprised approximately 4% of errors, compared with a figure of 11% in the data. The reason for this discrepancy is that subjects in Experiment 1 were instructed to group the lists in threes, for which the current version of SEM made no allowance (though see Fit 5). Indeed, almost half of the repetitions in the data were three positions apart, corresponding to interpositions between groups (Experiment 2). Thus, the low RMSE of 1.67% over the 10 data points in Figure 5-8 reflects the small frequencies involved, and belies considerable differences between the model and data owing to grouping effects. The important point of the fit however was that SEM produced the correct qualitative distribution of repetitions required by the repetition constraint (Chapter 4).

One might wonder why the first item was often recalled in both the first and the last position of a report, given that it was not strongly cued at the last position. The reason is similar to the reason why omissions increase towards the end of recall: Repetitions often follow cases where the last item has been recalled too early. Again, this can be illustrated in

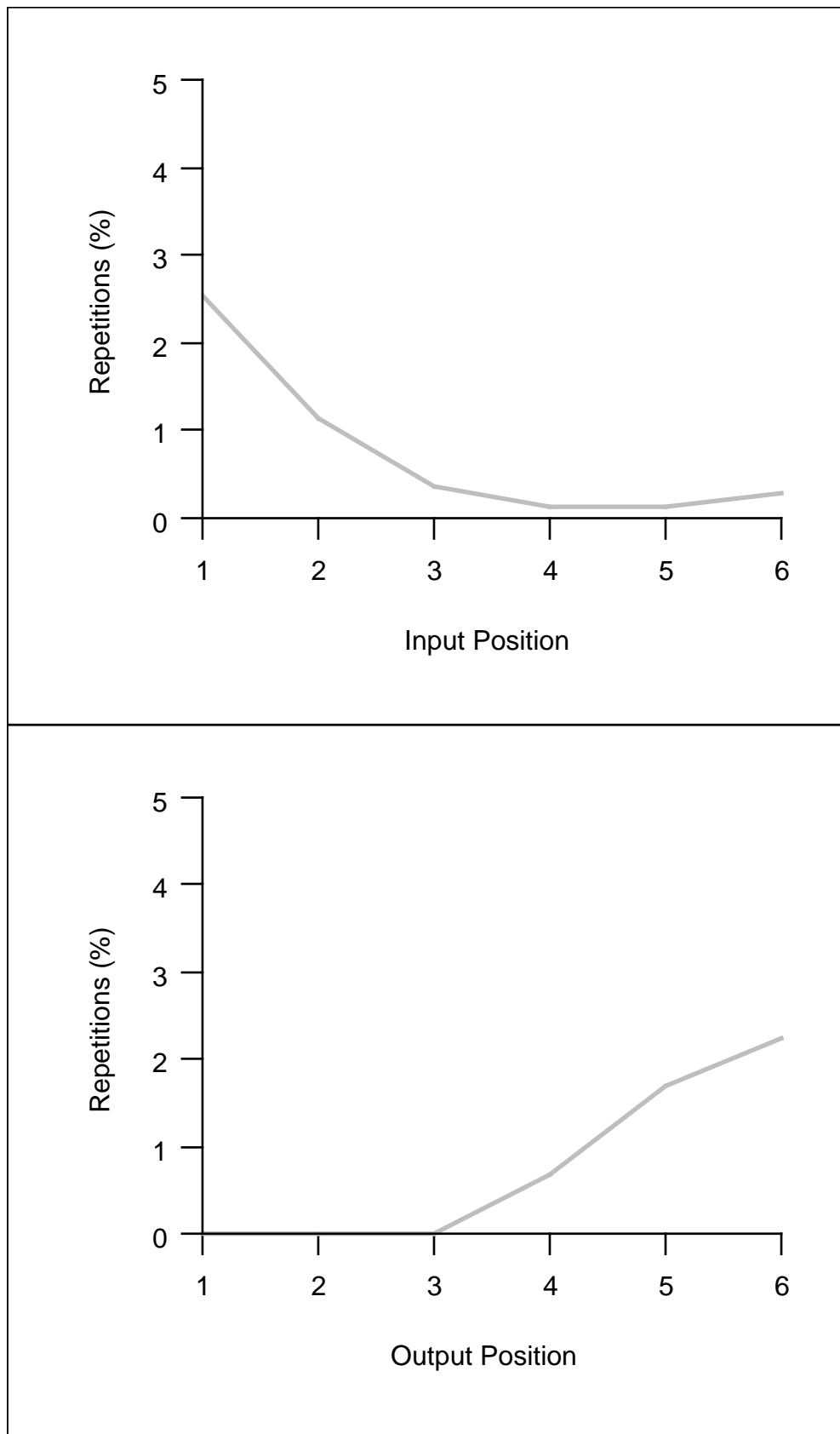


Figure 5-8: Repetitions by input position (upper panel) and output position (lower panel) from SEM in Fit 3.

competition space (Figure 5-9). Because the positional uncertainty function for the last position is so sharp, there is little difference in the strength with which the first few items are cued for the last position. Given that the first item has normally had slightly longer for suppression to wear off, then, if any item is to be repeated (i.e., Item 4 is not recalled), due to additional noise pushing it above the omission threshold, it is most likely to be the first item. Note that this pattern of errors would not be likely with the symmetrical coding of position in the Articulatory Loop Model (Burgess & Hitch, 1992, 1996b), where there would be virtually no overlap between positional codes for the first and last positions of reasonably long lists.

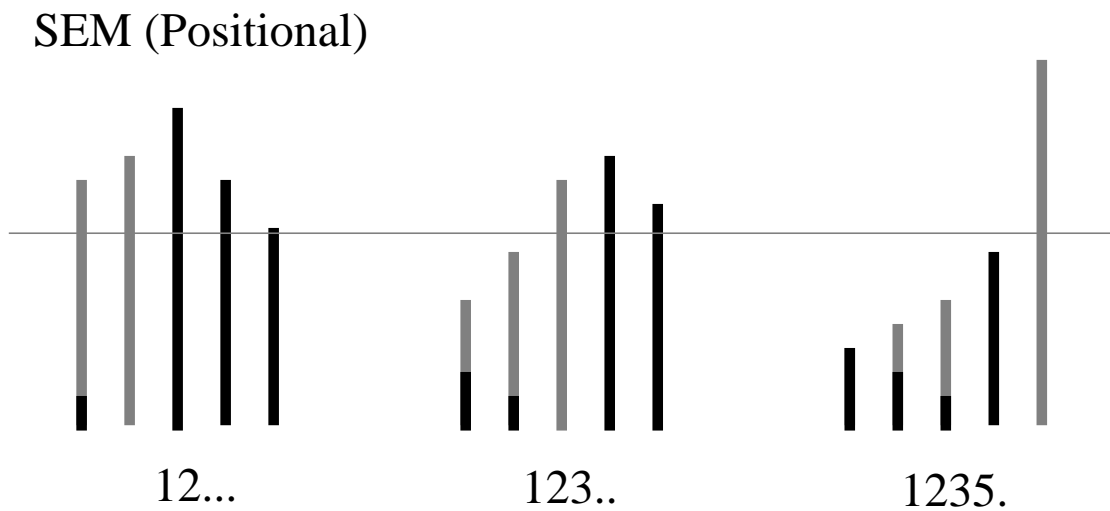


Figure 5-9: Competition space within SEM for the last three responses to a list *12345* recalled as *12351*, illustrating possibility of repetitions.

The exact frequency of repetitions depends on factors such as list length, the omission threshold and, in particular, guessing strategies (Chapters 6 and 7). However, the rate with which suppression decays remains the most important factor, particularly given that many interpositions are repetitions (Experiment 2): Because interpositions are accompanied by an decrease in overall errors, such repetitions are difficult to attribute to guessing strategies.

Finally, another general point about modelling emerges from this fit. Repetitions represent little more than 2% of responses in Experiment 1. Thus it is possible for a model to account for nearly all the variance in serial position curves, without producing any repetitions. Yet repetitions are not random errors; they are highly constrained in their distribution. In other

words, an excellent quantitative fit to serial position curves would not reflect a small, but reliable aspect of the data. This emphasises the importance of applying hypothesis testing to models as well as data. The addition of a fifth free parameter to SEM, R_S , is not justified in order to produce a smaller RMSE, but in order to explain an important subclass of errors.

Fit 4. Phonological Confusions

In addition to demonstrating the appropriate pattern of transpositions, omissions and repetitions, SEM must be able to produce phonological confusions. Moreover, such errors must be sensitive to the strong constraints shown in Chapter 1, which are troublesome for other models, and chaining models in particular.

To fit the alternating curves of Experiment 1, SEM borrows an assumption from the Primacy Model (Page & Norris, 1996b). This assumption is that phonological confusions happen at a second stage of response retrieval. An item is selected as before, but before it is output, its phonological representation is accessed, in order to articulate a response. Occasionally though, competition over such phonological representations may result in access to a similar, but incorrect item, resulting in a confusion error. This extra stage of phonological retrieval was simply added to the existing version of SEM (for more details, see Appendix 3).

The addition of phonological retrieval involves four new parameters. Parameters $0 < P_S, P_D < 1$ reflect the similarity between phonologically similar (confusable) and dissimilar (nonconfusable) items respectively. The item chosen after positional cuing activates its own phonological representation by an amount 1, similar ones by an amount P_S , and dissimilar ones by an amount P_D . The parameter $A_P > 0$ reflects the baseline activation of the phonological representations, and the parameter G_P reflects additional noise in these activations (similar to G_C).

The baseline activations of phonological representations are assumed to arise from phonological access during presentation or rehearsal of a list (Fit 6). They therefore provide an additional item memory. (In subsequent versions of SEM, these activations decay over time, resembling a short-lived phonological store.) Thus, though the activation of phonologically similar items can impair recall, the baseline activation of list items produces an overall beneficial effect, by keeping items above the omission threshold and reducing the incidence of

extralist intrusions (Fit 9 in Appendix 3).

In the following fit, only two of the new parameters, P_S and G_P , were free to fit the data; the values $P_D=0.00$ and $A_P=1.00$ were fixed. The remaining parameters were unchanged from Fit 3, except for T_O , which was increased to allow for the additional phonological activations. The optimal values of 3 free parameters were $P_S=0.75$, $G_P=0.30$ and $T_O=0.90$, producing the error position curves for each condition in Experiment 1, with A2 lists removed (upper panel of Figure 5-10; cf. lower panel of Figure 2-1). The RMSE over all 24 data points was 5.06%, and the fit did not differ significantly from the data, $T^2=8.36$, $F(24,24)=0.18$, $p=.99$. Most importantly, error frequencies on nonconfusable positions in alternating curves did not differ from those on nonconfusable positions in the nonconfusable curve (Chapter 2).

SEM also showed the correct interaction between phonological similarity and transposition distance. The transposition gradients for conditions PC and PN (lower panel of Figure 5-10; cf. Figure 2-3) revealed an underadditive effect of phonological similarity. This interaction arose because the competition amongst phonological activations is weighted by the positional grading of categorical activations (Appendix 3).

While the implementation of phonological similarity in SEM might appear complex, there are three fundamental reasons why SEM fits the alternating curves of Experiment 1, where other models (except the Primacy Model of Page & Norris, 1996b) have failed (Henson et al., 1996). The first is that items are stored as separate nonphonological tokens. This means that phonologically similar items do not interfere with each other during storage. Such interference does occur in distributed phonological stores, such as associative networks (Jordan, 1986; Lewandowsky & Li, 1994), or the original Articulatory Loop Model (Burgess & Hitch, 1992). The second reason is that order is not stored via associations between phonological representations of items. Thus there is no effect of similarity on cuing, in contrast to most chaining models (Chapter 2). The final reason is that suppression of categorical representations is independent of suppression of phonological representations (Page & Norris, 1996b). This can prevent an effect of errors on cuing (Chapter 2)⁵. These assumptions seem vital in order to model what has proved to be extremely constraining data.

5. though not always true of the data (Chapter 4), consistent with the multiple-trial version of SEM (Fit 9).

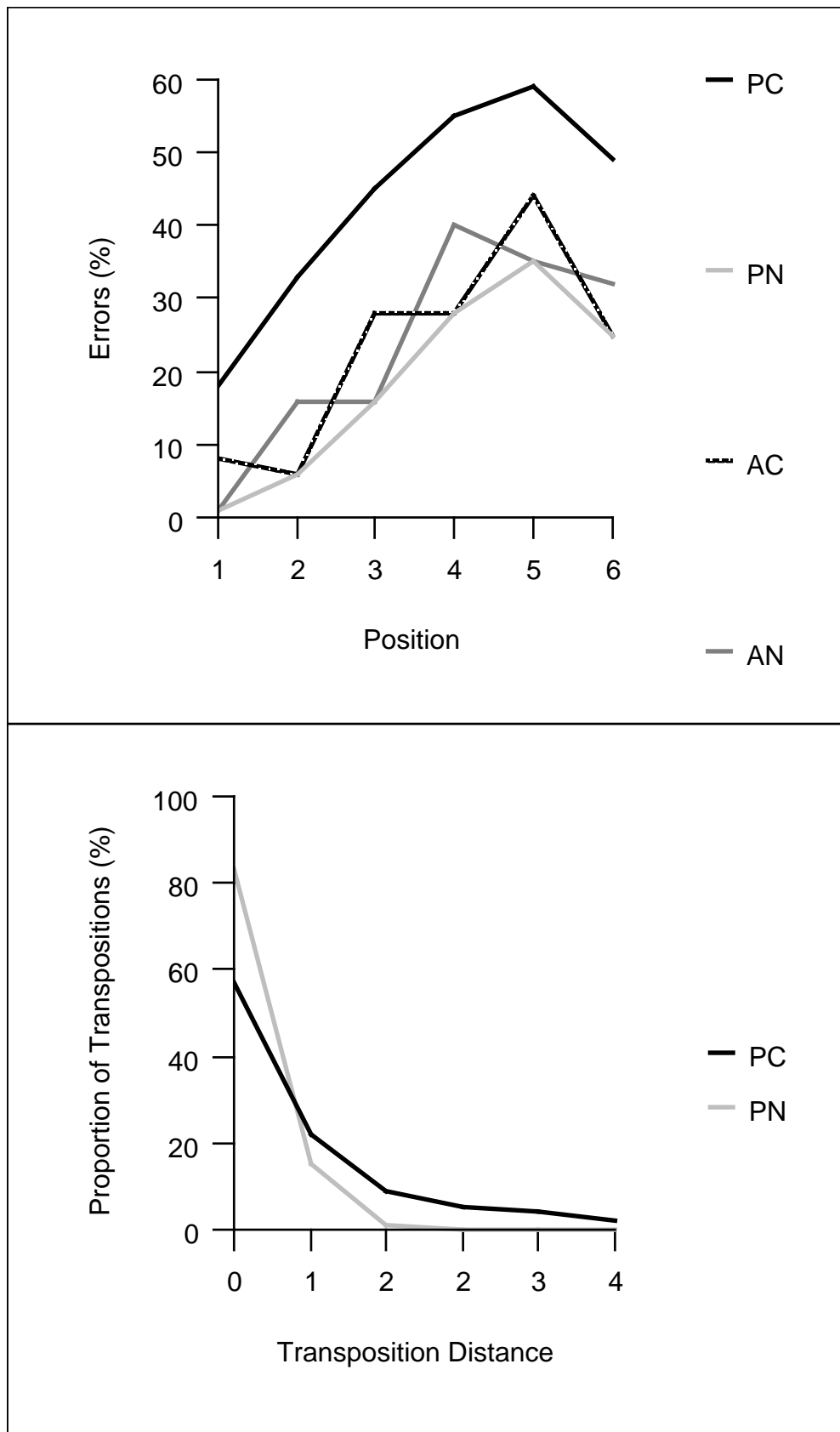


Figure 5-10: Errors by position (upper panel) and proportion of transpositions (including correct responses) by transposition distance (lower panel) from SEM in Fit 4.

Fit 5. List Length, Grouping and Interpositions

Previous fits showed that SEM can model serial recall of five and six items without changing start or end marker parameters. The next question was how SEM extends to serial recall of seven, eight and nine items, as in conditions U7, U8 and U9 of Experiment 3. In addition, SEM was fitted to the grouping in the G9 condition of Experiment 3. Extending SEM to grouped lists is a simple conceptual step, and one which illustrates the utility of start and end markers as anchor points in serial ordering.

The basic idea behind modelling grouping in SEM is that two dimensions of position are coded. The first is the position of an item in a group; the second is the position of a group in a list. These dimensions are coded with respect to two pairs of start and end markers, resulting in two positional codes for each token. For example, after encoding of the sequence *RMQ JHV*, short-term memory would contain six tokens like those depicted below:

<	{R}	< 1.00	0.36 >	< 1.00	0.60 >	>
<	{M}	< 0.60	0.60 >	< 1.00	0.60 >	>
<	{Q}	< 0.36	1.00 >	< 1.00	0.60 >	>
<	{J}	< 1.00	0.36 >	< 0.60	1.00 >	>
<	{H}	< 0.60	0.60 >	< 0.60	1.00 >	>
<	{V}	< 0.36	1.00 >	< 0.60	1.00 >	>

The leftmost positional code represents position of item-in-group; the rightmost code represents position of group-in-list (assuming the same start and end marker parameters in both cases). The cue for each response would also contain two such positional codes.

The effect of adding a second dimension of positional coding is shown in Figure 5-11 (where $S_0=E_0=1.00$ and $S=E=0.60$ for both item- and group-level markers). The positional uncertainty functions are obtained by multiplying the positional overlap between item-level and group-level codes (Appendix 3). The upper panel shows how the positional uncertainty functions for the middle position in an ungrouped list flatten as the list length increases from three to five to seven to nine. In other words, the positional uncertainty increases as list length increases. This is because the positional codes vary within fixed bounds of $\langle 1 \ 0 \rangle$ and $\langle 0 \ 1 \rangle$, and therefore have only a finite resolution. As the number of positions coded within this range

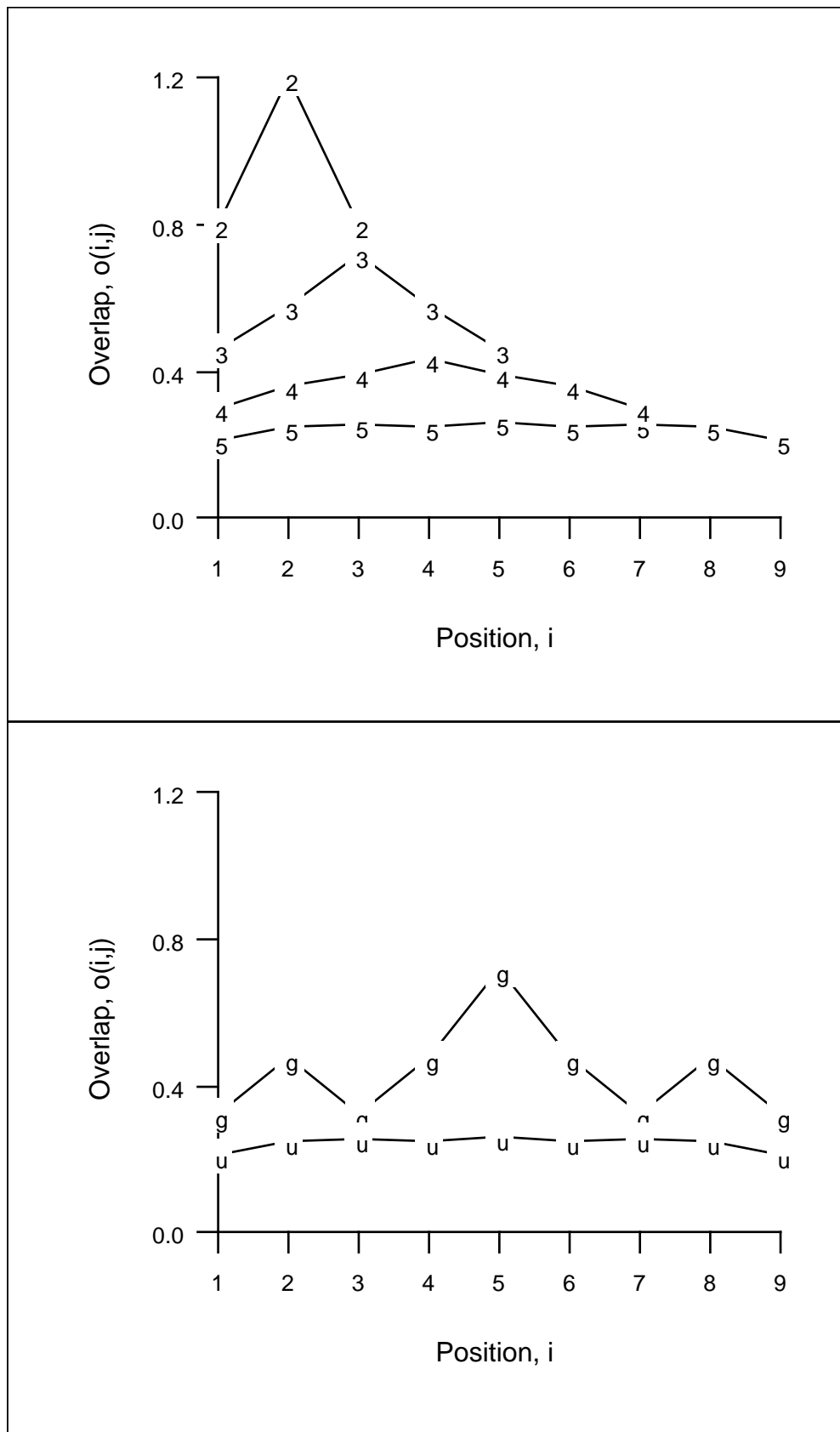


Figure 5-11: Positional uncertainty functions for the middle Position $j=2,3,4,5$ of three-, five-, seven- and nine-item ungrouped lists (upper panel), and the middle Position $j=5$ of ungrouped (line of 'u's) and 3-3-3 grouped (line of 'g's) nine-item lists (lower panel).

increases, the resolution of each code decreases. This is consistent with what evidence there is from positional probe tasks (e.g., Murdock, 1968). It is also an automatic consequence of employing start and end markers. It is not a property of other positional codes, such as the context signal of Burgess and Hitch (1992; 1996b).

It becomes virtually impossible to discriminate Position 5 in an ungrouped, nine-item list (the line of 'u's in the lower panel of Figure 5-11). When this list is grouped as three groups of three however (the line of 'g's'), the positional uncertainty function for Position 5 is much sharper, particularly with respect to immediately surrounding positions. This is because the start and end markers at the item-level are only coding three, rather than nine, positions. This means grouping improves discrimination of positions within groups (as well as between groups) explaining why the proportion of transpositions within groups is decreased by grouping (Experiment 2). This does not appear true of other models, such as Brown et al. (1996), Burgess and Hitch (1996b), or Lee and Estes (1981), where grouping only reduces the proportion of transpositions between groups (see also Frick, 1989). The slightly greater positional overlap between Position 5 and Positions 2 and 8 in the grouped list reflects the fact that these positions share the same code for position of item-in-group, and differ only in their code for position of group-in-list. It is these multiple peaks in the positional uncertainty function that produce interpositions (Chapter 3).

Adding a second set of start and end markers entails more parameters. Again however, the start marker parameters were fixed in all subsequent fits and the end marker parameters were expressed as a ratio of the start marker parameters. This produced four parameters: $F_{0,I}$ and F_I for item-level markers, and $F_{0,G}$ and F_G for the group-level markers.

A second major addition is the assumption of noise associated with positional codes. This noise is assumed to reflect random fluctuations in the encoding and reconstruction of positional codes (e.g., random shifts of attention across positions). Noise at the group-level is necessary to account for some degree of dependence between retrieval of items within the same group (Experiment 2). Positional noise was characterised by two new parameters, D_I and D_G , the standard deviations of zero-mean Gaussian noise at the item- and group-level respectively. A final addition is the assumption of two new thresholds, reflecting the minimum

degree to which the positional codes of tokens must overlap with those of the cue. Items or groups of items whose positional codes do not match this criterion do not enter the competition for output. This is necessary to explain why whole groups are occasionally omitted (Chapter 3). These thresholds are parameterised by M_I and M_G for the item- and group-level respectively (for more details, see Appendix 3).

SEM was fitted to all four conditions of Experiment 2, with a total of eight free parameters. The parameter values $F_{0,G}=0.60$, $F_G=1.00$, $M_I=0.40$, $D_G=0.08$, $M_G=0.85$ were constant across conditions. The parameters $F_{0,I}$, F_I and D_I changed between ungrouped and grouped conditions. For the ungrouped conditions, the parameters $F_{0,I}=0.60$, $F_I=0.75$, $D_I=0.04$ reflected people's ability to code position of an item in the list. For the grouped condition, the parameters $F_{0,I}=1.00$, $F_I=0.25$, $D_I=0.16$ reflected people's ability to code position of an item in a group.⁶ The remaining parameter values were the same as in Fit 4.

The fit to 33 data points from error position curves for each condition in Experiment 2 gave an RMSE of 12.58% (Figure 5-12; cf. Figure 3-1). The main reason for the poor fit was the presence of spontaneous grouping in the ungrouped conditions of the data (Chapter 2). Indeed, separate Hotelling T^2 -tests for each condition showed that the discrepancy was located mainly in the longer ungrouped lists, with $T^2=4.83$, $F(7,11)=0.45$, $p=.85$ for condition U7, $T^2=32.68$, $F(8,10)=2.40$, $p=.10$ for condition U8, $T^2=34.00$, $F(9,9)=2.00$, $p=.16$ for condition U9, and $T^2=5.59$, $F(9,9)=0.33$, $p=.94$ for condition G9. Another reason for the poor fit was that the current version of SEM does not allow for the delay between presentation and recall of each item, which is necessary to explain how list length exerts such a large effect on the first positions of recall (Experiment 2). An extended version of SEM that incorporates the effects of delay, giving better error position curves, is shown in Fit 10 of Appendix 3.

In spite of the poor fit to error position curves, the current version of SEM provided a good fit to the effects of list length and grouping on the different error types (Table 5-2; cf. Table 3-1). The RMSE over the 8 data points was only 1.46%. In ungrouped lists, the main effect of list length was to increase the incidence of omissions, with a smaller increase in

6. In fact, SEM treats a list as a large group. The difference in the item-level parameters in ungrouped and grouped conditions reflects the procedural differences between lists and groups.

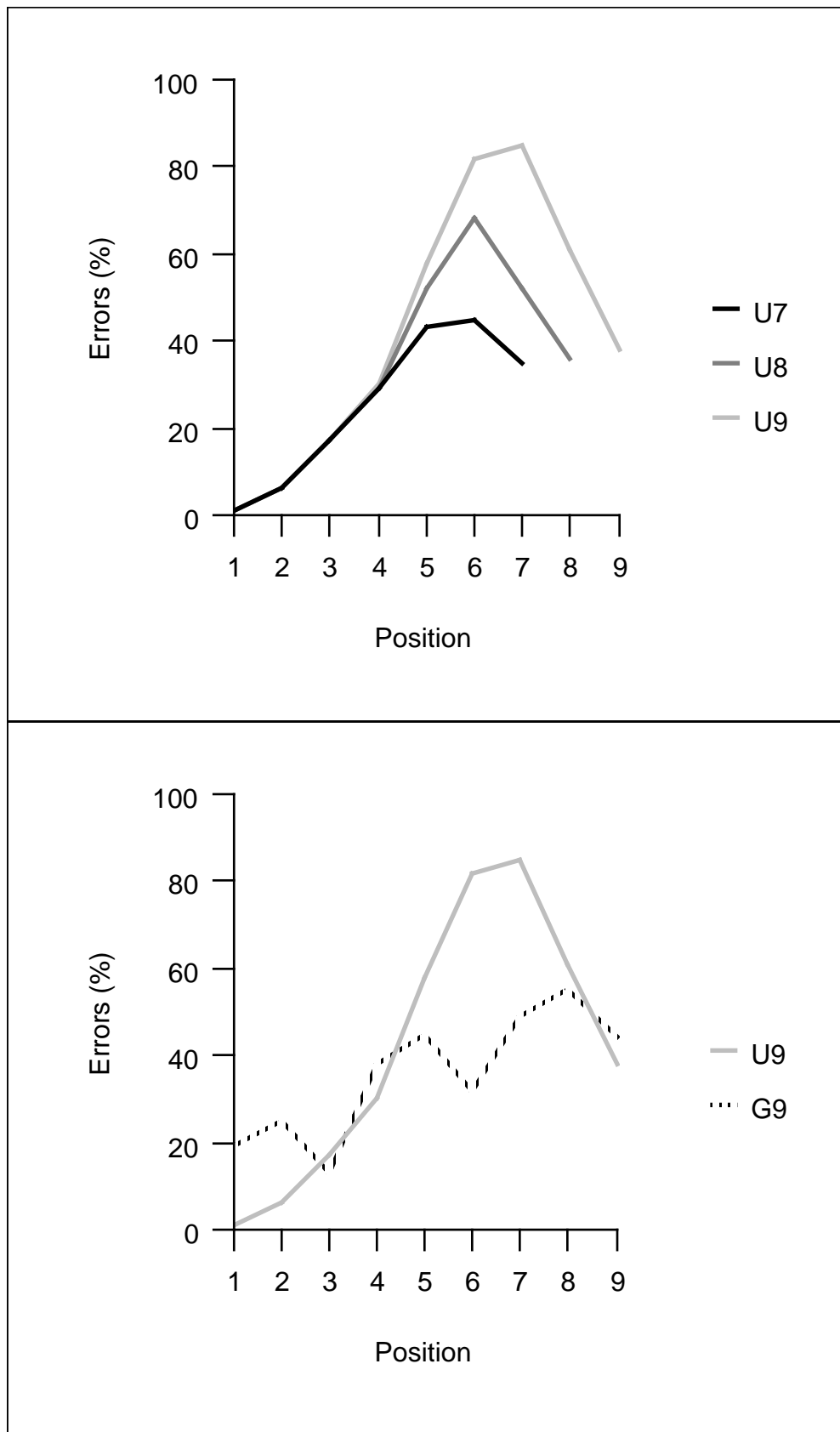


Figure 5-12: Errors by position for ungrouped lists (upper panel) and nine-item lists (lower panel) from SEM in Fit 5.

transpositions. In SEM, these increases arise because longer lists have both lower and flatter positional uncertainty functions (Figure 5-11). The lowering of these functions causes more omissions and the flattening causes more transpositions (i.e., a lower signal-to-noise ratio). Conversely, the effect of grouping was to decrease both omissions and transpositions, by raising and sharpening positional uncertainty functions (Figure 5-11).

Condition	Omissions	Transpositions
U7	.06	.19
U8	.12	.21
U9	.18	.24
G9	.13	.23

Table 5-2: Frequency of omissions and transpositions from SEM in Fit 5.

The error distribution in grouped lists is particularly important. The upper panel of Figure 5-13 shows the transpositions and omissions produced by SEM for condition G9. Transpositions showed the scalloped curves, with primacy and recency within each group, while omissions were flatter within groups, but increased across groups (cf. Figure 3-2 and Figure 3-3). The marked reduction in transpositions at the end of groups arose because of the strong and sharply tuned end marker at the item-level (providing accurate coding for the last position in group). This reflects the distinctive nature of the end of groups. The monotonic increase in omissions across groups arose because of the relatively weak end marker at the group-level. When combined with a high positional threshold, the latter can produce a failure to retrieve whole groups (Chapter 3).

The lower panel of Figure 5-13 shows the transpositions produced by the model for the U9 and G9 conditions. In the ungrouped condition, the locality constraint was respected, with a monotonic decrease in transpositions as transposition distance increased. In the grouped condition however, there was a peak for three-apart transpositions. This peak reflected transpositions between groups that maintain their position within groups (cf. Figure 3-4). Further analysis of these interpositions revealed that a greater proportion arose between the

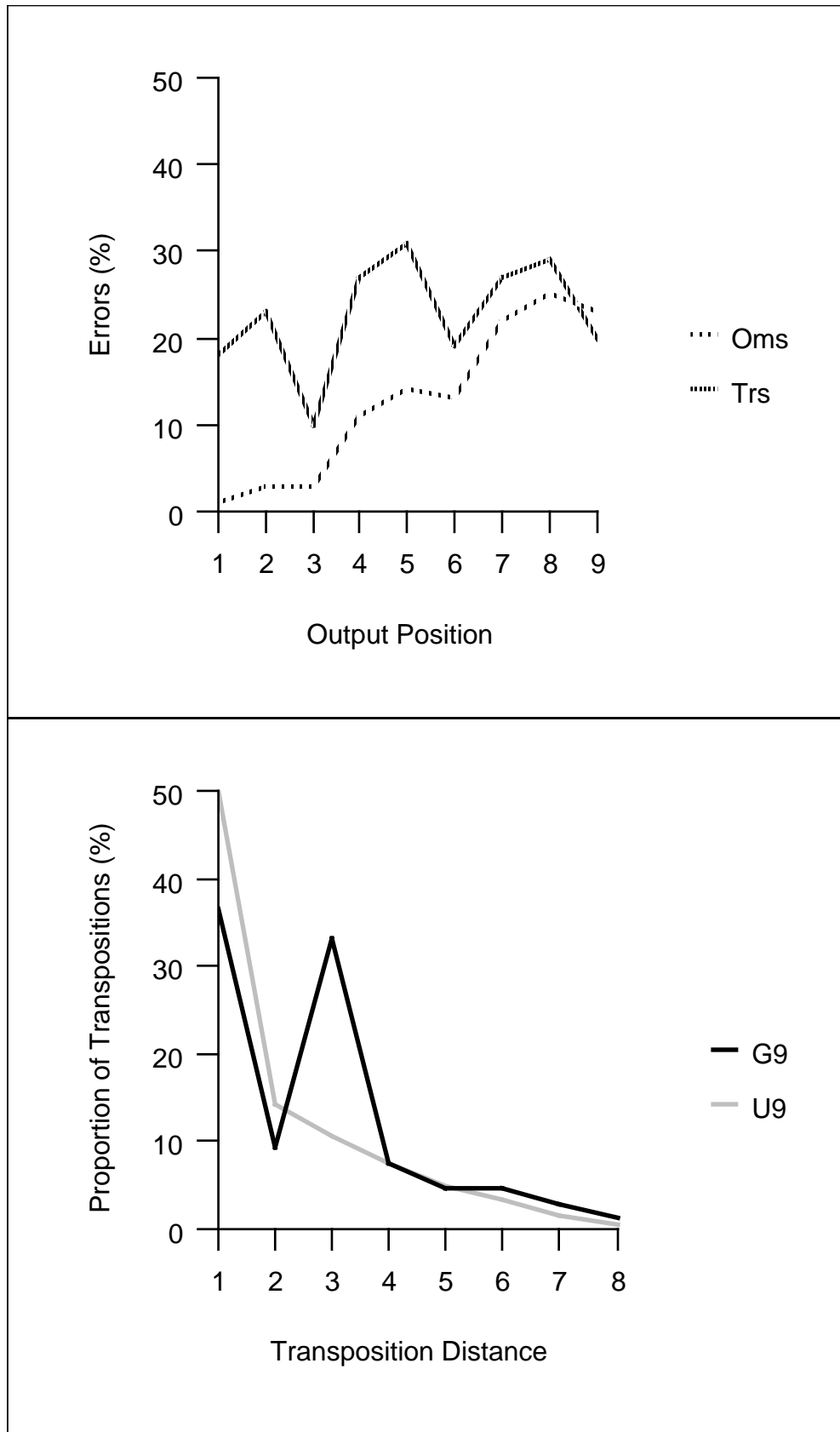


Figure 5-13: Omissions and transpositions by output position (upper panel) and proportion of transpositions by transposition distance (lower panel) from SEM in Fit 2. (Trs=Transpositions; Oms=Omissions.)

middle of groups (.37) than the start (.31) or end (.32) of groups, in agreement with the data (Experiment 2). This was because the peaks of positional uncertainty functions for middle positions of a sequence are lower than for start or end positions (Figure 5-2), meaning smaller differences in overlaps between middle positions of different groups than between terminal positions of different groups. This point is elaborated in the context of protrusions in Fit 6.

A further important property of the interpositions produced by SEM is that they arose singly, but not completely independently. Though they were rarely the result of whole groups swapping, 21% of interpositions in the present fit were followed by another interposition from the same group, a figure close to the 18% in the data, and significantly greater than the chance level of 11%. In other words, there was some dependency between recall of items in the same group. This dependency arose because of noise in the positional codes at the group-level. When the noise is great, and a group's position in the list is poorly encoded or poorly reconstructed at retrieval, recall of all items in that group is affected (and similarly, noise at the list-level can affect recall of whole trials). This is contrary to the independent perturbation assumption of the Perturbation Model (Lee & Estes, 1977; 1981).

Finally, though the figures were slightly higher than in the data, grouping decreased the proportion of transpositions within groups, from .49 in condition U9 to .40 in condition G9. This is consistent with the data from Experiment 2, but not with other models of grouping, which only predict a reduction in transpositions between groups (above; Chapter 3).

In summary, SEM gave an excellent fit to the full pattern of errors in the grouped condition of Experiment 2. Its fits to the ungrouped conditions were not so good, but this was to be expected, given the spontaneous grouping in these conditions, for which the current version of SEM made no allowance. In fact, the error position curves produced by SEM for ungrouped eight and nine item lists may be more imaginary than real: It may be that people can never recall such long sequences without spontaneously grouping them into smaller subsequences (Chapter 3). Indeed, this would be expected from SEM's positional uncertainty functions (Figure 5-11): Once the list length exceeds five or more items, the positional coding of middle positions becomes very hazy, and insufficient to support serial recall. However, by inserting additional anchor points within a sequence, in the form of additional start and end

markers, the positional uncertainty can be reduced. Thus SEM not only provides a rationale for the limited capacity of people's short-term memory for serial order, but also provides a rationale for people's spontaneous grouping of long lists.

Fit 6. Intertrial Interval and Protrusions

Previous fits have used the single-trial version of SEM (Appendix 3). To model intertrial effects such as proactive interference, the multiple-trial version is necessary. This version contains three further assumptions. The first assumption is another component to tokens, representing the general context during their encoding. This context is nonpositional (i.e., cannot be reinstated at recall) and represents all other intrinsic (e.g., mood) and extrinsic (e.g., environmental) factors that change over time. General context is modelled by a single value, a one-dimensional vector. For mathematical convenience, the current context is represented by the value 1, and older contexts are represented by decreasing values less than 1 (i.e., rather than updating the current context, the context of existing tokens in memory is multiplied by a parameter $E_C < 1$ for each contextual change).

With the addition of general context, tokens contain three contextual vectors, two positional (coding positions of item-in-group and group-in-list) and one nonpositional (general context). Immediately after presentation of the first of two groups, *RMQ*, short-term memory would contain three tokens like those depicted below:

<	{R}	<	1.00	0.36	>	<	1.00	0.60	>	<	0.96	>	>
<	{M}	<	0.60	0.60	>	<	1.00	0.60	>	<	0.98	>	>
<	{Q}	<	0.36	1.00	>	<	1.00	0.60	>	<	1.00	>	>

where the rightmost code represents the general context (with $E_C = 0.98$). The more recent the encoding of tokens, the less is the change in their general context.

During recall, the general context for each cue is always the current context. The overlap between the general context of the cue and the general context of each token is determined in exactly the same manner as for positional codes (via Equation 5-2), and the combined positional uncertainty functions are determined from multiplying the overlaps of the three contextual vectors (Appendix 3).

The addition of general context entails six new parameters. The first is the rate of contextual change, E_C . This change is assumed to occur between episodes, where C_P is the number of episodes between presentation of each item (a function of the presentation rate), C_D is the number of episodes during the delay before recall (a function of the retention interval), C_R is the number of episodes between recall of each item (a function of the recall rate) and C_I is the number of episodes between trials (a function of the intertrial interval). These four parameters are fixed by the experimental design. The last parameter, C_A , represents the amount of intrinsic contextual change between trials due to attentional shifts. An example attentional shift between trials is when one “thinks of something else”, in order to put the previous trial out of mind. This illustrates the difference between contextual change and real-time change: Contextual change may be either slower or faster relative to the passage of time. A great deal of cognitive activity may take place in the few seconds during or between trials, resulting in large differences in intrinsic context over a small length of time. Thus the notion of context used in SEM is not just a case of relabelling time (Baddeley & Hitch, 1993).

With the simple example of $E_C=0.98$, $C_P=C_D=C_R=0$ and $C_I=C_A=20$, the peaks of the positional uncertainty functions for two successive lists of five items are shown in Figure 5-14 (ignoring any group-level codes and using the same start and end marker parameters as in the upper panel of Figure 5-2). The line of ‘c’s represents the overlap between the cue for Position i and a token at Position i in the current trial; the line of ‘p’s represents the overlap between the cue for Position i and a token at Position i in the previous trial (the overlap for different positions within trials will always be less). The peaks of the positional uncertainty functions for the previous trial are lower than for the current trial, to the extent that the general context has changed between trials (due to the multiplicative nature with which overlaps are combined). Occasionally however, the difference between positional overlaps for tokens in the two trials is bridged by additive noise, resulting in an intrusion (most often a protrusion).

The second major assumption is that recall of an item creates a new token in short-term memory. Importantly, the item recalled is recoded in its output position (which may or may not be correct), and its general context is updated to the current context (Figure 5-15). This process of “reperception” also reactivates the phonological representation of the item, akin to

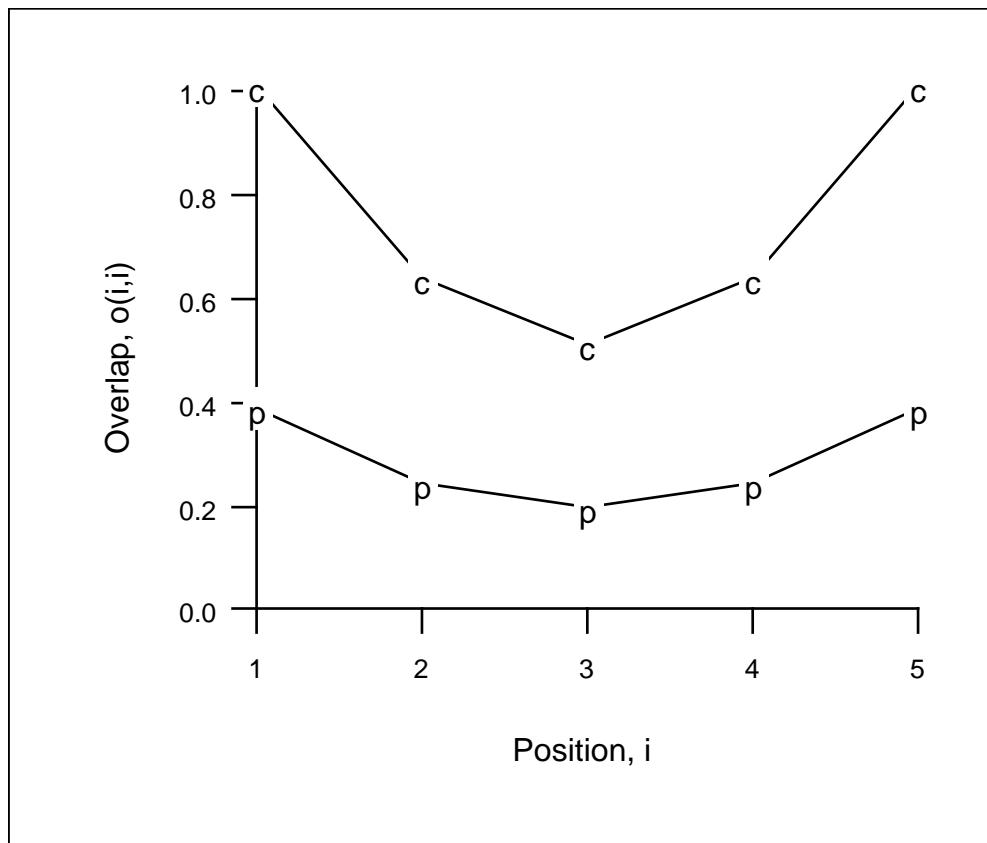


Figure 5-14: Example peaks of positional uncertainty functions for the current trial (line of 'c's') and previous trial (line of 'p's').

the refreshing role of Baddeley's (1986) notion of subvocal rehearsal. In other words, the continual updating of positional and item information corresponds to maintenance rehearsal in short-term memory (and rehearsal and recall are equivalent in this sense).

The final assumption is that the activations of SEM's phonological representations also decay during presentation, retention, recall and intertrial intervals. Like the decay of suppression, the decay of phonological activations is exponential and assumed to occur in real-time (Appendix 3). The rate of decay is characterised by the last new parameter, R_p . In the present fit, this decay operates during the same intervals characterised by C_p , C_D , C_R and C_I . Decay during presentation produces a "recency-gradient" of phonological activations, as might be expected from item recognition tasks (e.g., Monsell, 1978).

Existing data suggest that the decay of phonological information is quite rapid (Baddeley, 1986). For example, phonological confusions disappear after a short, filled delay

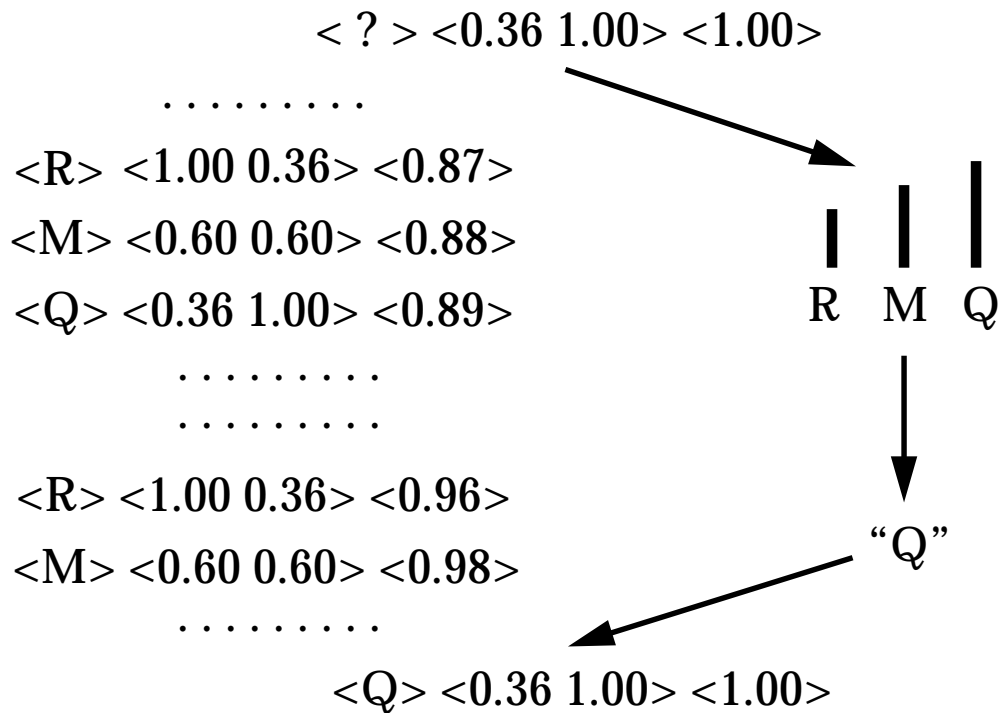


Figure 5-15: Schematic representation of SEM's rehearsal process.
(Group-level codes not shown.)

(e.g., Conrad, 1967; Bjork & Healy, 1974; Estes, 1973). In SEM, confusions disappear when phonological activations have decayed completely (Fit 9 in Appendix 3). Phonological decay and contextual change during presentation and recall also afford SEM closer fits to list length effects (Fit 10, Appendix 3) and word-length effects (Fit 11, Appendix 3).

The multiple-trial version of SEM was fitted to both conditions of Experiment 3, with a total of 5 free parameters. The parameters C_P , C_D , C_R and C_I were fixed by the experimental design. Specifically, the values $C_P=1$, $C_D=3$, $C_R=1$ were constant across conditions, while C_I reflected the length of the filled intertrial interval, with $C_I=2$ for the Short condition and $C_I=20$ for the Long condition (Experiment 3). These values reflected the number of episodes, where an episode corresponded to the presentation or recall of an item (C_P , C_R), or the presentation of a distractor (C_D , C_I). The free parameters E_C , R_P and C_A were set to $E_C=0.98$, $R_P=0.05$ and $C_A=20$. The remaining two free parameters were set to $G_C=0.10$ and $T_O=0.70$, whose values were changed from Fit 5 because of the new assumptions of SEM (e.g., phonological decay). The remaining parameter values were identical to those in Fit 5.

The fit to the 10 data points in the error position curves of each condition in Experiment 3 gave an RMSE of 6.43%, a difference that was not significant, $T^2=5.63$, $F(10,8)=0.27$, $p=.97$. SEM produced the full range of transpositions, omissions, repetitions and intrusions, with the RMSE to 40 data points from error position curves for each error type being only 3.86%.

The most important errors in the present fit were intrusions. The frequency of intrusions in the Short condition (.07) was greater than in the Long condition (.03). The majority of these were immediate intrusions from the previous report, owing to the recoding of items during recall. The proportion of such intrusions that were output protrusions was .46 and .34 for the Short and Long conditions respectively (cf. Table 3-3). The higher frequency of protrusions with the short rather than long intertrial interval was not found in the data, but this may reflect the considerable noise in the data, particular with the small numbers involved and effects of guessing (Experiment 3). A version of SEM that allowed for guesses, in a manner to be described in Chapter 6, produced an even closer fit to the data.

Intrusion gradients for each position are shown in Figure 5-16, collapsed across Long and Short conditions. SEM produced a pattern of intrusions similar to that in the data (Figure 3-6). Indeed, the RMSE over all 25 points in the lower panel was only 6.71%, which was a good fit given the noise associated with the relatively small numbers in the data.

Comparison of the two panels of Figure 5-16 reveals that, though there were fewer intrusions on the first position than the middle position, the proportion that were protrusions was greater. This pattern is in agreement with the data (Figure 3-6) and follows from SEM because of the following reasons. Positional uncertainty functions for the previous trial are flatter than for the current trial, meaning that the difference between trials is larger for the first position than for the middle position (Figure 5-14). Because larger differences are harder to bridge with additive noise, there will be fewer intrusions on the first position than the middle position. A similar reasoning explains the higher frequency of interpositions between the middle than start or end of groups (Fit 5; Experiment 3) and why the proportion of errors that are protrusions decreases as retention interval increases (Conrad & Hull, 1966, as confirmed in Fit 9 of Appendix 3). However, because the positional uncertainty functions within trials are

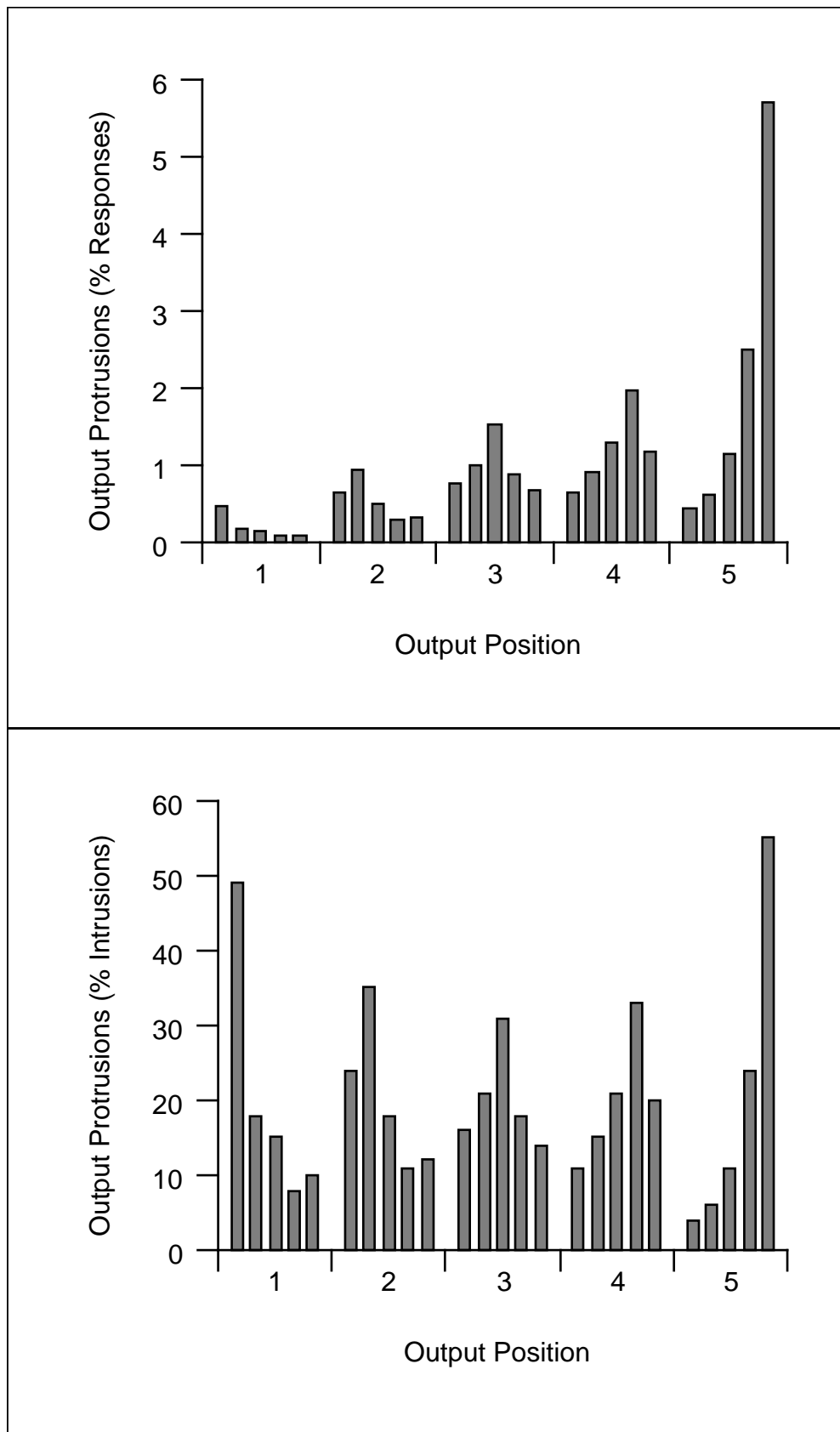


Figure 5-16: Output intrusions as a proportion of responses (upper panel) and as a proportion of intrusions per output position (lower panel) from SEM in Fit 6.

sharper for the first position than the middle position (Figure 5-2), a greater proportion of intrusions that do occur on the first position will be protrusions.

The pattern of intrusions on the last position is complicated by the effect of errors on earlier positions. For example, if the last item of the current trial is recalled too early and suppressed, there is a greater likelihood of an intrusion following on the last position (for the same reason that omissions and repetitions are likely on this position; Fit 2 and Fit 3). Hence, most intrusions in Figure 5-16 were on the last position. This is not true of the data from Experiment 3, though it is true of other data, such as that from Experiment 5 (Figure 6-4 in Chapter 6). The reason why it is not true of the data from Experiment 3 is unclear, though one possibility may reflect the difficulty of indicating omissions appropriately in spoken recall. Nonetheless, the important aspect of the present fit is that SEM can be readily extended to proactive interference between trials, producing the appropriate pattern of intrusions from the previous report as a function of the intertrial interval.

Summary of SEM's Fits

The six fits above show that SEM can model the effects of primacy, recency, phonological similarity, list-length, grouping and proactive interference in short-term memory. More specifically, SEM can capture the complete pattern of errors, including transpositions, omissions, repetitions, confusions, interpositions and protrusions, and the important constraints on their distribution (i.e., all nine constraints in Chapter 4). It was argued that other models of short-term memory fail to meet one or more of these constraints.

To allow such coverage, the full, multiple-trial version of SEM has a considerable number of parameters. However, only a fraction of these were free to fit each data set. The remaining parameters were constrained by the experimental procedure (e.g., the intertrial interval, C_I), or kept constant throughout fits (e.g., the decay of suppression, R_S). Some parameter changes across fits were necessary because of the incremental exposition of SEM (e.g., the decrease in T_O with the introduction of phonological decay in Fit 6). The parameters that were truly free across fits (e.g., noise in the competition stage, G_C) were necessary to accommodate differences between experiments beyond present concerns (e.g., the particular

stimuli, presentation modality, or method of recall employed).

Finally, despite some variation in parameter values required for different quantitative fits, the qualitative behaviour of SEM is fairly robust to parameter changes. For example, the nine empirical constraints are met under a wide range of parameter values (within sensible limits). This robustness results from SEM's core assumptions of positional coding, separate storage of tokens, and a recall process of noisy choice and suppression.

Extension to Other Phenomena

SEM can also be extended to other important phenomena in short-term memory. Though demonstrating fits to such data would exceed the present remit, the general approach which SEM might take is outlined below.

Serial Recall

Much research on serial recall has been performed under the *working memory* framework (Baddeley, 1986). In particular, research has focused on the *phonological loop*, a component of working memory assumed to underlie short-term memory for verbal material. The phonological loop has two components: a short-lived phonological store susceptible to decay, and an articulatory control process, which allows rehearsal of material in the phonological store and which is required to encode visual material in that store. In general terms, the transient phonological activation in SEM corresponds to the phonological store, while rehearsal in SEM corresponds to use of the articulatory control process. This bipartite approach proves useful in providing a unified account of the interactions between articulation rate, phonological similarity, irrelevant sound and articulatory suppression.

Articulation Rate

If rehearsal prevents decay of phonological representations, the rate of rehearsal will be an important determinant of short-term memory. Rate of rehearsal appears related to rate of articulation. Evidence for this comes from the *word-length effect*: Span is smaller for words that take longer to articulate, even when balanced for number of syllables and phonemes (Baddeley, Thomson & Buchanan, 1975). The most convincing demonstration of this effect is that the digit span of bilinguals is greater in the language in which the digits are articulated faster (Ellis & Hennesly, 1980). The relationship between span and articulation rate is linear

and implies that span for verbal material is approximately equal to the number of items that can be articulated in two seconds (Baddeley, 1986).⁷ Thus span is not simply a fixed number of chunks (Miller, 1956/1994; Schweikert & Boruff, 1986), as might be suggested from previous fits of SEM.

A first approximation to modelling word-length in SEM is by varying C_P and C_R . The longer the words, the greater C_P and C_R , reflecting a greater opportunity for phonological decay during presentation and recall. Fit 11 in Appendix 3 demonstrates that SEM can produce a relationship between span and articulation rate that is close to that in the data (Hulme, Maughan & Brown, 1991). Like the Primacy Model (Page & Norris, 1996b), decay during recall explains the greater impairment when long words are recalled before short ones (Cowan et al., 1992) and decay during presentation explains the effect of word-length on the first item recalled (Page & Norris, 1996a).

The above fit is only a first approximation because it does not take into account covert rehearsal during presentation and recall. Most subjects report some attempt at rehearsal during these intervals. Indeed, according to the working memory theory, covert rehearsal is necessary to explain why memory can extend beyond presentation, retention and recall intervals longer than a few seconds. Rehearsal during presentation also explains why presentation rate has little effect on serial recall: the greater potential for phonological decay with slow presentation is offset by a greater opportunity for rehearsal. When rehearsal is prevented by concurrent articulatory suppression, slow presentation rates do impair recall (Baddeley & Lewis, 1984).

Three different rehearsal strategies can be distinguished. During the pause between presentation of Item $N-1$ and Item N , rehearsal can be *repetitive*, where Item $N-1$ is repeated as many times as possible, *associative*, where Items $N-2$ and $N-1$ are repeated together as many times as possible, and *cumulative*, where as many items from Item 1 onwards are rehearsed as possible. Without instruction, the modal strategy is cumulative (Page & Norris, 1996a). With instruction, cumulative rehearsal is generally superior to associative rehearsal (Palmer & Ornstein, 1971; Ferguson & Bray, 1976).

7. Though rehearsal is associated with articulation in the working memory theory, this is not actually enforced by the correlation between span and articulation rate, because anarthric children show normal word-length effects (Bishop & Robson, 1989), suggesting that rehearsal and articulation may both rely on more central processes.

One argument for cumulative rehearsal is that it minimises the delay between each item's input and output, reducing phonological decay. SEM also suggests a further reason. For an item to be coded in a position, it must appear in a sequence of items. Neither repetitive nor associative rehearsal allow this (associative rehearsal simply codes the relative order of two items). Only cumulative rehearsal allows coding of position. Furthermore, the nature of the positional code in SEM will change as the number of items rehearsed increases (owing to the influence of the end marker). This may be important if the list length is unknown (Chapter 6).⁸

Thus the effect of word-length in Fit 11 may be better described as the effect it has on covert rehearsal: fewer long words than short words can be rehearsed covertly between presentation and recall of items. Though covert rehearsal has not been modelled explicitly in SEM, it could be modelled implicitly in the values of C_P , C_D and C_R , by making the same assumption of "time since last rehearsal" of Page and Norris (1996b).

Because phonological decay is assumed more rapid than contextual change, the word-length effect in SEM is attributable mainly to the former. Thus a word-length effect would not be expected after 15 seconds of distraction between items (Cowan, Wood & Borne, 1994), because the large delay between an item's presentation and rehearsal means that phonological activations will have decayed almost completely (and any difference in general context between short and long words will be negligible compared to that between positions). Because recall can still be supported by contextual and positional cues however, performance will remain above the chance-levels predicted by the phonological loop in such situations. With an approximate half-life of two seconds, the phonological store, unlike SEM, is not able to support serial recall when rehearsal is prevented for more than a few seconds.

In summary, SEM appeals to the same decay-based account of word-length effects, mediated by rehearsal, as the phonological loop. Though there are other accounts of the word-length effect that do not appeal to decay (Neath & Nairne, 1995) or rehearsal (Brown & Hulme, 1995), there is little to favour these accounts, particularly since they have overlooked

8. The nature of positional codes in SEM does present a problem when there is not enough time between each item for a complete cumulative rehearsal. For example, if Items 1-6 are rehearsed before Item 7, but there is no time to rehearse Items 1-7 before Item 8, then the positional codes for Items 1-6 (particularly later items) will reflect a six- rather than eight-item list. This is more likely for supraspan lists, and may be another reason why such lists are spontaneously grouped. Most models have problems modelling such displaced rehearsals explicitly.

the related problem of serial order. However, because SEM does not rely on phonological activations for serial recall, it can explain short-term memory for serial order in situations beyond those explicable by the phonological loop (e.g., Fit 9 in Appendix 3).

Phonological Similarity

Fit 4 demonstrated how SEM models phonological similarity, an effect contingent on transient activation of phonological representations. These activations may correspond to Baddeley's phonological store, though one that stores mainly item rather than order information. With a short filled-delay, decay of these activations is appreciable, explaining the rapid forgetting over the first few seconds (e.g., Peterson & Peterson, 1959). In SEM, this forgetting reflects an increase in omissions and transpositions, together with a reduction in confusions (Fit 9 in Appendix 3), consistent with the data (e.g., Bjork & Healy, 1974).

The phonological store is not specific in how phonological similarity affects its contents. In SEM, the locus of the phonological similarity effect is a second stage of item retrieval, an assumption shared with other models (Lee & Estes, 1977; Page & Norris, 1996b). This was necessary to explain why phonological confusions do not affect surrounding nonconfusable items. This assumption has support from models of speech production in which lexical retrieval precedes independent phonological retrieval (Levelt, 1989).

The role of phonological information in SEM is also consistent with that suggested by Tehan and Humphreys (1995). They observed that immediate recall of subspan lists showed no detectable proactive interference, but clear evidence of phonological confusions. With a short delay however, proactive interference emerged and phonological confusions disappeared. They attributed this to short-lived phonological information that overcomes any proactive interference. In SEM, this information corresponds to the rapidly-decaying phonological activations, which aid discrimination of items between lists, because more recent items have more active phonological representations, but impair discrimination of items within lists, because more active phonological representations are more easily confused.

SEM's treatment of phonological information is simplified however. Phonological similarity clearly requires more than the simple metric p in SEM (Appendix 3). With lists of nonsense syllables, similarity is a function of syllable structure and distinctive phonemic

features (Ellis, 1980). Confusions involve the movements of consonants rather than vowels, particularly onsets (Drewnowski, 1980b). These movements respect position within syllables, so that onsets are only likely to swap with other onsets, to form new syllables (Treiman & Danis, 1988). Even with the familiar items in the experiments considered in Chapter 4, there was evidence for a similar type of blend error. *Blends* are intrusions that are recombinations of the phonemes of list items (Drewnowski & Murdock, 1980). For example, when a list contains *J* and *V*, a common blend is *G*, containing the onset of *J* and the rhyme of *V*. Though rare, such intrusions are more common than intrusions of other similar letters, such as *B*, *P*, or *T*. Phonological retrieval is clearly a more complex process than currently modelled in SEM.

Other effects of phonological similarity, such as its interaction with the modality effect (Drewnowski, 1980b; Murray, 1967; Watkins, Watkins & Crowder, 1974) and grouping (Frick, 1989) require further simulations of SEM. More problematic is the suggestion that the redundancy of vowels over trials is a critical factor, and more important than their similarity (Drewnowski, 1980b). These issues are yet to be addressed fully by any model.

Irrelevant Sound

Concurrent irrelevant speech during a serial recall task impairs performance, to a greater extent than comparable noise levels, and sometimes as a function of phonological similarity between relevant and irrelevant material (Salame & Baddeley, 1982). According to the working memory account, the irrelevant material has automatic access to the phonological store, where it interferes with the relevant material.

There are problems for this account however. Phonological similarity between relevant and irrelevant material does not always have a significant effect, and is small compared to the effect of similarity within the irrelevant material (Jones & Macken, 1995b). An impairment comparable to that found with speech has also been found with tones (Jones & Macken, 1993), especially if the tones change in pitch, location, or rhythm. This suggests an alternative “changing-state” account of the *irrelevant sound effect* (Jones & Macken, 1995a).

In SEM, irrelevant sound might increase the noise in the encoding and retrieval of tokens (e.g., the parameter M_I), rather than noise in phonological activations per se. This would cause an impairment independent of the similarity between relevant and irrelevant

material. The impairment would also be confined mainly to order rather than item errors, as suggested by the absence of an irrelevant sound effect on free recall (Salame & Baddeley, 1990). The additional noise may reflect difficulty in encoding or reconstructing positional codes; a difficulty related to the amount of change in the irrelevant stream. Irrelevant sound showing rapid changes over time (e.g., abrupt vowel transitions in speech) may interfere with the ability to mark the start and end of sequences. In particular, if irrelevant tones disrupt the ability to group (Hitch, Burgess, Shapiro, Culpin & Malloch, 1995), the results of Macken and Jones (1993, 1995a) may have arisen because the tones prevented spontaneous grouping.

In sum, SEM may be able to incorporate the changing-state account of irrelevant sound, and make contact with recent research on irrelevant tones and grouping.

Articulatory Suppression

Concurrent articulation of an irrelevant item (e.g., repeating “the, the, the...”) also impairs serial recall (Murray, 1967). More interestingly, under visual presentation, such *articulatory suppression* removes the effects of word-length (Baddeley, Thomson & Buchanan, 1975), phonological similarity (Peterson & Johnson, 1971) and irrelevant sound (Salame & Baddeley, 1982). Under auditory presentation, articulatory suppression removes the effect of word-length (providing it continues throughout presentation and recall), but does not remove the effects of phonological similarity (Baddeley, Lewis & Vallar, 1984) or irrelevant sound (Hanley & Broadbent, 1987). According to the working memory theory, articulatory suppression commandeers the articulatory control process. This not only prevents rehearsal, removing the word-length effect, but it also prevents the recoding of visual material into the phonological store. The latter removes effects of phonological similarity and irrelevant sound for visual material, which requires recoding, but not for auditory material, which has automatic access to the phonological store.

If articulatory suppression prevents covert rehearsal, SEM can explain its interaction with word-length in a similar manner. By also making the assumption that articulatory suppression prevents activation of phonological representations for visual material, SEM can explain its interaction with phonological similarity: With no activation of phonological representations, there is no effect of phonological similarity.

However, by assuming that the irrelevant sound effect arises from positional noise, it is not immediately clear how SEM can explain why the effect is removed for visual material under articulatory suppression. One possibility is that irrelevant sound and articulatory suppression exert similar, but not additive, effects. If the combined extent of impairment is limited, then an interaction between the two effects will depend on how much impairment is caused by each effect alone. If the impairment due to articulatory suppression alone is greater for visual than auditory material, there will be a stronger interaction in the visual case. Though not as simple as the working memory account, this account has greater explanatory power when applied to the effects of varying the suppression material. Macken and Jones (1995) found that articulatory suppression of changing material had a greater effect than unchanging material, and only the former removed the irrelevant sound effect. In SEM, articulatory suppression of changing material is likely to cause greater disruption of positional coding, and hence might predict a greater interaction with irrelevant sound.

Finally, one problem faced by the working memory theory is that some recall of visual material remains possible under articulatory suppression. This cannot be attributed to the phonological store, because recoding of the visual material is prevented. One possibility is to appeal to a second store, such as a visuospatial sketchpad (Baddeley, 1986). SEM does not have to appeal to additional means of storing serial order however. Though prevention of phonological activation impairs recall, items can still be recalled via their positional tokens.

In summary, SEM can be extended to most of the data supporting the working memory theory by borrowing some of its assumptions. Furthermore, it can explain why serial recall, though impoverished, remains possible both under suppression and after much longer intervals than predicted by the working memory theory. This is attributable to longer-lasting, nonphonological, positional information, necessary, for example, to explain protrusions after a filled delay of 20 seconds between trials (Experiment 3). By assuming a relation between the ease of generating positional codes and the rate of change of irrelevant material, SEM may also allow some reconciliation between the working memory and changing state theories. However, considerable work remains, especially regarding the detailed nature of phonological information in SEM.

Influence of Long-term Memory

Other important phenomena concern the effects of long-term memory on short-term serial recall. Foremost is the *lexicality effect*, whereby serial recall of lexical items (e.g., words) is superior to nonlexical items (e.g., nonwords, such as nonsense syllables, or words in an unfamiliar language), even when articulation rate is controlled (Hulme, Roodenrys, Brown & Mercer, 1995). The lexicality effect is usually additive on linear span-rate functions, affecting the intercept but not the slope (though not always, Multhaup, Balota & Cowan, 1996). The effect is reduced when subjects are trained on nonwords (Hulme et al., 1991).

In SEM, long-term memory determines the level at which an “item” is defined in short-term memory. For example, each word in a list represents a single item, or chunk (Miller, 1956/1994). Each nonword on the other hand may be better represented as a group of items, where each item is a phoneme. Both the order of nonwords and the order of the phonemes within nonwords must be stored in short-term memory, much like the groups of items in a grouped list.⁹ This extra requirement may explain the lexicality effect, though further simulations of SEM will be required to determine its exact relationship to span-rate functions.

SEM’s proposal that LTM determines the level of encoding contrasts with other explanations of the lexicality effect, where LTM affects retrieval, or *redintegration* (Brown & Hulme, 1995; Schweikert, 1993). The redintegration approach assumes that the representation in memory is sublexical, and lexical information aids reconstruction of this representation during retrieval (Frick, 1988a). Both encoding and retrieval accounts can explain why new lexical representations improve memory for unfamiliar words, but they differ in other respects. SEM’s encoding approach lends itself better to errors in recall of nonlexical items. For example, the swapping of initial or final phonemes in recall of nonwords might correspond to interpositions between groups of phonemes (though additional phonotactic constraints clearly play a role; Hartley & Houghton, 1996; Chapter 8).¹⁰ The redintegration approach lends itself

9. Indeed, one way of distinguishing groups and chunks in SEM might be whether the same start and end markers are used (for groups), or different start and end markers are used (for chunks). Interference between positional codes results in the former case (e.g., interpositions) but not the latter.

10. One way of capturing phonotactic constraints might be to model suppression at the level of articulatory features (i.e., the physical movements of articulators), rather than at the level of phonemes. Having articulated a phoneme, suppression of its articulatory features may temporarily inhibit recall of phonemes that share those features, and hence constrain the set of possible phonemes that can follow.

better to errors in recall of lexical items, such as the blends described above. More likely, lexicality affects both encoding and retrieval processes, with order stored concurrently at several levels (e.g., words, phonemes, articulatory features; Houghton, Hartley & Glasspool, 1996). In any case, “vertical” extension of SEM to multiple levels of representation is clearly an important area for further work.

Other influences of LTM include the effects of predictability (Chapter 1), semantic similarity (Brooks & Watkins, 1990; Poirier & Saint-Aubin, 1995), word-frequency (Watkins, 1977) and word-likeness (Gathercole & Martin, 1996). The effect of semantic similarity is to allow additional means of organising items in STM, though such organisation is normally secondary to serial organisation (Seamon & Chumbley, 1977), and much of the effect may be attributable to guessing strategies (Crowder, 1979). The effect of word frequency might reflect different baseline activations of categorical or phonological representations in SEM. The effect of predictability and word-likeness are harder to explain. They appear to reflect the number of similar sequences in LTM, clearly beyond the current scope of SEM. These more subtle interactions between STM and LTM pose problems for most models of serial recall.

Modality and Suffix Effects

SEM is currently silent on the issue of modality effects in short-term memory. It is well-known that auditory or vocalised presentation leads to better recall than silent, visual presentation, particularly for the last few items in a list (e.g., Conrad & Hull, 1968; Margrain, 1967). One possibility is an additional source of auditory information, like the Precategorical Acoustic Store (PAS) of Crowder and Morton (1969), which held a temporary “echo” of the most recent items. This store was assumed to have a small capacity, because an irrelevant item suffixed at the end of a list impaired recall of the last few items, removing the modality effect.

However, the original PAS account of modality and suffix effects proved too simple. The auditory advantage can extend over several items, and is long-lasting in the absence of further auditory input (Penney, 1989; Tell, 1971). This suggests an acoustic store that can hold several items for considerably longer than originally imagined. Interpretation of the suffix effect is not so simple because it also arises with mouthed or lipread stimuli, and appears to exert more than one effect (Baddeley & Hull, 1979; Penney, 1985, 1989).

An alternative explanation of the auditory advantage might be superior representation of serial order (Drewnowski & Murdock, 1980). One possibility is a directional auditory trace with stronger interitem associations (Drewnowski, 1980a; Penney, 1989), though this seems unlikely (Metcalf & Sharpe, 1985). Another possibility is better temporal resolution of auditory than visual material (Glenberg & Swanson, 1986), or even better positional coding (Neath & Crowder, 1990). Better positional coding would explain why Frankish (1985) found an auditory advantage on most positions of grouped lists (rather than just the last positions), particularly the end of groups, which is difficult to explain in terms of the original PAS.

One way to model better positional coding in SEM is to increase the strength or sharpness of SEM's start and end markers. The inherent temporal properties of auditory information may allow better definition of the start and end of a sequence. In fact, a stronger end marker in SEM will not only improve coding of final positions (Bunt, 1976; Glenberg, 1990), but also improve item memory for later items, particularly the last (Page & Norris, 1996a), as demonstrated in Fit 12 (Appendix 3). A stronger end marker at both the item- and the group-level would explain modality effects at the end of groups (Frankish, 1985) and perhaps differences in visual and auditory grouping (Chapter 3). If auditory presentation also entailed a stronger start marker, the modality effect may extend to the first as well as last few items, as found in probed recall (Greene & Crowder, 1988; in serial recall, the advantage for the first few items may be masked by a ceiling effect).

Though obviously a somewhat ad hoc assumption, given the currently unspecified nature of the start and end markers (Chapter 6), this approach would also explain some subtleties of the suffix effect. The auditory suffix effect is generally attenuated when the suffix differs to list items, in voicing, location, or rhythm (Frick, 1988b). The magnitude of the suffix effect may therefore depend on the degree to which it is perceptually grouped with list items (Frankish & Turner, 1984; Kahneman, 1973; LeCompte & Watkins, 1995), in agreement with the conditions for a visual suffix effect (Frick & De Rose, 1986). In SEM, perceptual grouping may determine whether the end marker includes or excludes the suffix item in coding the last position of lists. If the suffix is included, recall of the last few items will be impaired, as shown in Fit 12 (Appendix 3). Again, this can apply equally well to the coding of the last position in

groups, explaining the effect of a suffix after each group (Frankish, 1985). An additional effect of any suffix in SEM will be to increase the delay before recall, producing a slight impairment across all positions, owing to greater phonological decay (Baddeley & Hull, 1979). Though there remain aspects of the suffix effect that are difficult for a grouping account (Penney, 1978, 1985), and additional effects of semantic similarity (e.g., Routh & Frosdick, 1978), the assumptions in Fit 12 appear a reasonable first step.

Thus SEM offers a promising approach to modelling both modality and suffix effects. Nonetheless, other aspects of auditory information are necessary to explain interactions with recall order (Broadbent, Cooper, Frankish & Broadbent, 1980), precategorical properties (Crowder, 1978; Frankish, 1996), modality of other list items (Greene, 1989), and why the auditory advantage is restricted to undegraded speech sounds (Surprenant, Pitt & Crowder, 1993). This requires relating models of STM like SEM to the processes of speech perception.

Tasks other than Serial Recall

Many tests of short-term memory do not require conventional serial recall of short lists. The most obvious case is free recall, where there is no requirement for serial order.

Free Recall

Free recall also shows primacy and recency effects, but these may arise for different reasons than in serial recall, particularly for long lists. There is a large literature on free recall, which exceeds the present remit. However, in relation to SEM and the problem of serial order, two points are worth making. Firstly, with free recall instructions, actual recall order depends on list length. For short span-length lists, subjects will normally default to serial recall; for longer lists, the last few items are often recalled first, followed by the first few items (though the exact order varies between subjects and depends on factors such as modality). In SEM, positional codes are sufficient to support serial recall of short lists, but not for long lists, where codes for middle positions become indistinguishable (Fit 5). The ability to distinguish middle positions may therefore underlie the transition between serial and nonserial recall. Nonetheless, even when middle items cannot be distinguished, recall of the first few items may still be mediated by the start marker, and the last few items by the end marker (or by phonological activations). Middle items can only be weakly cued by the overlap in general

context, and so will not be recalled well, producing bowed serial position curves. Indeed, the assumption of contextual overlap makes SEM compatible with theories that explain primacy and recency in free recall in terms of contextual distinctiveness (e.g., Glenberg & Swanson, 1986; Greene, 1986). Thus, some of SEM's assumptions are applicable to free recall as well as serial recall, if only at the hand-waving level.

Probed Recall

Another task is probed recall, as introduced in Chapter 1. In the case of item-probed successor recall, SEM, possessing no item-item associations, may have to appeal to covert serial recall. Nonetheless, this is what the data suggest (Chapter 1; Palmer & Ornstein, 1971; Sternberg, 1967). In the case of item-probed position recall, the probe item may be used to cue the positional code of SEM's corresponding token (i.e., the reverse process to that in serial recall). The case of position-probed item recall is less clear, because a position probe has no necessary relation to the internal positional codes in STM. With a numerical position probe for example, an additional translation process will be required to convert the probe into start and end marker values in SEM. With a spatial position probe, Chapter 1 described some evidence suggesting more direct access to internal positional codes. However, this appears true only when spatial and temporal positions are correlated (Hitch, 1974), suggesting that a "spatiotemporal probe" may be a better description.¹¹

Even with spatiotemporal probes however, direct access may be limited. Though latency data suggest that the first and last item of a list (Sanders & Willsemsen, 1978a) or group (Hendriks, 1984) can be accessed directly, the longer latencies for middle items suggest that they are accessed via serial search from the terminal positions. This evidence for serial search is not conclusive however, because SEM suggests an alternative reason. If response latency were related to cued strength, such that strengths had to increase above a threshold level before providing in direct access, then the lower peak strengths for middle items in SEM's positional uncertainty functions would predict the same latency profiles. Better evidence for serial search is the fact that the word-length effect, though diminished, is still

11. Interestingly, Hitch found an advantage when an item probe was combined with the spatiotemporal probe in successor recall. However, it was not clear whether this advantage was any greater than expected from the smaller number of possible responses resulting from the provision of one item as the probe.

found in spatiotemporal probed recall (Avons, Wright & Palmer, 1994). One possibility is that positional codes can be reinstated directly, but that it is often easier to reinstate codes for only the first and last positions directly, and reinstate the rest serially (perhaps explaining some of the individual differences found by Sanders and Willsemsen, 1978a). Alternatively, even spatiotemporal probes may not map simply enough onto internal positional codes to allow direct access. Thus the data on position-probed item recall suggest that direct access is sometimes possible, but are far from decisive.

In item recognition tasks (where the task is simply to state whether the probe item was somewhere in the list), latency data was originally taken to support serial search (Sternberg, 1969). More recent data however demonstrate a recency effect that is better explained by direct access (McElree & Doshier, 1993). In contrast to other probing techniques in SEM, the item recognition task could be achieved simply by checking the activation of the phonological representation of the probe item. This would produce direct access and a recency effect (e.g., Corballis, 1967), though the complete story may not be so simple (Monsell, 1978).

Finally, there is the question of whether item-probed position recall and position-probed item recall are symmetrical. Initial evidence suggested not (Jones, 1976), but more recently, symmetry was found when item information was controlled (Nairne, Whiteman & Woessner, 1995). SEM does not make explicit claims about symmetry, though symmetry between item and positional codes is consistent with its current formulation of tokens. More troublesome are data suggesting asymmetrical effects of phonological similarity (Hitch, 1972), but clear implications require a better understanding of how probe tasks are performed.

Backward Recall

The difficulty in reinstating positional codes in any order is supported by data on backward serial recall (e.g., Madigan, 1971). This task is normally harder than forward recall (Henson, 1995), though once item information is equated, the difference can disappear (Farrand & Jones, 1996). The latter authors argue that their data imply a single process underlying forward and backward recall, though others argue the opposite, with backward recall using spatial information (Li & Lewandowsky, 1993, 1995). These discrepancies may reflect strategic differences in the way people attempt backward recall, the most common

strategy depending on procedural details (e.g., whether recall is immediate or delayed, or whether there are intralist distractors, as in Li & Lewandowsky's experiments).

Clearer evidence on backward recall comes from latency measures. Longer latencies in immediate backward recall (Anders & Lillyquist, 1971) suggest that it may involve successive forward searches, reporting the last item after each search (Page & Norris, 1996b). This implies that positional codes in SEM can only be reinstated in a forward order, from the first through to the last. Again however, further data suggest some direct reinstatement of positional codes is possible. Error data in Henson (1995) suggest that people may be able to retrieve groups directly, even if they must retrieve items within those groups in a forward order. This is supported by closer inspection of latency data (Anders & Lillyquist, 1971) and is consistent with data from spatiotemporal probed recall (Hendriks, 1984). Thus evidence from backward recall, much like that from probed recall, suggests a combination of covert serial search and direct access via positional codes, which is not necessarily problematic for SEM.

Spatial Recall

So far, serial order has been restricted to the temporal dimension, where serial recall implies recall of temporal order (*temporal recall*). Serial order may also be defined along a spatial dimension. The question considered below is whether SEM could be extended to recall of spatial order (*spatial recall*).

Mandler and Anderson (1971) showed that a constant temporal order across four presentations of a sequence aided temporal but not spatial recall of the last presentation (where temporal and spatial order were uncorrelated). Constant spatial order on the other hand aided spatial but not temporal recall. They suggested therefore that the two dimensions are independent (in agreement with Hitch & Morton, 1975; Slamecka, 1967). Independence was further supported by superior temporal than spatial recall, and the fact that only temporal recall showed a recency effect.

An independence between temporal and spatial recall is not problematic for SEM. Spatial position might be encoded in tokens together with temporal position, and one or other cued independently. Furthermore, there is no reason why start and end markers could not be used to define spatial as well as temporal position (e.g., Nelson & Chaiklin, 1980). Spatial

position might be coded relative the left and right extremes of a linear sequence for example (though it is unclear why this does not produce a recency effect). Some suggestion of positional uncertainty associated with spatial positions was found by Hitch (1974) in spatial-probed recall, though it was not as clear as for temporal position. However, more recent research reveals the relation between spatial and temporal information to be far more complex.

Healy (1977) reported that spatial recall showed effects of temporal as well as spatial position. A similar result was reported for spatial-probed recall (Murdock, 1969). However, Healy failed to find the phonological confusions in spatial recall that typify temporal recall. This suggests a better distinction is between phonological and spatiotemporal coding: Only spatiotemporal coding applies to spatial recall, whereas both spatiotemporal and phonological coding apply to temporal recall. Phonological coding serves mainly to improve item recall (Healy, Cunningham, Gesi, Till & Bourne, 1991), as in SEM, though why it applies only to temporal recall remains unclear. Moreover, the nature of the spatiotemporal coding is also unclear. Healy (1982) showed that visual similarity of items had negligible effect on spatial recall, which could be achieved equally well with identical items. This suggests the underlying spatiotemporal representation is not a literal “movie”, but an abstract memory for a temporal series of locations. This is supported by similarities between temporal recall of verbal items and temporal recall of spatial locations (Jones, Farrand, Stuart & Morris, 1995; Smyth & Scholey, 1996). Nonetheless, the relation between spatial and temporal order clearly requires further research before models like SEM can be applied. For the moment, SEM is confined to temporal recall of items presented sequentially, in the absence of spatial information.

Running Span

In the running span task (Pollack, Johnson & Knaff, 1959), subjects are presented with a long list of items and have to recall as many of the most recent items in order as possible. Though lower than conventional spans, running memory spans are at least 3-4 items. *Prima facie*, this task would appear difficult to model in SEM, since the start and the end of the sequence are undefined. However, there is no reason why subjects cannot impose their own subjective starts and ends of subsequences, and use these to define position.¹² In other words,

12. Alternatively, subjects (and SEM) might use decaying phonological activations, reordering these on recall.

they may continually update the start of a group of items they intend to remember. Indeed, such spontaneous grouping is apparent (Pollack et al., 1959), and may explain why running memory span is greater when the total list length is known in advance. By assuming a variable, subjective start marker, this task is not necessarily problematic for SEM.

Other Tasks

In the case of perceptual matching of spatial sequences, performance for sequences differing by an adjacent transposition is worse than for those differing by a remote transposition (Ratcliff, 1981). Ratcliff fitted his accuracy and reaction time data by using positional uncertainty functions produced by the Perturbation Model (Lee & Estes, 1981). However, positional uncertainty in this model requires perturbations over time, and the same data may be equally well fitted using SEM's positional uncertainty functions, which do not require temporal perturbations (the positions may be anchored by spatial markers at the left and right of the sequence, as suggested above). A similar account may apply to position-specific priming, rather than assuming perfect initial coding of position and subsequent crosstalk (Peressotti & Grainger, 1995).

In the temporal domain, recognition is likewise poorer for sequences differing by an adjacent transposition than a remote transposition (Jahnke, Davis & Bower, 1989). These authors also fitted their data by assuming positional uncertainty functions, though the functions were taken from data on a second task of item-probed position recall, rather than being generated by a model. Nonetheless, these functions resembled those produced by SEM. Thus perceptual matching, priming and recognition of sequences all provide data consistent with the positional coding of SEM.

Long-term Learning

Finally, the most important question for SEM concerns long-term learning, or transfer from STM to LTM. For example, it is unclear how SEM would model the serial learning task introduced in Chapter 1. Given the episodic nature of SEM's storage, there is no incremental effect of learning the same sequence again and again. In the absence of rehearsal, a long-enough retention interval (i.e., enough contextual change) will cause complete forgetting of sequences. Nonetheless, there is evidence suggesting such forgetting is not atypical of STM,

and a secondary system is responsible for long-term learning (hence the distinction between temporary STM and permanent LTM in Chapter 1).

Examples of long-term learning in the serial recall task include the Hebb effect (Hebb, 1961). Hebb found that a list repeated every few trials showed improved recall with each repetition. However, the Hebb effect does not arise simply with repeated presentations, even with vocalisation (Cunningham, Healy & Williams, 1984). The effect is contingent on active rehearsal or recall (Kidd & Greenwald, 1988). A distinction between active and maintenance rehearsal seems necessary to explain why Healy, Fendrich, Cunningham and Till (1987) found an advantage of precuing over postcuing recall only when precuing before presentation; precuing at the start of a rehearsal interval between presentation and recall showed no advantage over postcuing immediately before recall. In other words, maintenance rehearsal alone does not improve recall (Brown, 1958). A lack of maintenance rehearsal explains why incidental learning reduces overall performance, but not the rate of forgetting (Cunningham, Healy, Till, Fendrich & Dimitry, 1993; c.f. Muter, 1980). Thus SEM's rehearsal process is appropriate for maintenance rehearsal, and a different process appears necessary for active rehearsal and long-term learning. Without active rehearsal, forgetting from STM is consistent with that predicted by SEM.

A further example of long-term learning in serial recall is the McNicol effect (McNicol, 1978). McNicol found a small but significant increase in recall of items that maintained the same position across successive trials, but not for two items that maintained only relative order. In general terms, this favours positional over chaining theory, suggesting some strengthening of position-item associations. However, recent replications (Page & Norris, 1996a) show the effect to be no greater than expected from the fact that protrusions can no longer be detected as errors. In other words, the McNicol effect could be no more than a scoring bias, in which case it is not incompatible with SEM, which has no strengthening of positional associations. Nonetheless, McNicol did find larger increases for items that maintained relative order over 10 or more trials, particularly with instructions for semantic elaboration. This may reflect the additional process of active rehearsal suggested above.

Unfortunately however, there is little evidence to discern the nature of active rehearsal

and long-term learning. It clearly involves the process of chunking subsequences of a repeated list (e.g., Bower & Winzenz, 1969; Martin, 1974). There may be a role for strengthening of position-item associations (Burgess & Hitch, 1996b), but, as a means of transfer to LTM, this could not overcome the interference problem as soon as several sequences of the same items are learned (Chapters 1, 8). Another possibility is that associations are learned to a different start and end marker for each sequence (Houghton, 1990). This requires numerous pairs of start and end markers available for the learning of new sequences. Alternatively, long-term learning may involve a different means of storing serial order. The extension of primacy-gradient ideas (Grossberg, 1978; Nigrin, 1993; Page, 1994) would appear to be a promising approach. Since the interest in serial learning has waned (Slamecka, 1985), further data are required to constrain models of this fundamental aspect of human cognition.

In summary, SEM requires considerable extension to model long-term learning, an issue related to the problem of serial order in LTM (Chapter 8). Nonetheless, the study of STM suggests that sequences are initially stored by positional codes, but that these codes soon become ineffective in the absence of maintenance rehearsal. Transfer to LTM may involve a secondary process of active rehearsal and chunking of these sequences.

Comparison with Other Theories

SEM can be briefly related to existing theories of short-term memory. The theories can be divided into general theories, and more specific models.

General Theories

SEM is a model of short-term memory, as defined in terms of temporary rather than permanent storage (Chapter 1). Such memory is assumed to span seconds to minutes, exceeding the classical extent of primary memory (Waugh & Norman, 1965). This is necessary to explain above-chance recall after several seconds of distraction; performance that one would not necessarily want to attribute to long-term (secondary) memory, in the sense of permanent storage. SEM's phonological activations are more akin to the notion of primary memory. As activation of LTM representations (Cowan, 1993), their transient nature explains the rapid forgetting over the first few seconds of retention (Peterson & Peterson, 1959).

Forgetting in SEM is both interference-based (Keppel & Underwood, 1962; Melton, 1963), in the retrieval of tokens, and decay-based (Brown, 1958; Baddeley & Scott, 1971; Conrad, 1967), in the retrieval of phonological forms. Decay occurs during storage, and both proactive and retroactive interference occur during retrieval, from competition between items encoded with similar positional and general contexts (i.e., an overload of start and end cues, Sanders, 1975). More generally, SEM is an example of theories that assume memory is related to contextual distinctiveness, with similar principles applying to both STM and LTM (Crowder, 1993; Neath, 1993a, 1993b). As a model specifically of STM however, it maintains the STM/LTM distinction of the modal model (Healy & McNamara, 1996).

In relation to organisation in STM, SEM's grouped structure is a matrix rather than hierarchy (Broadbent, 1981), in that one cue (e.g., position of item-in-group) applies to several items, and no item is dependent solely on one cue (Broadbent, Cooper & Broadbent, 1978). In other words, positions are coded along multiple dimensions and recall along one dimension does not require of recall along others, in contrast to hierarchical models (e.g., Johnson, 1972). When comparing hierarchical and matrix models of short-term memory, McNicol and Heathcote (1986) found that a hierarchical model fitted their data better when the items were familiar, such as digits, letters or musical notes, but a matrix model fitted their data better when items were unfamiliar, such as nonalphanumeric characters (e.g., \$, #, @). However, their matrix model assumed independence between each dimension (e.g., Lee & Estes, 1981). With the assumption of noise at each level of positional coding, SEM is a matrix model that does not assume complete independence (Fit 5). By including a nonindependent matrix model like SEM, McNicol and Heathcote may have been able to fit their data for all types of item (given that neither model they considered fitted particularly well).

Finally, SEM takes an interdependent stance on the relation between item and order information (e.g., Healy, 1974, 1982; Murdock, 1976). *Prima facie*, SEM's tokens store order information, via positional codes, while SEM's phonological activations store additional item information. This is not strictly true though, and the two types of information are not independent: Tokens also store item information, and phonological activation determines not only the number of item errors, but also the number of order errors (given the nature of the

competition process; Appendix 3). This contrasts with empirical measures of item and order information, which are often taken to imply independence. For example, Healy (1974) showed that serial position curves for four items were bowed when only order had to be remembered, but virtually flat when only the items had to be remembered. However, this may be an artifact of such a short lists; when item errors are plotted for longer lists, they are not flat, and can be bowed (Chapter 4). More importantly, an empirical dissociation does not imply independence (as apparent from the different patterns of item and order errors in Fit 2). Though item and order information may be useful concepts, their independence is not (Crowder, 1979).

Specific Models

SEM shares many assumptions with previous models. Of its three core assumptions, SEM's positional coding is based on the work of Houghton (1990). Houghton implemented start and end markers as nodes in a connectionist network, which was used to model sequential effects in speech production (though the network's inner-product metric of positional overlap lacks the important qualities of SEM's Euclidean metric; Equation 5-2). SEM's storage of separate tokens is based on the multiple-trace ideas of Hintzmann (1976), which appear better suited to explaining repetition effects in episodic memory (Chapters 7, 8). SEM's retrieval processes of additive noise, omission thresholds and response suppression are shared with the Primacy Model (Page & Norris, 1996b) and allow closer fits to detailed error patterns.

Of SEM's other assumptions, the coding of positions at multiple levels is based on the work of Lee & Estes (1977; 1981), though their notion of trial-level codes differs from SEM's notion of general context. Indeed, SEM's distinction between reinstatable (positional) and non-reinstatable (general) contexts is more akin to the ideas of Hintzmann, Block and Summers (1973). The assumption of maintenance rehearsal and phonological activations is based on the phonological loop (Baddeley, 1986), as described above.

However, SEM also differs from previous models in important ways. Firstly, in relation to positional coding, there are distinctiveness models of memory (e.g., Johnson, 1991; Murdock, 1960; Neath, 1993a; Neath & Crowder, 1996). The model of Johnson (1991), for example, assumes that serial position is represented along a single dimension, much like a physical property (e.g., loudness). Expressing magnitudes on this dimension in relation to

others allows a parameter-free estimation of positional overlap or distinctiveness. Though appealing, given the several parameters SEM uses to characterise positional codes, these models have only been used to produce general, qualitative results, such as primacy and recency. It is not clear that they can provide quantitative fits to data (particularly to detailed error patterns). More importantly, these models are descriptive models rather than process models. In other words, they only characterise long-run statistics of recall, and cannot produce an example recall protocol in the way SEM can.

The Perturbation Model (Lee & Estes, 1981) is better specified than distinctiveness models, and captures positional uncertainty with a single parameter, the perturbation rate. However, this perturbation only arises during storage: The model presumes that people can initially code position perfectly. More importantly, the unspecified codes can be extended arbitrarily, and provide no rationale for the limited resolution of positional coding (Fit 5). The Perturbation Model is also another descriptive model that does not fully simulate the recall process (Page & Norris, 1996b; see Nairne & Neath, 1994, and Mewhort, Popham & James, 1994, for a similar criticism of TODAM). For example, by assuming that items within a sequence perturb independently, the Perturbation Model predicts impossible situations where more than one item is supposedly stored at the same position. Moreover, by assuming that items perturb independently between sequences, it cannot explain the small dependencies found in the data (Experiment 2; Nairne, 1991). Finally, its assumption that omissions arise when items perturb “out of the trial” (Lee & Estes, 1981) is incompatible with the present pattern of omissions and repetitions (Chapter 4).

The attribute model of Drewnowski (1980a) extends to several aspects of short-term memory, including effects of list length, familiarity and phonological similarity. In this model, several attributes of items are coded, such as identity, position, auditory features and interitem relations. During recall, these attributes are addressed in a predetermined order of priority. Though appealing however, these ideas have little justification or explanatory power. For example, effects of list length are a simple consequence of “item load” in memory. Moreover, its assumption of only four positional codes is incorrect (Chapter 3), its assumption of interitem associations in the auditory trace is doubtful (Chapter 2) and, most importantly, it

does not produce appropriate transposition gradients. A similar model is the feature model (Nairne, 1988, 1990), which addresses recency, modality and suffix effects. However, the feature model has no explicit representation of serial order and, like the attribute model, fails to meet the fundamental locality constraint on transpositions. These models, like the Perturbation Model, are better regarded as frameworks than as detailed models of serial recall.

The ability of TODAM (Lewandowsky & Murdock, 1989) and its various extensions (Murdock, 1993, 1995) to model serial recall from short-term memory has already been questioned in Chapters 2 and 4 (though this does not necessarily detract from its application to other aspects of memory). It is unable to fit much of the data fitted by SEM, and, as a chaining model, cannot explain positional errors. Schneider and Detweiler (1988) have a more complex, distributed model of short-term memory, but unfortunately it is couched at a level which makes its application to present data unclear.

The Primacy Model of immediate serial recall (Page & Norris, 1996b) is appealing in its simplicity, though it is yet to be extended to grouping and intertrial effects. Like SEM, it fits detailed error patterns such as transpositions and omissions (though not perfectly in either case; Fit 1 and Fit 2). Indeed, the separate stages of categorical and phonological retrieval in SEM (Fit 4) are based on the two stages of the Primacy Model, allowing the correct pattern of confusions. However, being an ordinal rather than positional model, the Primacy Model cannot produce interpositions (Fit 5) or protrusions (Fit 6). The model needs to be supplemented with additional positional information, such as that employed in SEM.

Alternatively, the Primacy Model may be combined with SEM, in relation to SEM's phonological activations. These activations are currently seen as comprising an unordered item store, rather than the Primacy Model's ordered store. Indeed, SEM's phonological activations comprise a recency-gradient, rather than the primacy-gradient of Page and Norris. Nonetheless, the presence of a recency gradient is not a core assumption of SEM, and is not essential to fit present data. If serial order is also stored in the phonological store, it might be fruitful to combine the primacy-gradient ideas of Page and Norris with the positional-coding ideas of SEM. This is an area for future work.

Of all current models however, SEM is most similar to the Articulatory Loop Model of

Burgess and Hitch (1992). This model and its revisions (Burgess & Hitch, 1996a, 1996b) give reasonable qualitative fits to error data, such as transpositions, omissions, and phonological substitutions. It can also provide a qualitative fit to positional errors such as interpositions and protrusions (Burgess & Hitch, 1996b), though not to the same level of detail as SEM. Unlike SEM, the Articulatory Loop Model is implemented as a neural network, though the advantages of this are debatable. By remaining computational, but not connectionist, SEM is able to ignore this level of complexity. This means that the core assumptions of SEM are more transparent (expressible as simple equations in Appendix 3), making predictions clearer and allowing the model to be more readily testable. A connectionist framework does not appear to contribute much at this level of cognition (for a similar argument, see Page & Norris, 1996b).

Nonetheless, there remains an important difference between SEM and the Articulatory Loop Model. This reflects the nature of the positional codes. The moving context window assumed by Burgess and Hitch (Chapter 1) codes absolute position (e.g., *first*, *second*, *third*, etc.), irrespective of list length. Indeed, the coding of absolute position would appear a property of any model that codes position via temporal oscillators (e.g., Brown, Preece & Hulme, 1996). The coding of absolute position also seems implicit in the Perturbation Model (Lee & Estes, 1981). SEM on the other hand codes position relative to both the start and the end of a sequence, a coding which is sensitive to list length. The difference between absolute and relative position is testable, allowing the models to be distinguished empirically. This is the purpose of the experiments in Chapter 6.

Chapter Summary

This chapter introduced a new, computational model of short-term memory for serial order, the Start-End Model (SEM). The core assumptions of SEM are: 1) the position of an item in a sequence is coded relative to the start and end of that sequence, 2) items are stored in memory as position-sensitive tokens, and 3) items are retrieved by reinstating the positional codes for a response, and letting tokens compete in parallel for output. Additional assumptions that the influence of the start marker is stronger and longer-lasting than that of the end marker, that items are temporarily suppressed after output, that response selection is supplemented by additional phonological information, and that not all context is reinstatable at recall, allows SEM to give excellent quantitative fits to the data from Experiments 1, 2 and 3. No other model can reproduce the complete pattern of errors in this data. Moreover, SEM is readily extendable to other phenomena, such as the effects of retention interval, list-length, word-length, articulation rate, presentation modality and a redundant suffix (Fits 9-12; Appendix 3).

Two main issues remain for SEM however. Firstly, the psychological correlates of SEM's start and end markers are unspecified. In particular, the question remains of how the influence of the end marker can extend backwards in time. If SEM is to be useful as a psychological model, it must specify experimental manipulations that affect the behaviour of its start and end markers. Secondly, SEM must go beyond fitting existing data, and make novel predictions to be tested. These issues are tackled in Chapters 6 and 7.