# *Short-term Memory for Serial Order*


Richard Nevill Astley Henson

St. John's College



Dissertation submitted for the degree of Doctor of Philosophy

University of Cambridge

November 1996

# *Declaration*

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration. No part of it has been submitted for any other degree or qualification.

Chapter 2 is based on the paper by Henson, R. N. A., Norris, D. G., Page, M. P. A. & Baddeley, A. D. (1996), "Unchained memory: error patterns rule out chaining models of immediate serial recall", in *The Quarterly Journal of Experimental Psychology, 49A*, 80-115.

The experiments in Chapter 7 were presented at the Experimental Psychology Society Bristol Meeting, March, 1996, and, together with the model in Chapter 5, at the Second International Conference on Memory in Padua, July, 1996. The experiments in Chapter 7 are also included in the manuscript submitted for publication by Henson, R. N. A. (1996), "Item repetition in short-term serial recall: Ranschburg repeated".

# *Acknowledgements*

The work described in this thesis was supported by a Studentship from the Medical Research Council and generous grants from St. John's College, Cambridge. The work was conducted at the MRC Applied Psychology Unit, which brings together many special people to whom I owe thanks, particularly the other postgraduates, the computer support team, the library staff and the statisticians Ian Nimmo-Smith and Peter Watson. I was particularly fortunate to have three PhD supervisors at the Unit: Alan Baddeley, Dennis Norris and Mike Page. I would like to thank Alan for his remarkable experimental insight and Dennis for his emphasis on the value of clear writing. Most of all, I would like to thank Mike for his invaluable help and friendship. Many ideas in this thesis developed from memorable discussions with Mike during our regular lunch-time walks.

I am also fortunate to have many supportive friends in Cambridge; too many to name here, but I hope they know who they are. One friend in particular has always been there for me, Cheryl Smythe, to whom my gratitude comes with the deepest love.

This thesis is dedicated to my parents, Ken and Sue Henson, for their unfailing support and continued interest throughout my numerous years of study, particularly the last three.

# *Abstract*

How do we remember the order of a novel sequence of items, such as the digits in a telephone number? This problem is addressed by eight experiments on serial recall of temporal sequences. These experiments are used to develop a new model of short-term memory for serial order, the Start-End Model (SEM).

Existing approaches to the problem of serial order include chaining, positional and ordinal theories. Chaining theories, which store order in a chain of interitem associations, face problems explaining the pattern of errors in serial recall (Experiment 1). Positional theories, which store order by associations between items and their positions, are consistent with these errors, particularly those between sequences that maintain their position within a sequence (Experiments 2 and 3). Such positional errors cannot be explained by ordinal theories.

Previous positional models cannot explain the detailed pattern of errors however, as revealed by a meta-analysis of serial recall experiments. These errors are summarised as a set of empirical constraints, which are used to develop SEM, a computational model that simulates serial recall. SEM assumes (a) position is coded relative to the start and end of a sequence, (b) these codes are stored together with items as position-sensitive tokens, and (c) items are retrieved in order by cuing with codes for each position. SEM produces excellent quantitative fits to data from Experiments 1, 2 and 3.

SEM predicts that errors between sequences of different length will maintain position relative to the end of those sequences, in contrast to other positional models, which predict that such errors will maintain absolute position. SEM's predictions are confirmed in Experiments 4 and 5. SEM also predicts effects of repeated items in serial recall, which are examined in Experiments 6, 7 and 8. These complex effects of repetition pose important challenges for models of serial recall.

SEM is a model of short-term memory. However, the associated theory of relative position extends beyond short-term memory, and is a promising approach to the problem of serial order in general, particularly with respect temporal order in episodic memory.

# *Table of Contents*

# *List of Tables*

# *List of Figures*

# Chapter 1: Introduction

## *The Problem of Serial Order*

How do we store and retrieve a sequence of items in the correct order? For example, how do we recall a six-digit telephone number with apparent indifference to the other 719 permutations of the digits? Moreover, how do we recall that number with apparent indifference to the numerous permutations that correspond to other telephone numbers, bank codes, etc? This is the *Problem of Serial Order* (Lashley, 1951). The present thesis attempts an answer to this problem, at least with respect to short-term memory for temporal order.

More generally, all our actions are sequenced in time. Consider the articulation of a single word: This involves the ordered production of a number of phonemes. Given a typical production vocabulary over 50,000 words in English, this is a huge number of sequences drawn from a set of only 45 or so phonemes. Yet we have little problem articulating words rapidly and correctly. For example, it is extremely rare for the four phonemes in “style” to be misordered to produce articulation of “slight” instead. How is this huge capacity stored in memory, virtually free from interference?

We can normally remember the order of major events in our lives without recourse to external aids such as diaries. Here we are more prone to error, forgetting whether something happened yesterday or the day before, but there is certainly the impression of an autobiographical continuum, along which events recede hazily into the past. This continuum may be punctuated by factual knowledge, such as dates of important occurrences, but these aside, how are events ordered on such a continuum?

These questions illustrate the fundamental nature of the problem of serial order; an unresolved problem that cuts across many traditional psychological distinctions of memory. Clearly, phone numbers heard once may be forgotten completely an hour later, whereas others, such as those of our first home, may stay with us forever. This difference is often thought to reflect temporary storage in short-term memory versus more permanent storage in long-term memory (e.g., Baddeley, 1986). Spelling a word aloud may involve declarative memory, whereas the sequence of motor commands used to write the word involves procedural memory (e.g., Squire, 1994). Finally, episodic memory for the order of past events can be distinguished from semantic memory for facts (e.g., Tulving, 1983), such as the order of British Monarchs for example. There is no a priori reason why such different memories should use the same means of representing serial order; there is similarly no reason why they should not.

This thesis is concerned with the problem of serial order in short-term memory. More specifically, the domain is short-term, episodic memory for the temporal order of verbal material, a memory tapped by the task of serial recall. With respect to the first example, this is the memory that allows one to retain an unfamiliar telephone number long enough to dial it shortly afterwards, even though the same number might be forgotten minutes later. The advantage of this domain is that many variables that may affect memory for serial order can be readily manipulated and controlled in the laboratory. Though the answers that emerge from this restricted domain may not generalise to all of the examples above, they represent a good starting point. The extent to which the answers do generalise is discussed in Chapter 8.

## Theories of Serial Order

There are three basic theories of serial order: chaining theory, positional theory and ordinal theory. Each theory is introduced below, in the general terms of ordering a sequence of elements, where those elements might be digits in a telephone number, movements in an complex action, or events in autobiographical memory.

### Chaining Theory

This theory assumes order is stored by the formation or strengthening of associations between successive elements. The order is retrieved by stepping along these associations in a process called *chaining*, where each element cues the recall of its successor.

Chaining is probably the oldest approach to serial order (Ebbinghaus, 1964) and certainly the most intuitive. It is a simple extension of stimulus-response theory, where each response can become the stimulus for the next (Lashley, 1951). In its various guises, it has remained popular in several different models (e.g., Elman, 1990; Jones, Beaman & Macken, 1996; Jordan, 1986; Lewandowsky & Murdock, 1989; Richman & Simon; 1994; Wickelgren, 1965b). However, chaining theory faces several problems, as discussed below.

The simplest chaining models assume only pairwise associations between adjacent elements of a sequence (e.g., Wickelgren, 1965b), and a cue which consists entirely of the immediately preceding response (upper illustration in Figure 1-1). There are several immediate objections to such *simple chaining models*. For example, how do they handle sequences with repeated elements, in which two or more different elements will share the same cue? Or how do they allow recovery from error, because once an error has been made, the cue for all subsequent responses will be incorrect? This should lead to a cascade of further errors ("a chain is only as strong as its weakest link").

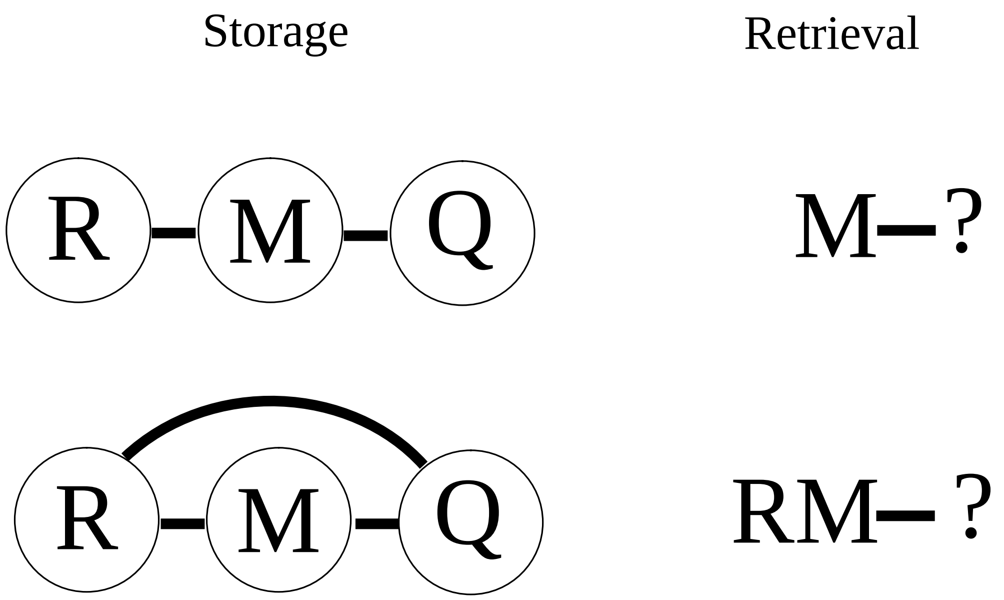



Figure 1-1: Simple and compound chaining models
(upper and lower illustrations respectively).

These questions are less problematic for models that assume remote associations as well as adjacent ones (e.g., Ebbinghaus, 1964; Slamecka, 1985). In these *compound chaining models* (lower illustration in Figure 1-1), the cue consists of a number of preceding elements, an approach which is popular in recurrent neural networks (e.g., Elman, 1990; Jordan, 1986). These compound cues allow disambiguation of elements repeated in a sequence, by virtue of the additional context of elements preceding the repeated elements. They also mean that a single error is less devastating, the additional context allowing recovery from that error.

There are other possible solutions. With respect to the problem of repeated elements, one can appeal to the type/token distinction, so that two occurrences of the same type have nonidentical token representations. For example, representations of each element may be embedded in different temporal or spatial contexts, allowing the same element to function as a different cue at different positions in a sequence, as in the case of Wickelgren's "allophones" (Wickelgren, 1969). This distinction is discussed in Chapter 7.

With respect to the problem of errors in recall, the TODAM model (Murdock, 1983) assumes only pairwise associations[1], but allows recovery from errors by only cuing with the previous response if it is correct. Otherwise, a cue approximating the correct one is used (Lewandowsky & Murdock, 1989). Though this may be appropriate when feedback of the correct response is provided, it is inappropriate for most situations (such as serial recall), where one does not always know whether each response is correct. Nevertheless, this approach illustrates an important distinction in chaining theory: whether the cue consists of the preceding elements recalled, which may be erroneous, or whether the cue consists of the preceding elements stored, irrespective of whether or not they are recalled correctly. This distinguishes *closed-loop* chaining models, where responses are fed back as cues, from *open-loop* models, where there is no feedback and therefore no necessary detrimental effect of errors. This distinction is particularly important in Chapter 2.

Nevertheless, there are many arguments against the sufficiency of chaining theory as a general account of sequential behaviour (e.g., Johnson, 1972; Lashley, 1951). One argument concerns the interference predicted by most chaining models as soon as several sequences of the same elements are stored (at least for those that assume type representations). For example, how could the order of letters in the words *pat*, *apt* and *tap* be retrieved if each letter is associated with almost every other (Houghton & Hartley, 1996)? The scale of this *interference problem* is apparent in the above example of 50,000 words drawn from a set of 45 phonemes: The degeneracy of associations assumed by chaining theory would surely predict much greater interference in speech production.

Another argument concerns the feedback of responses in closed-loop models. Though

1. (though see Murdock, 1993, 1995 for extensions of TODAM that use more than pairwise associations)

"...the only strictly physiological theory that has been explicitly formulated... postulates chains of reflexes, in which performance of each element of the series provides excitation of the next" (Lashley, 1951, p. 114), Lashley argued that many actions are executed so fast that there is simply not enough time for feedback to cue the next response. The finger presses of a skilled typist, for example, are too fast for proprioceptive feedback (Rumelhart & Norman, 1982). This argument has had a profound effect on motor control theory (Bruce, 1994), leading to the idea of motor programs that are independent of feedback, and rejecting the idea of reflex chains (though clearly there is some role for feedback; MacKay, 1982).

Lashley's most persuasive arguments concern everyday observations that order involves more than the linear structure of chaining theory. Many sequences are ordered hierarchically, such as the order of words in a sentence, the order of syllables in each word, the order of phonemes in each syllable, etc. Moreover, the order of elements typically respect syntactic or schematic structure beyond interitem associations. For example, speech errors show that a noun is far more likely to swap with another noun than a verb, even if the noun is further away in the sentence (Levelt, 1989). Such structure is reflected at all levels of speech production. For example, the structure of a syllable imposes constraints on the order of its constituent phonemes that is clearly beyond chaining theory (Hartley & Houghton, 1996).

Finally, the question remains as to what cues the first element in a sequence, in order to "kick-start" the chaining process. Many chaining models appeal to a separate contextual cue (Murdock, 1995) or plan unit (Jordan, 1986). Thus, chaining theory is clearly not a sufficient answer to the problem of serial order; later chapters will question whether it is even necessary.

**Positional Theory**

This theory assumes order is stored by associating each element with its position in the sequence. The order is retrieved by using each position to cue its associated element. In other words, rather than using the item-item associations of chaining theory, positional theory uses position-item associations.

The simplest example of a positional theory is Conrad's "box" model of short-term memory (Conrad, 1965). Conrad assumed that people possess a number of boxes in memory, into which elements of a sequence can be placed (upper illustration in Figure 1-2). The order of elements can be retrieved by stepping through the boxes according to a predetermined routine. This model does not have a problem with repeated elements, because they are stored in separate boxes, nor with recovery from errors, because the retrieval mechanism can continue to the next box irrespective of whether the contents of the previous box were correct (i.e., no feedback is required). This is the method by which conventional (Von Neumann) computers store and retrieve order, through routines accessing separate addresses in memory.

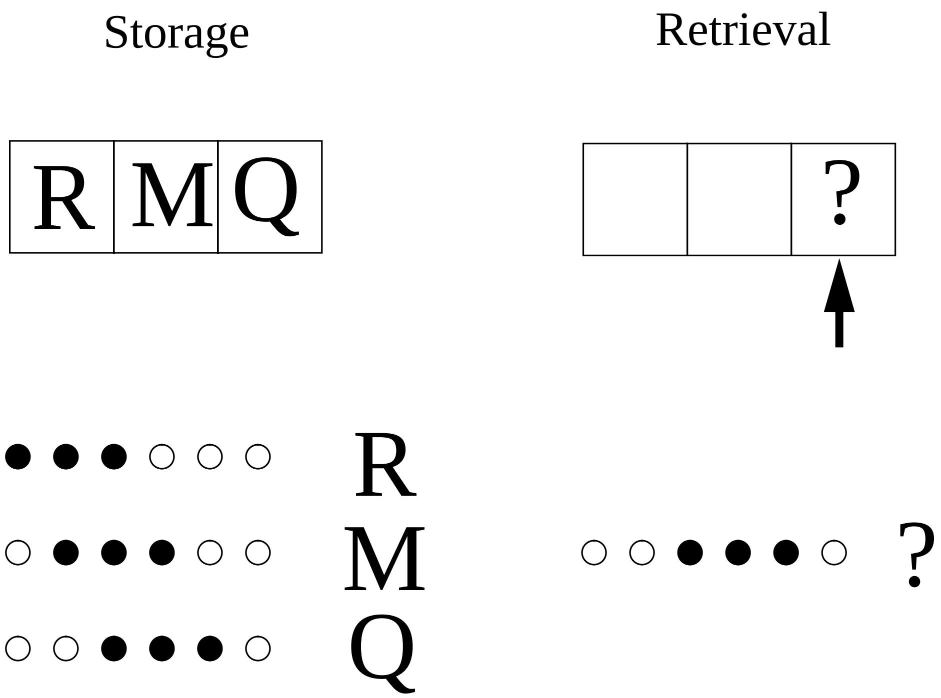



Figure 1-2: Positional models of Conrad (1965) and Burgess and Hitch (1992) (upper and lower illustrations respectively).

As a psychological model however, such a simple model is inadequate. Firstly, how many boxes do people have in short-term memory: five, six, seven, or more? If a new box is created for each element in a sequence, there would be no limit to the length of sequences people could hold in short-term memory, which is clearly not the case. Secondly, the model provides no immediate rationale for the errors people make when they misremember sequences: People are more likely to confuse elements close together in a sequence than far apart (Chapter 2). There is no reason for this with the perfect coding and retrieval of positions assumed by Conrad (1965).

One way to explain the above errors is that positional codes become confused over time. For example, the Perturbation Model (Lee & Estes, 1977; 1981) assumes the positions of elements are initially coded perfectly, but can get perturbed during storage such that nearby elements swap. Another way to explain such errors is that positional codes are not perfect, but overlap, in the sense that the code for one position is similar to the codes for nearby positions (lower illustration in Figure 1-2). This is the approach taken by the Articulatory Loop Model of Burgess and Hitch (1992). In this model, the circles in Figure 1-2 represent nodes in a connectionist network. The filled nodes are active nodes; the unfilled nodes are inactive nodes. The "window" of active nodes moves from left to right for each position in a sequence, and is associated with other nodes (not shown) representing each element. However, because there is some overlap in the set of active nodes for nearby positions, elements at these positions can sometimes be confused during retrieval.

Positional theory can be extended to a hierarchy of positional codes (e.g., Lee & Estes, 1981). For example, a phoneme can be coded for its position in a syllable, a syllable can be coded for its position in a word, a word can be coded for its position in a sentence, etc. As a general solution to the problem of serial order however, the status of positional theory remains unclear. There is a sense in which the problem is not solved, but circumvented. This sense reflects the question of how the order of the positional codes, rather than the order of the elements, is stored and retrieved from memory. This question cannot be answered without specifying the nature of the positional codes. One suggestion is that the codes are successive states of internal oscillators in the brain (Brown, Preece & Hulme, 1996; Burgess & Hitch, 1996a, 1996b). In these models, elements are associated with different states of the oscillators, and these states can be reconstructed simply by resetting the oscillators and letting them change under their own dynamics. In other words, the oscillators represent a biological clock, which can be rewound in order to retrieve a sequence from memory.

Positional theory does not solve the interference problem however. If a positional model is to store and retrieve the order of letters in *pat*, *apt* and *tap*, then it must employ different positional codes for each word (otherwise each letter will be associated with several positions). If these positional codes were states of internal oscillators for example, a different set of oscillators might be required for every word we know. Furthermore, the syntactic constraints on serial order, such as those in speech production, involve more than simply positional information (Chapter 8).

Nonetheless, this thesis provides good evidence that people do use positional codes in short-term memory for serial order, and describes a new model that attempts better specification of the codes. The question of whether such a positional model is sufficient for long-term memory for serial order is resumed in Chapter 8.

### Ordinal Theory

This theory assumes order is stored along a single dimension, where that order is defined by relative rather than absolute values on that dimension. Order can be retrieved by moving along the dimension in one or other direction. This theory need not assume either the item-item nor position-item associations of the previous theories.

For example, Grossberg (1978) assumed that order is stored in a primacy gradient of strengths in memory, such that each element is stronger than its successor. The order of elements is retrieved by selecting the strongest element, suppressing it, selecting the next strongest, suppressing it, etc. (Figure 1-3; suppression indicated by the broken lines). This idea has been incorporated into the Primacy Model of short-term memory (Page & Norris, 1996b), where the strengths might represent the degree of association to the start-of-sequence context, or even simply activations of item representations in memory.

The original Perturbation Model (Estes, 1972) is also an ordinal theory, where order is inherent in the cyclic reactivation of elements. Perturbations in the timings of reactivations lead to erroneous reorderings of the elements, like shifts in the relative phases of a series of oscillations.[2] Yet another ordinal model is that of Shiffrin and Cook (1978). This model assumes associations between each element and a “node”, but only the nodes are associated with one another (i.e., unlike chaining models, where it is the elements themselves that are associated with each other). By moving inwards from nodes representing the start and end of the sequence, the associations between nodes allow the order of items to be reconstructed.

2. Later developments of this model (Lee & Estes, 1977, 1981) assumed the perturbations were of positional attributes of items, rather than the timing of the items themselves, making it a positional model.

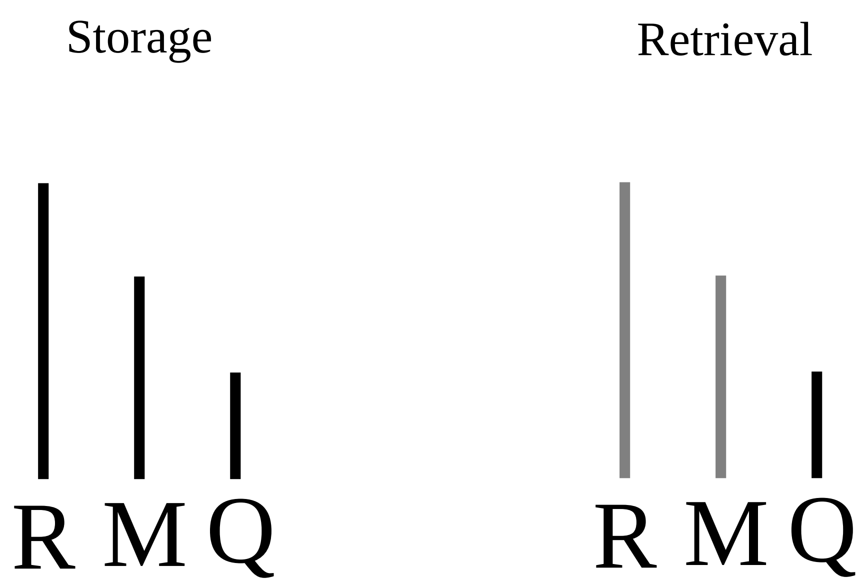



Figure 1-3: An ordinal model (e.g., Page and Norris, 1996b).

Ordinal models like the Primacy Model require token representations in order to handle repeated elements: The order of repeated elements could not be represented over type representations with a single strength. As regards errors in recall, ordinal models imply that errors will cooccur, in the sense that one error will cause another (because order is defined relationally). For example, if an element becomes stronger than its predecessor in the Primacy Model (owing to random noise), then the elements will transpose, causing two errors. This is an attractive property, because such transpositions are common in people too (Chapter 2).

Ordinal models do not require feedback of responses, and a process like suppression can operate independently of errors occurring at later stages of output. Some ordinal models have also been extended to hierarchical structuring in a manner that allows them to overcome the interference problem (Nigrin, 1993; Page, 1994). This is achieved by assigning a new node to each sequence, and associating elements to that node with a strength proportional to their relative order (via a primacy gradient). Because each sequence is associated with a unique node, the order of different sequences can be stored without interference (Chapter 8). This is a simpler solution than assuming separate positional codes for each sequence.

Thus ordinal theory escapes some of the criticisms of chaining and positional theories. Nonetheless, the present thesis will argue that ordinal theory is insufficient as an account of people's short-term memory for serial order.

## Theoretical Differences

In spite of the various strengths and weaknesses of specific models discussed above, important differences remain between the three theories of serial order. The difference between chaining and positional theories is obvious: the retrieval cue in the former is the previous element; the retrieval cue in the latter is some (abstract) positional code. The difference between positional and ordinal theories is less obvious, but relates to whether the position of an element in a sequence can be defined independently of its surrounding elements. In positional theories, it can; in ordinal (and chaining) theories, it can not. The consequence is that, in ordinal models, the middle element in a sequence can only be retrieved after retrieval of its predecessors (or successors, in the model of Shiffrin & Cook, 1978). In positional models however, it is possible to retrieve the middle element directly, by reinstating the appropriate positional cue. This is crucial in explaining a class of positional errors found in short-term, serial recall (Chapter 3). More specific differences between these theories are detailed in Chapters 2, 3 and 4.

There are, of course, means of storing serial order that do not fit easily within these theories. For example, order may be stored propositionally, in the form of statements such as “R was before M” and “Q was after M” (though these might be regarded as a form of interitem association). While appropriate perhaps for semantic memory (e.g., the order of British Monarchs), this approach is less appropriate for short-term memory for serial order. Remembering that “R was before M” may supplement short-term memory, but such memories do not appear necessary, and no explicit models have been developed along these lines. Subsequent chapters are therefore restricted to the three theories outlined above.

## Empirical Differences

The three theories of serial order can be distinguished empirically, though not always easily. Most of the relevant experiments have been performed within the domain of short-term memory (see Chapter 8 for evidence from other domains). In these experiments, the sequences are typically novel lists of verbal items to be recalled shortly after their presentation. These experiments fall into three main types: serial learning, probed recall and serial recall.

### Serial Learning

Chaining and positional theories were first pitted against one another within the serial learning paradigm, under the guises of item-item versus position-item associative theories (e.g., Young, 1968). The serial learning paradigm uses lists of 10 or more items, beyond normal memory span (i.e., the lists are rarely recalled correctly on the first attempt). Subjects are given repeated presentation and recall trials until they learn a list to criterion. The manner in which they have learned the order of items is normally investigated by the amount of transfer to a second task. For example, list learning might be followed by paired associate learning of items that were either adjacent or nonadjacent in the list. If a list *RMQJHV* were learned by forming associations between successive items, as predicted by chaining theory, positive transfer might be expected in paired associate learning of *RM* and *QJ*, but not *HR* or *VM* (e.g., Young, 1962). Alternatively, if serial learning of the list *RMQJHV* were followed by serial learning of a second list *RJQVHM*, or *JHVRMQ* (e.g., Ebenholtz, 1963; Keppel & Saufley, 1964; Slamecka, 1964), then positional theory, but not chaining theory, might predict positive transfer in the former case (because three of the items retain their positions from the original list), while chaining theory, but not positional theory, might predict positive transfer in the latter case (because most of the items retain the same predecessor from the original list).

Unfortunately, conclusions from the serial learning paradigm were mixed (Ebenholtz, 1972; Jensen & Rohwer, 1965; Young, 1968). Indeed, they led some researchers to propose that both item-item and position-item associations are utilised in serial learning (Battig, Brown & Schild, 1967; Houston, 1976). Alternatively, the mixed results may reflect a problem with the paradigm, in that the two theories are hard to dissociate fully. For example, in the paired associate task, any transfer for the associates *RM* and *QJ* could equally well reflect mediation of similar positional codes (given that they come from nearby positions), rather than interitem associations. Positive transfer in such cases does not rule out the positional theory therefore. In fact, given the proximity required for interitem associations, the chaining and positional accounts will always be confounded in these situations. Another problem is that transfer between the two tasks presumes that the associations underlying the tasks are identical, and yet there may be quite separate mechanisms underlying serial and paired associate learning.

The second example of transfer between two serial learning tasks is not without problems either. One problem is that, without a more detailed model of how serial learning is achieved (via item-item or position-item associations), the exact predictions of the two theories remains unclear. For example, if new item positions impaired learning at the same time as old positions improved learning, then a lack of transfer from *RMQJHV* to *RJQVHM* could reflect a balance between positive and negative transfer for individual items. Similarly, a lack of transfer from *RMQJHV* to *JHVRMQ* could reflect a large negative effect of unlearning old interitem associations before the learning of new ones is possible, particularly if those associations are remote as well as adjacent. Another problem is that the results from this transfer task are sensitive to the exact method of serial learning, such as whether subjects are aware of the relationship between the original and derived list (Maisto & Ward, 1973).

In summary, the use of transfer tasks has several methodological problems, and the serial learning paradigm has not proved a fruitful means of testing theories of serial order.

**Probed Recall**

Another means of testing theories of serial order is with probed recall (e.g., Murdock, 1968). For example, given a list followed by a probe item from that list, chaining theory predicts it to be a simple matter to recall the item that followed the probe in the list (via the association between them). If, on the other hand, the probe were a number corresponding to a position in the list, positional theory might predict it to be a simple matter to retrieve the item at that position. Ordinal theory predicts both tasks to be somewhat harder.

The main problem with the probed recall paradigm however is that there are many ways to perform the task. For example, in the first task above of *item-probed successor recall* (e.g., Waugh & Norman, 1965), performance might not be based on a direct item-item association, but rather an indirect series of item-position-item associations, by first retrieving the position of the probe and then cuing with the next position. The fact that people can perform the task of *item-probed position recall* (e.g., Jahnke, Davis & Bower, 1989; McNicol, 1975) is consistent with this hypothesis. The basis of performance on the second task above of *position-probed recall* (e.g., Nairne, Whiteman & Woessner, 1995) is uncertain because there may not be any simple transformation of the probe into a positional code (i.e., no direct mapping between numbers and internal positional codes).

The uncertainty in the processes underlying probed recall is reinforced by examining response latencies. Sternberg (1967) showed that the latency in item-probed successor recall increases linearly with position of the probe in the list. This suggests serial search from the start of the list, until the probe is encountered. In other words, performance on this probe task is probably based on covert serial recall of the whole list. This is exactly what the ordinal theory would have to predict. Nonetheless, performance is improved if the probe item is accompanied by a spatial, positional probe (Hitch, 1972), which attenuates the effect of probe position on latency (Monsell, 1973). This suggests some role for positional information, consistent with data from experiments using a spatial probe alone. Sanders and Willemsen (1978a), for example, showed that response latency with a spatial probe is a nonmonotonic function of position, with a recency advantage in addition to the primacy advantage of Sternberg's task. A spatial probe may therefore ameliorate the need for serial search, at least for terminal items, supporting positional theory. The problem with this approach however is that it remains unclear how a spatial probe interacts with memory for temporal order, and whether the results would hold in the absence of spatial information (Chapter 5).

In summary, the probed recall task has also faced problems distinguishing theories of serial order, particularly for ordinal and positional theories, mainly owing to uncertainty in the processes underlying performance of the task.

### Serial Recall

A better way to test theories of serial order is with serial recall. This task is simply to recall the whole list in a forward order, from the first to the last item. It proves particularly fruitful to study the pattern of errors when recall fails.

One example of errors in serial recall are *associative intrusions* (Wickelgren, 1966). These are transpositions between the items immediately following repeated items, and are more common than corresponding transpositions following nonrepeated items. Wickelgren used such errors to support chaining theory: Repeated items are ambiguous cues for the items that follow them because they are associated with more than one such item. These errors are reexamined in Chapter 7, though it is argued that they do not, in fact, constitute evidence for chaining theory. Another example are *serial order intrusions* (Conrad, 1960). These are errors of items that occurred at the same position in the previous trial, and are significantly more common than predicted by chance. Such errors clearly support positional theory, and are examined in more detail in Chapter 3.

Thus, as Estes (1972) observed: “When retention is imperfect, the confusion errors that occur are highly systematic” (p. 161). Though a single error may reflect a temporary failure to realise an accurate representation in memory, large numbers of errors show striking patterns in their distribution. These patterns shed light on the mechanisms subserving serial recall and hence the underlying representations of serial order over which the mechanisms operate (Conrad, 1959). Errors in serial recall are seldom random guesses.

There are several close cousins to serial recall, such as backwards recall (where items are recalled in reverse order), or positional recall (where items must be placed in the correct positions, but the order of recall is unconstrained), which may also shed light on theories of serial order. However, serial recall remains the most important task, given that it underlies the basic index of short-term memory, *memory span*, and many of the relevant empirical dissociations in short-term memory research (Baddeley, 1986).

Serial recall is also an everyday cognitive activity (e.g., recalling a telephone number). Indeed, a forward order is the default (and optimal) recall order from short-term memory, at least for up to six or seven items. Serial recall is therefore less likely to be contaminated by specialised strategies than are other unusual, and perhaps artificial, laboratory tasks, like probed recall. Moreover, because people are trying their best not to make errors (as opposed to trying their best to make use of a probe), error analysis is more likely to reveal the underlying representation of serial order, particularly given that people are often unaware of their errors (Chapter 6). Not surprisingly therefore, experiments in this thesis are confined to serial recall.

## Measuring Serial Recall

The earliest measurement of serial recall was the proportion of lists recalled correctly (e.g., Crannell & Parrish, 1957). This measure underlies the span index of short-term memory: One’s memory span is usually defined as the length of list that one can recall correctly 50% of the time. However, this measure ignores differences in the recall of each item in a list. Murdock (1968) pioneered the use of *serial position curves*, which plot the proportion of items recalled correctly at each position of a list. These curves are bowed, with an advantage in recall for the first and last few items. The advantage for early items is termed *primacy*, and the advantage for later items is termed *recency*. In immediate serial recall, primacy is normally more pronounced than recency (Chapter 4).

Later work distinguished two main types of error: *order errors* and *item errors* (e.g., Estes, 1972). Order errors are list items recalled in the wrong position; item errors are list items not recalled anywhere in the report. Estes analysed order errors further by comparing an item's position in the list with its position of recall. This analysis showed that erroneous items are clustered around their correct position, rather than being randomly distributed.

However, few studies actually go beyond measuring the proportion of lists correct or plotting serial position curves, let alone analysing the distribution of order errors. This thesis attempts a more comprehensive classification of errors. Indeed, a major theme behind the thesis is that a great deal of information is available through analysing error patterns in more detail. This may be why Conrad is reputed to have said that error analysis is "the royal road to memory" (A. D. Baddeley, personal communication, 1995).

### Classification of Errors

The classification of errors used in the present thesis is described below, with examples given in Table 1-1. This classification distinguishes an item's position in a list, its *input position*, from its position in a subject's report of that list, its *output position*.

When scoring by output position (i.e., taking each response in a subject's report), errors can be broadly categorised into *omissions* and *substitutions*. Omissions arise when no item is given for a position; substitutions arise when an incorrect item is given. Substitutions may be either *transpositions* or *intrusions*. Transpositions are list items in the wrong position; intrusions are items that were not present in the list. Intrusions may be items outside the experimental vocabulary (the set of items from which all lists are constructed), but most often they are items appearing on previous trials. Those intrusions that come from the immediately preceding trial are called *immediate intrusions*.

| Error Type | List (Input Positions) | Report (Output Positions) |
|---|---|---|
| Omissions | R M **Q** J **H** **V** | R M **K** J **-** **-** |
| Transpositions | R M Q **J** **H** V | R M Q **H** **J** V |
| Intrusions | R M **Q** J **H** V | R M **F** J **Y** V |
| Confusions | R M Q **J** H **V** | R M Q **K** H **P** |
| Repetitions | **R** **M** Q **J** H **V** | R M Q **R** H **M** |
| Associates | R M **Q** **J** **H** **V** | R M **J** **H** **V** **Q** |
| Interpositions | R M **Q** J H **V** | R M **V** J H **Q** |
| Protrusions | F P **Y** K **Z** W | F P Y K Z W |
| | R M **Q** **J** **H** V | R M **Y** **P** **Z** V |

Table 1-1: Example errors in serial recall.
(Errors are in bold; items corresponding to a particular error type are underlined.)

A special class of substitutions are phonological *confusions*. These are incorrect items that are phonologically similar to the correct item, and are common in tests of immediate memory. A special class of transpositions are *repetitions*. Repetitions are items that occur more than once in a report, even though they only occurred once in the list. The distribution of repetitions sheds light on the retrieval processes underlying serial recall (Chapter 4). Another special class of transpositions are *associates*. These are items recalled in the correct order relative to the previous item recalled, albeit in the wrong position (e.g., Wickelgren's associative intrusions). These errors are predicted by chaining models (Chapter 2).

Two further types of *positional errors* can be identified. *Interpositions* arise when lists are split into groups (e.g., by a pause between presentation of every third item). They are transpositions between groups that maintain their position within a group. *Protrusions* are similar errors, but maintain position between trials rather than between groups (i.e., Conrad's serial order intrusions). Note that the definition of protrusions is orthogonal to that of transpositions and intrusions, in that a protrusion may be either an intrusion or a transposition with respect to the current trial (which is why the term is preferable to Conrad's). Positional errors are predicted by positional models (Chapter 3).

Unless stated otherwise, errors are classified by output position. However, additional information is provided by scoring against input position (under which categorisation of errors is similar, and fairly self-evident; Table 1-1). For example, when omissions are scored against input position, they represent items that are not recalled anywhere in a report. This distribution of omissions can differ from that plotted against output position (Chapter 4). The distinction between item and order errors also generally refers to input position (e.g., Healy, 1974), though this distinction is not used often in the present thesis.

This concludes the majority of errors distinguished in this thesis. Though the complete classification might appear somewhat complex (certainly more comprehensive than conventionally attempted), each type of error plays an important role in constraining models of serial recall, as subsequent chapters will demonstrate.

### Additional Terminology

In the present thesis, *error position curves* are plotted instead of conventional serial position curves. These show the percentage of responses that are errors at each position. When plotting any error, these curves correspond to an inversion of serial position curves about the 50% performance line (and the distinction between input and output position is irrelevant). However, errors of different types can also be plotted separately, against either input or output position, so providing more information than conventional serial position curves.

Transpositions can be categorised with respect to both input and output position. This information can be represented in matrix form, where the entry in the *ith* row and *jth* column represents the number of items in Position *i* of a report that came from Position *j* in the list. The serial position curve corresponds to the main diagonal of this matrix, while the rows can be plotted to give *transposition gradients*. These gradients are generally peaked, with the peaks representing correct items ($i=j$) and the number of transpositions decreasing with increasing transposition distance $|i-j|$ (e.g., Figure 2-2 in Chapter 2). Immediate intrusions can also be classified in this way, where the *jth* column represents the position in the previous trial (either the previous list or the previous report). Such *intrusion gradients* show a similar, though flatter, pattern to transposition gradients (e.g., Figure 3-6 in Chapter 3).

Another error of interest is the *first error* to occur in a report. This allows calculation of the conditional probability of an error, given that previous responses are correct (Henson, Norris, Page & Baddeley, 1996). Using survival analysis (Appendix 1), this measure reveals more subtle changes in error probabilities across positions, such as the effects of group boundaries (Chapter 3).

Finally, the terms *Short-term Memory* (STM) and *Long-term Memory* (LTM) used in the present thesis may have different meanings in other contexts. STM and LTM are not meant as theoretical constructs, such as the *Primary* and *Secondary* memories of Waugh and Norman (1965). Though STM (or *Working Memory*, Baddeley, 1986) may correspond to a distinct memory system (Schacter & Tulving, 1994), the only distinction used here is that information in STM is temporary, being forgotten after a matter of minutes, while information in LTM is permanent. The relationship of STM to other aspects of memory is discussed in Chapter 8.

## Experimental Design

The eight experiments described in this thesis employed serial recall of between five and nine items. The items were familiar, so that only memory for their order was required, which changed every trial. The items were presented sequentially, in the middle of a computer screen, each one replacing its predecessor. Thus serial order was only defined temporally; there was no spatial information. Presentation rates were between one and two items per second. Subjects were instructed to guess if unsure, but could omit if no item came to mind.

Apart from the main manipulations of interest, the experiments differed in whether the items were digits (Experiment 2), letters (Experiments 1, 6, 7, 8) or words (Experiments 3, 4, 5). Some experiments required vocalisation of the items as they appeared (Experiments 3, 6, 7, 8); the others required items to be read in silence (Experiments 1, 2, 4, 5). Some required recall immediately after the last item had disappeared (Experiments 1, 2, 4, 5); the others delayed recall by three, vocalised distractor digits (Experiments 3, 6, 7, 8). Some required written recall (Experiments 1, 2, 4, 5, 6, 8); the others required spoken recall (Experiments 3, 7). These procedural differences were not of primary importance and their implications are only mentioned in passing.

The main manipulations were the phonological similarity of items (Experiment 1), the length and temporal grouping of lists (Experiment 2), the length of the intertrial interval (Experiment 3), the relative size of different groups (Experiment 4), the relative length of different lists (Experiment 5), the presence of repeated items (Experiments 6, 7) and the effect of guessing instructions (Experiments 4, 5, 8). These manipulations arise from predictions of different theories of serial order and different models of serial recall.

## Statistical Tests

Many analyses in the present thesis concern proportions or probabilities (e.g., the proportion of responses that are errors). Such proportional scores are suspect to floor and ceiling effects when they are close to 0 or 1, sometimes producing skewed distributions. To make some allowance for this, an empirical, log-odds transform is used (Appendix 1). This transform is particularly useful for proportions of errors of a certain type (e.g., the proportion of errors that are transpositions), given that some subjects make more errors than others: The log-odds can be weighted, giving more weight to scores from subjects who make more errors. In other words, the mean proportion can be weighted by sample size.[3]

The means and standard deviations of proportions are given in tables (the latter in brackets). When the sample size is fixed across subjects, these statistics normally represent

3. This weighting is restricted to pairwise comparisons of log-odds in the present thesis. Weighted ANOVAs can be performed on log-odds, but they lack orthogonality, making interpretation of interactions ambiguous. ANOVAs in the present thesis were therefore restricted to unweighted log-odds.

untransformed data. When sample sizes vary across subjects however, the statistics are calculated by retransforming weighted log-odds, in order to give a truer indication of population statistics. This retransformation is always indicated in the table captions. (One caveat with this approach is that retransformed proportions will not always sum to exactly 1.) When summary statistics refer to approximate proportions, they are expressed as percentages.

Finally, a comparison between two mean proportions is deemed statistically significant for alpha levels below *.05*. When making multiple pairwise comparisons, significance levels are based on Holm's method for adjusting alpha, which is essentially an iterative application of the Bonferroni correction (Appendix 1).

## Computational Modelling

This thesis combines both experimental and modelling approaches. In particular, many of the analyses performed in Chapters 2, 3 and 4 were used to develop the computational model of short-term memory described in Chapter 5. This model makes predictions which were used to guide further experiments in Chapters 6 and 7. This interaction between empirical and computational approaches illustrates the important role of modelling in psychological research.

There are many advantages of computational modelling over more traditional verbal theorising (e.g., Hintzman, 1991). The main advantage is that computational models can be specified unambiguously. This reduces misinterpretation and makes their predictions clearer, improving their testability (indeed, provision of the relevant computer program allows anyone to validate a model's predictions). Computational models also allow direct quantitative fits to data, rather than the qualitative hand-waving made by many verbal theories (e.g., they can not only predict an interaction, but also the size of the interaction). In particular, the ability of computational models to simulate nonlinear, probabilistic processes means that complex interactions can arise from relatively simple mechanisms; interactions that are hard to predict a priori (or analytically, in the case of mathematical models). Given the sequential dependency between responses in serial recall (Henson et al., 1996), this ability proves particularly important in explaining the complex interactions between different error types (Chapter 5).

## Overview of Thesis

In Chapter 2, the predictions of chaining theory are tested in immediate serial recall of lists of phonologically confusable and nonconfusable items. The data from Experiment 1, together with those in Henson et al. (1996), provide no support for chaining theory.

In Chapter 3, the effects of list length, grouping (Experiment 2) and proactive interference (Experiment 3) are examined. The pattern of errors found between groups and between trials is explicable by a positional theory of serial order, but not an ordinal theory.

In Chapter 4, meta-analyses of error data are performed on a number of experiments performed recently at the Applied Psychology Unit. These analyses produce a set of empirical constraints that any model of short-term, serial recall must meet. No previous model can.

In Chapter 5, a new, computational model of serial recall is developed, the Start-End Model (SEM), which meets the empirical constraints of Chapter 4, and provides quantitative fits to the data from Experiments 1, 2 and 3. This model is an example of a positional theory. Extension of SEM to other phenomena in STM is discussed, as is its relationship to previous models. Most importantly, SEM predicts an new property of positional errors in serial recall.

In Chapter 6, Experiments 4 and 5 confirm the predictions of SEM, and pose a serious challenge to other positional models, particularly those that assume positional codes are generated by internal oscillators in memory.

In Chapter 7, Experiments 6, 7 and 8 examine the effects of repeated items in serial recall. The results, together with those in Henson (1996b), are consistent with the basic assumptions of the SEM, but suggest that several additional processes are involved in memory for repeated items.

In Chapter 8, the more general assumptions of a positional solution to the problem of serial order are discussed. It is concluded that, while not denying other representations of serial order, particularly in procedural memory, positional theory appears a promising approach to the problem of serial order, particularly in episodic memory.

# Chapter 2: Chaining Theory

## *An experiment testing Chaining Models*

This chapter describes an experiment involving recall of lists of alternating confusable and nonconfusable items. The results, together with those in Henson et al. (1996), are troublesome for chaining models of serial recall. The chapter also includes a detailed analysis of transpositions in serial recall, which is used to test the alternative model in Chapter 5.

### Phonological Similarity

An abundance of empirical data suggests that representations underlying performance in most verbal short-term memory tasks are speech-based. The order of items that are pronounced similarly (even if they are read in silence), such as *B, D, P,* is harder to recall than the order of items that are pronounced differently, such as *C, F, J* (e.g., Baddeley & Ecob, 1970; Conrad & Hull, 1964). This *phonological similarity effect* (Baddeley, 1986) occurs in spite of the fact that the items themselves are more likely to be recalled when similar, albeit in the wrong order, as can be demonstrated by comparing serial with free recall (Watkins, Watkins & Crowder, 1974).

Wickelgren (1965b) offered an explanation for the phonological similarity effect in terms of a simple chaining model, where items are stored by pairwise associations between their constituent phonemes. Assuming each phoneme has a single (type) representation in memory, repeated phonemes, such as the vowel */i:/* in the list *B, D, P,* are associated with more than one successor (i.e., */d/* and */p/*). Such lists are therefore formally equivalent to lists with repeated items, and the phonological similarity effect occurs for the same reason as associative intrusions (Chapter 1; Wickelgren, 1966). That is, phonological similarity acts on the cuing of items, because repeated phonemes are ambiguous cues for their successors.

A similar prediction would appear true of other chaining models. The most obvious way to model phonologically similar items in TODAM (Murdock, 1983) and recurrent networks (e.g., Jordon, 1986) would be to assume overlapping (nonorthogonal) vector representations. This would also produce an effect of similarity on cuing. Indeed, a general property of such distributed, associative memories is that “...errors are more likely when discriminations must be made between similar states...” (Jordan, 1986, p. 37). The exact predictions of a compound chaining model that chains along phonological representations are shown analytically in Henson (1994).

Baddeley (1968; Experiment V) tested whether phonological similarity affects the cuing of items, as suggested by chaining models, or whether it affects the retrieval of items. He used immediate serial recall of lists of six items, where the items were drawn from a set of letters pronounced similarly (the *confusable items*) and a set of letters pronounced differently (the *nonconfusable items*). With lists in which confusable and nonconfusable items alternated, error position curves revealed a “sawtooth” shape, where the peaks of the sawteeth represented errors in recall of confusable items, and the troughs represented fewer errors in recall of adjacent nonconfusable items (e.g., Figure 2-1). The sawteeth for these *alternating lists* were confined within more conventionally bowed curves for two *pure lists*: the *confusable lists*, which contained only confusable items, and the *nonconfusable lists*, which contained only nonconfusable items. While the peaks of the sawteeth lay below the curve for confusable lists, the troughs were virtually coincident with the curve for nonconfusable lists.

Baddeley argued that the fact that most errors in recall of alternating lists occurred for confusable items, rather than the nonconfusable items that followed them, favoured the idea of phonological similarity acting on retrieval rather than on cuing. Indeed, the fact that the confusable items in alternating lists had little to no effect on recall of the nonconfusable items, when compared with those in nonconfusable lists, suggested that there is no effect of phonological similarity on cuing.

Disregarding chaining models on the basis of these results is premature however. Sawteeth on their own are certainly insufficient. This is because chaining models could predict an effect of similarity on retrieval as well as on cuing (e.g., at the deblurring stage of TODAM; Lewandowsky & Li, 1994). Sawteeth could then result if the effect of phonological similarity is simply greater on retrieval than on cuing. The apparent coincidence of alternating and nonconfusable curves, for recall of nonconfusable items, is harder to reconcile with chaining models. However, this coincidence was not found in Experiment VI of the same paper, which used auditorily presented words. Moreover, a more sophisticated, probabilistic analysis shows the combined effect of phonological similarity on cuing and on retrieval can, in principle, reconcile chaining models with Baddeley's data (Henson, 1994; Henson et al., 1996).

## Experiment 1

The first aim of Experiment 1 was to replicate Baddeley's results with a more powerful design geared towards detecting an effect of phonological similarity on cuing. The most important comparison was between recall of nonconfusable items in alternating lists (where they were preceded by a confusable item) and recall of nonconfusable items in nonconfusable lists (where they were preceded by another nonconfusable item). An impairment in recall of nonconfusable items when their predecessors were phonologically similar to other list items would constitute evidence for chaining models.

A second aim was to conduct a more thorough analysis of subjects' responses. Though Baddeley reported errors by position, he did not examine the actual types of error, such as whether the errors were omissions or substitutions. Such analysis addresses further theoretical questions. For example, some theories suggest that similar representations degrade faster than dissimilar ones, as in Posner and Konick's (1966) "acid bath" theory. In this case, the peaks of the sawteeth in Baddeley's data may have reflected a greater incidence of confusable items being omitted, or being substituted for a random guess. However, if phonological similarity acts through response competition during retrieval, then the majority of these errors should be confusions; that is, one confusable item being substituted for another (e.g., Bjork & Healy, 1974; Conrad, 1965). As shown in Henson et al. (1996), this type of substitution is important if chaining theories are to be reconciled with Baddeley's data.

One modification in design of the present experiment was to generate the lists from a small experimental vocabulary, and to block the conditions separately, rather than intersperse them randomly as in Baddeley's experiment. This ensured that all lists in a block contained the same six items (conforming to the "order only" condition of Healy, 1974). With such a design, subjects know in advance which particular items will be presented, and need only concentrate on the order in which they occur. Consequently, minimal numbers of intrusions and omissions were expected, making transpositions the most likely errors. This allowed the simplifying assumption that reports were permutations of list items, and hence determination of the chance probabilities of certain responses.

A further interest was the distribution of associate errors (Chapter 1). Given an effect of phonological similarity on cuing, phonological chaining models predict that associates will be more frequent for nonconfusable than confusable lists. This is because a nonconfusable item is more likely to cue its successor in the list than is a confusable item (which partially cues other items; Henson, 1994). More generally, any closed-loop chaining model predicts that the frequency of associates should exceed that expected by chance, irrespective of phonological similarity. This is because the erroneous item, even if only part of the cue for the next response, will still increase the probability of recalling its successor, rather than the correct successor. These constituted two more specific tests of chaining theory.

A final modification in design was that subjects in the present experiment were encouraged to group the six items into two groups of three. Baddeley did not give such instruction to his subjects. However, grouping strategies are often brought to bear on the most simple of span tasks, and can have dramatic effects on the pattern of errors (Chapter 3). Particular advantage is conveyed to recall of the first and last items in a group, revealed as primacy and recency effects within each group. Indeed, a suggestion of such spontaneous grouping is apparent in Baddeley's error position curves, particularly for confusable lists. The concern was that grouping strategies might interact with the structure of alternating lists. For example, a choice of grouping in twos rather than threes may affect the nature of errors made in recall of alternating lists. Thus the explicit instruction to group in threes in the present experiment was to encourage a single, consistent grouping strategy across subjects.

In summary, the aims of the experiment were: 1) to reproduce and make explicit tests of Baddeley's findings, specifically the sawtooth error position curves for alternating lists; and 2) to conduct a more thorough analysis of errors.

## Method

### Subjects

Forty-eight subjects from the APU Subject Panel were tested, of whom seventeen were male and thirty-one were female. Their mean age was twenty-seven years.

**Materials**

Stimuli were lists of six, single-syllable consonants, generated from a vocabulary of twelve. The letters were classified according to their confusability; that is, whether they were phonologically similar to any other letters in the vocabulary. The six confusable letters shared a common rhyme when pronounced: *B, D, G, P, T, V;* the six nonconfusable letters possessed unique rhymes: *H, K, M, Q, R, Y.*

The two pure list types were the confusable lists, containing all six confusable letters, and the nonconfusable lists, containing all six nonconfusable letters (conditions PC and PN respectively). Two alternating list types (A1 and A2) were identified according to the two mutually exclusive sets of three confusable and three nonconfusable letters in the vocabulary (Table 2-1). These lists comprised the two alternating conditions, according to whether the alternation began with a confusable or a nonconfusable item in the first position (conditions AC and AN respectively). Conditions AC and AN were nested inside list types A1 and A2, such that a block of A1 or A2 lists contained six lists of condition AC and six of condition AN. With the randomised order of lists within blocks, this nesting was to reduce the chance of subjects' detecting a pattern of confusable-nonconfusable alternation.

The lists were generated according to the following constraints: None of the lists contained obvious acronyms (nor cooccurrence of letters in alphabetical order), each letter appeared equally often (twice) in each position, and the frequency of adjacent letter pairs was made as uniform as possible, after the above considerations had been met. In other words, first- and second-order contingencies between items were close to being balanced.

**Procedure**

Every subject attempted recall of 4 blocks of 12 lists, each block containing lists of one of the list types PC, PN, A1 or A2. Before the first list of each block, the six letters that would appear in the following 12 lists were presented in a circle, in order to familiarise subjects with the set of possible responses. Subjects were told that the lists contained no repeated letters. The trial order of the 12 lists within blocks was randomised and the block order was fully counter-balanced across subjects.

The experiment was run on an IBM PC, with the capitalised letters appearing in the

| Condition | List Type | List Structure | Letter Set (Example List) | Number of Lists |
|---|---|---|---|---|
| PC | PC | CCCCCC | BDGPTV | 12 |
| PN | PN | NNNNNN | HKMQRY | 12 |
| AC | A1 | CNCNCN | DQTMPK | 6 |
| | A2 | CNCNCN | BHGYVR | 6 |
| AN | A1 | NCNCNC | QDMTKP | 6 |
| | A2 | NCNCNC | HBYGRV | 6 |

Table 2-1: Composition of lists in Experiment 1.
(C=confusable item, N=nonconfusable item).

centre of a monochrome VDU, each letter approximately half an inch high and replacing the previous one. Presentation rate was two letters per second (400-ms on; 100-ms off). Subjects were instructed to read the letters in silence, and immediate recall was prompted after the last item disappeared, with letters written left to right across a row of six boxes provided on a response sheet. A minimum of 10 seconds was required between trials, after which subjects pressed a key to start the next trial. A short break of a minute occurred between blocks.

Subjects were instructed to write down answers immediately and, if unsure, told to “write the first letter that comes to mind”. If no letter came to mind, they were asked to put a line through the appropriate box. Subjects were reminded to recall in a forward manner, writing from left to right on the response sheet, and to resist the temptation to recall the last few letters first. Finally, subjects were advised that grouping the six letters into two groups of three may aid their retention; an example of such 3-3 grouping of a telephone number was given. Three practice trials then followed. The whole experiment took about 20 minutes.

**Results**

In brief, the results replicated those of Experiment V in Baddeley (1968), though there was a confounding effect of the predictability of lists (Henson et al., 1996). Nevertheless, there were significant sawteeth in alternating conditions, in addition to normal primacy and recency. Closer analysis showed that the peaks of these sawteeth reflected confusable items transposing with one another, and that such confusions were sensitive to transposition distance. Most importantly, there was little evidence for an effect of similarity on cuing, even taking into account predictability, while there was clear evidence for an effect on retrieval. This was apparent in both error position curves and more detailed analyses of associate errors.

**Overall Performance**

Approximately 20% of PC lists, 58% of PN lists, 55% of AC lists and 51% of AN lists were recalled correctly. Omissions comprised approximately 5% of errors, while intrusions amounted to only 3%. The rarity of such errors reflected the small experimental vocabulary, and probably accounts for the higher level of performance than in Baddeley’s experiment.

**Predictability**

Error position curves (upper panel of Figure 2-1) replicated the main features of Baddeley’s. The effect of phonological similarity extended over all positions in the pure confusable lists, but just the positions of confusable items in the alternating lists. Importantly, there was no evidence of more errors for nonconfusable items in alternating lists than in pure nonconfusable lists. In fact, nonconfusable items were recalled slightly better in alternating lists (i.e., the sawteeth straddled the nonconfusable curve, rather than sitting on top of it).

As reported in Henson et al. (1996), closer inspection of the stimuli suggested a reason for this: The letters in different list types differed in their *predictability* (e.g., how often the letters cooccur in the English language; Baddeley, Conrad & Hull, 1965). Letters in the A2 lists were especially predictable. This might explain why performance in alternating conditions AC and AN was slightly better than expected, compared to condition PN.

The counterbalanced design of lists meant that predictability should not affect tests within conditions. However, predictability did potentially confound tests across conditions. Though the larger experimental vocabulary in Baddeley's experiments made such a confound less likely, a similar caution should apply to his results also. Without equating predictability across conditions, one cannot be sure that performance on nonconfusable items in alternating lists was truly unaffected by the presence of confusable items.

Two further experiments in Henson et al. (1996) controlled for the predictability of letters (as did all subsequent experiments in the present thesis). The approach taken here, when comparing across conditions, was to remove the A2 lists from analysis, so that the AC

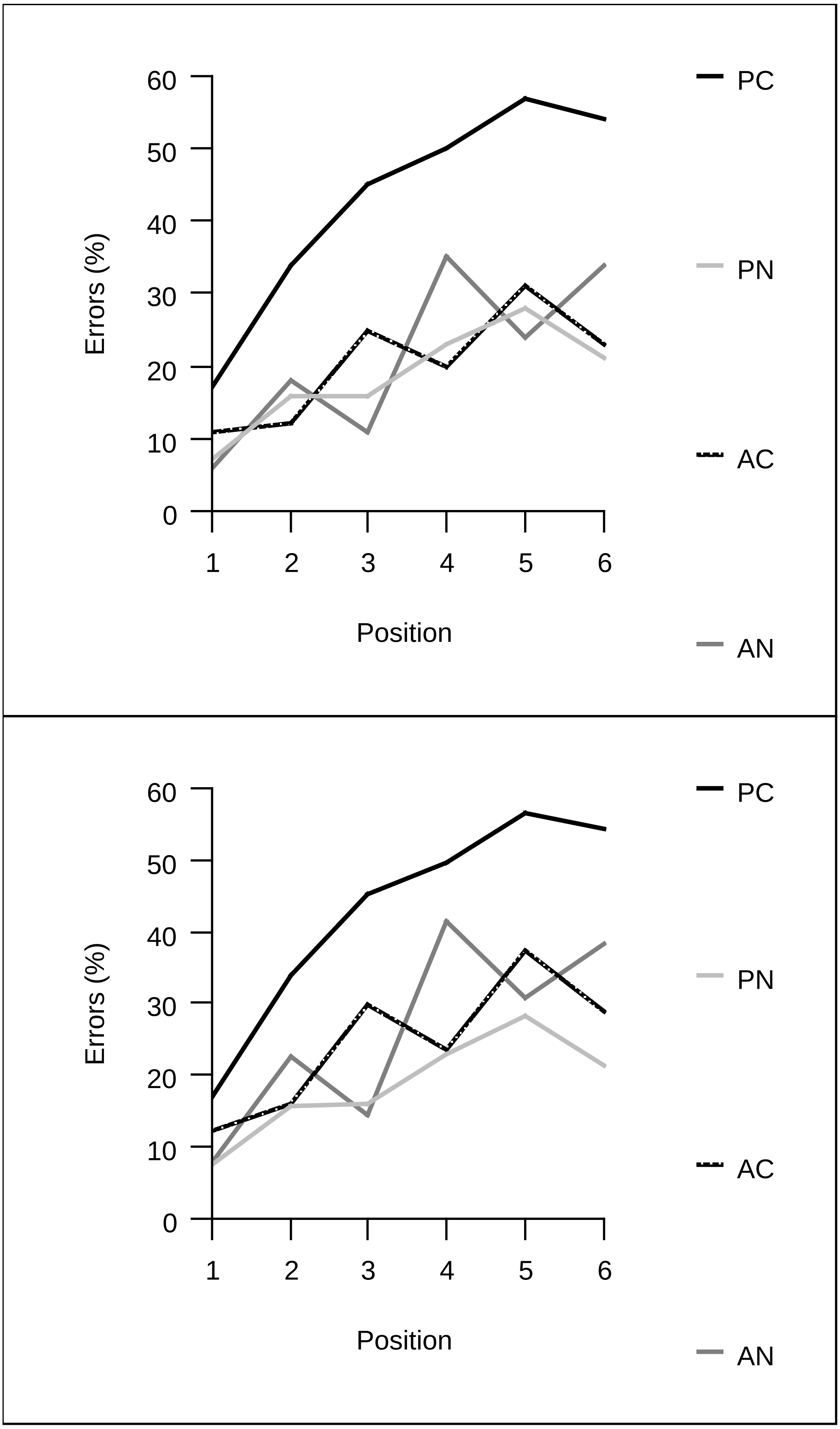



Figure 2-1: Errors by position including A2 lists (upper panel) and excluding A2 lists (lower panel) in Experiment 1.

and AN conditions were calculated from the A1 list type only (lower panel of Figure 2-1). The predictability of letters in A1 lists was less than that in PN and PC lists (Henson et al., 1996), which will tend to increase errors in alternating lists compared to the pure lists. A failure to find significantly more errors for nonconfusable items in conditions AC and AN than PN can not therefore be attributed to different predictabilities.

**Comparisons within Conditions**

In order to test Baddeley's findings statistically, planned comparisons were performed on the log-odds of an error (Chapter 1) across each position in the upper panel of Figure 2-1 (i.e., including A2 lists) in a separate ANOVA for each condition.

Two linear, orthogonal comparisons for the pure lists, PC and PN, tested for primacy (the average error score on Positions 1 and 2 compared with the average on Positions 3 and 4) and last-item recency (the error score on Position 6 versus Position 5). Both confusable and nonconfusable curves showed significant primacy, $F(1,235)>18.00$, $MSE<0.52$, $p<.001$, but only the nonconfusable curve showed significant recency, $F(1,235)=8.18$, $MSE=0.50$, $p<.01$ ($F<1$ for the confusable curve).

Three comparisons for alternating lists AC and AN tested the significance of the sawtooth shape (the error score on confusable positions compared to adjacent nonconfusable positions). A fourth contrast looked for an effect of primacy over the first four positions (as defined above). For both alternating curves, errors were significantly more common on confusable positions than adjacent nonconfusable positions, $F(1,235)>4.62$, $MSE<0.57$, $p<.05$ in all cases, except between the first two positions of condition AC, $F<1$. The latter probably reflected the opposing effect of primacy, which was significant in both AC and AN conditions, $F(1,235)>42.25$, $p<.001$.

**Comparisons between Conditions**

To test the predictions of phonological chaining models, the weighted log-odds of an error on nonconfusable positions in alternating lists was compared to that on nonconfusable positions in nonconfusable lists. Including A2 lists, there was no significant difference on any of the six positions, $Z(48)<0.39$, $p>.70$. This may have reflected the less predictable nature of the A2 lists. Even excluding A2 lists however, there was still no greater probability of an error in alternating lists for any position, *Z(48)<1.81, p>.07*, except the last, *Z(48)=3.16, family-wise p<.01* (using Holm's correction for the multiple comparisons). Thus, when examining the troughs of the sawteeth in Figure 2-1, there was only evidence for an effect of similarity on cuing for one of the six positions, providing A2 lists were excluded.

In addition, the weighted log-odds of an error on confusable positions in alternating lists was compared to that on corresponding nonconfusable positions in nonconfusable lists. Both including and excluding A2 lists, there was a significantly greater probability of errors on all six confusable positions, *Z(48)>3.21, family-wise p<.01*. Thus, when examining the peaks of the sawteeth in Figure 2-1, there was evidence for an effect of similarity on retrieval for all positions, whether or not A2 lists were excluded.

Finally, the weighted log-odds that adjacent transpositions were associates was calculated for confusable and nonconfusable lists, for the 33 subjects who made at least one pair of adjacent transpositions in both conditions. There was no evidence for a greater probability of associates in nonconfusable lists (*M=.23*, *SD=.22*) than confusable lists (*M=.22, SD=.15*), *Z(33)=0.08, p=.94*; another failure to find any effect of similarity on cuing. Nor did these probabilities differ significantly from a chance probability of *.20* (assuming the second error could be one of five list items), for either nonconfusable lists, *Z(33)=0.23, p=.82*, or confusable lists, *Z(33)=0.25, p=.80*, contrary to closed-loop chaining models.[1]

**Transpositions**

Transposition gradients were also calculated for each condition, collapsing across subjects (Figure 2-2; the six bars for each output position represent the percentage of responses from each input position, from left to right, the tall bars being correct responses). For PC and PN lists, transpositions decreased monotonically with increasing distance between input and output position. This monotonic decrease was remarkably lawful: The rank ordering of transpositions for each output position would be expected only 1 in 120 times if subjects guessed list items at random. The only exception to this monotonic decrease was transpositions from the first to the last position in PN lists; further inspection revealed that these were mainly repetitions (Chapter 4). The transposition gradients for the AN lists did not

1. Even if the mean probability were *.30*, the null hypothesis could not be rejected at the *.05* level (*power=99%*).

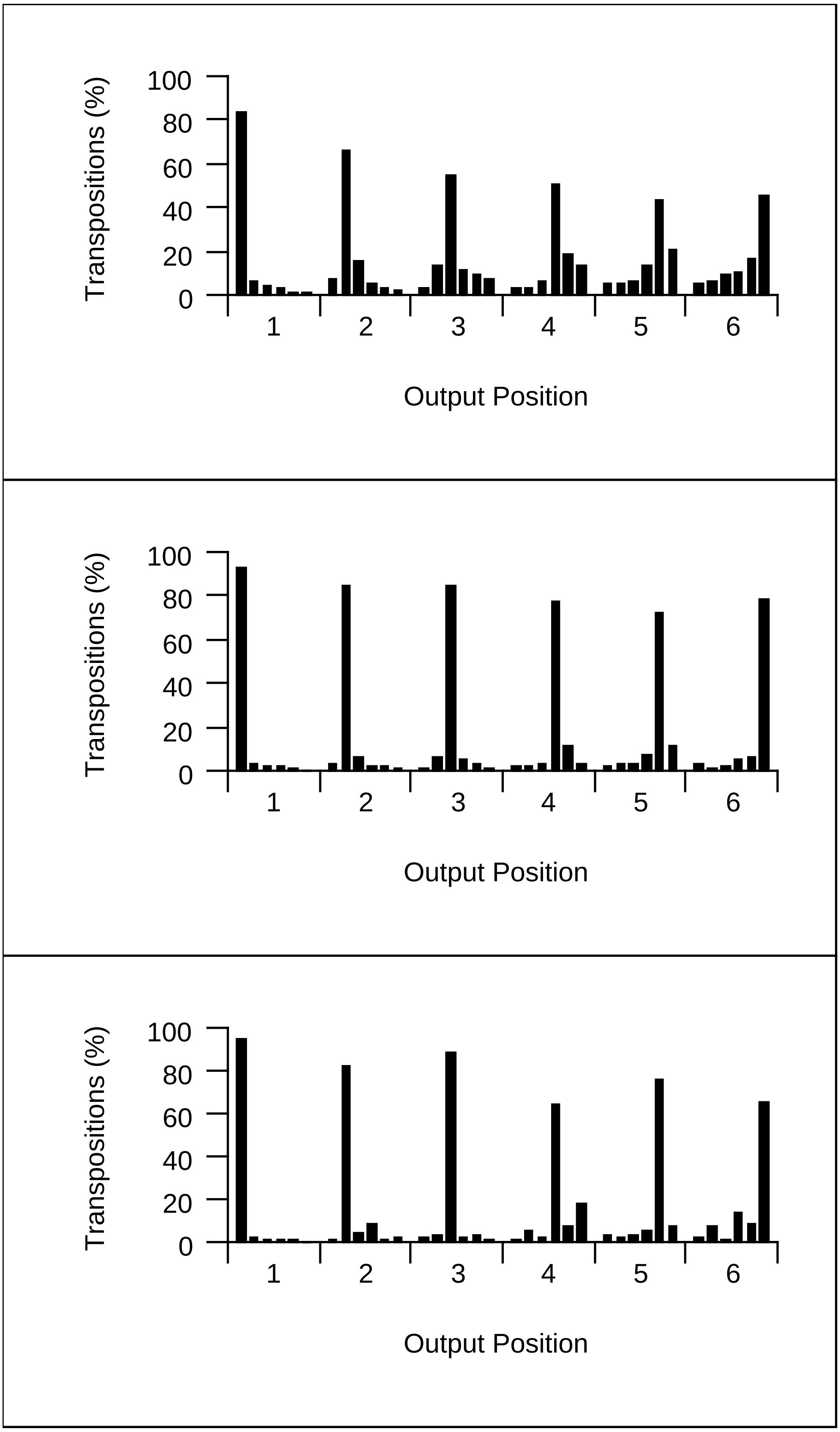



Figure 2-2: Transposition gradients for conditions PC (upper panel), PN (middle panel) and AN (lower panel) in Experiment 1.

decrease monotonically, but were a function of the phonological similarity between the correct and transposed item. Thus transposition gradients for output positions that corresponded to input positions of confusable items (Positions 2, 4 and 6) were punctuated by peaks for input positions of other confusable items. The same pattern arose for AC lists. Because the majority of reports were in effect permutations of list items, given that most errors were nonrepeated transpositions, the sawteeth shape of error position curves logically requires that the majority of transpositions in alternating lists were confusable items transposing with one another.

When transpositions were replotted against transposition distance, the gradients for confusable and nonconfusable lists were not parallel: The gradients were steeper for confusable curves (Figure 2-3), a finding confirmed in Chapter 4. In other words, the effect of phonological similarity was not additive, implying that phonological confusions do not arise independently of position (Chapter 5).

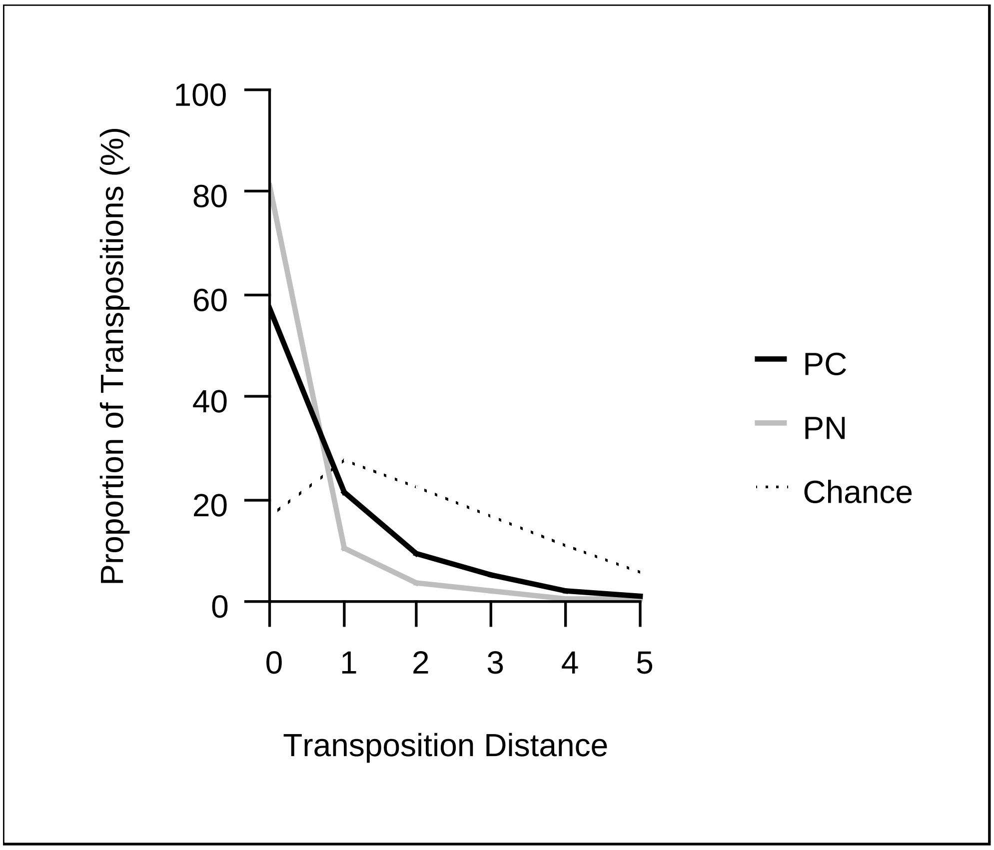



Figure 2-3: Proportion of transpositions (including correct responses) by transposition distance, together with chance levels, in Experiment 1.

**Discussion**

Results from the present experiment, and others in Henson et al. (1996), consistently fail to provide any support for chaining theory. In fact, they are difficult to reconcile with any current chaining model.

The prominent sawteeth in the error position curves for alternating lists reflect significantly more errors in recall of confusable than nonconfusable items. Though chaining models can be constructed that are compatible with this sawtooth pattern, they remain unable to explain the fact that, whether unconditional (or conditional, Henson et al., 1996) error probabilities are examined, the presence of confusable items in a list most often has no detectable effect on the probability of recalling following nonconfusable items. In the present experiment for example, only one nonconfusable position in six showed significantly more errors in alternating curves that nonconfusable curves, and this could owe to the less predictable nature of the alternating lists (with A2 lists removed). Though one can never be certain whether the failure to find a significant difference on the other five positions reflected a lack of statistical power, the same experiment was powerful enough to show significantly more errors on all six confusable positions in alternating curves. These findings meet Frick’s (1995) criteria for accepting the null hypothesis at least (though see Chapter 4).

Nevertheless, a number of questions might be asked of the present results. Firstly, there is the question of whether the phonological confusions arose during encoding rather than retrieval. Confusions during encoding would seem unlikely with visual presentation however; none was observed when subjects read items aloud during presentation (Henson et al., 1996, Experiment 3). Even with auditory presentation, an encoding account seems insufficient (Baddeley, 1968, Experiment IV). Secondly, there is the question of strategic effects: Subjects might treat confusable and nonconfusable items differently in alternating lists (e.g., by streaming them, particularly with auditory presentation; Jones, 1992). This also seems unlikely, given that not one subject in the present experiment detected the alternating pattern in A1 and A2 lists. Finally, there is the question of generality: The present results did not hold perfectly for subjects with low memory spans (Henson et al., 1996), who showed less well-defined sawteeth. This may reflect subjects abandoning phonological coding when they approach the limit of their memory (Salame & Baddeley, 1986). More likely, the differences reflect “knock-on” effects when recall becomes difficult and subjects “give up” (Chapter 4). In any case, conditional probabilities of first errors (Henson et al., 1996), which remove knock-on effects of prior errors, showed the same pattern as the present results.

These questions aside, there are two aspects of the present data that are troublesome for chaining theory: 1) recall of nonconfusable items was little affected by whether the previous item was confusable, and 2) recall of nonconfusable items was little affected by whether or not the previous item was confused. In other words, there was little evidence for an effect of similarity on cuing, or an effect of errors on cuing.

The lack of any effect of similarity on cuing was reinforced by the probability of associative errors, which did not depend on whether or not the previous item was confusable. More generally, the present results suggest that confusable items have little effect on any surrounding nonconfusable items. A similar conclusion was reached by Bjork and Healy (1974): “...the presence of two acoustically similar items in the same to-be-remembered stimulus does not increase the loss of order information for all letters in the stimulus string but rather produces rapid loss of order information specific to the two similar letters...” (p. 91).

The lack of an effect of similarity on cuing is troublesome for models that chain along phonological representations, such as that of Wickelgren (1965b). It appears troublesome for more recent models too, as soon as they adopt phonological representations (e.g., Murdock, 1983; Jordan, 1986). In particular, the inability of TODAM to simulate the Baddeley (1968) data was confirmed by Baddeley, Papagno and Norris (1991). The problems with chaining along phonological representations were also confirmed by Burgess and Hitch (1992), who obtained their best fits when associations between phonemes were minimised (in favour of position-item associations). The fact that present results hold when stimuli are vocalised (Henson et al., 1996, Experiment 3) is also problematic for theories that restrict chaining to the auditory modality (e.g., Drewnowski, 1980a; Penney, 1989).

One might argue that models that chain along nonphonological representations (e.g., context-sensitive tokens, Wickelgren, 1969) would not have to predict an effect of similarity on cuing. For example, TODAM might retain its normal assumption of random vector representations of items, and model phonological similarity as affecting only retrieval, during the subsequent "deblurring" of the results of chaining (Lewandowsky & Li, 1994). Another example is Richman and Simon's (1994) EPAM model. This model chains along unitised representations, or *chunks*, and locates phonological similarity in the retrieval of chunks. However, both these models face problems with the second aspect of the present data. As closed-loop chaining models, they still predict more errors on nonconfusable positions in alternating lists than in nonconfusable lists, because previous responses are more often in error in alternating lists. The additional errors on confusable items in alternating lists mean that the cue for the following nonconfusable item is correct less often than in nonconfusable lists. This prediction for an effect of errors on cuing was supported by neither error position curves, nor the incidence of associate errors, which were not significantly above chance.

The only type of chaining model consistent with present data would be an open-loop, nonphonological model (i.e., one that chained along nonphonological representations independently of feedback of previous responses). Such a specific model loses some of the intuitive appeal of chaining theory (e.g., that each response becomes the stimulus for the next). Moreover, given that there does not appear to be any data necessitating item-item chaining (Chapters 1, 4), and yet there is data necessitating positional information in short-term memory (Chapter 3), such a model does not seem worth pursuing.

The present results have in fact proved difficult for many models of serial recall, whether or not they employ chaining. For example, Burgess and Hitch (1992) stated in their abstract: "the model was unable to simulate human memory for sequences containing mixtures of phonemically similar and dissimilar items". This, together with comments in Henson et al. (1996), led to revision of the model (Burgess & Hitch, 1996a, 1996b). As such, the sawtooth shape of alternating curves is a bench-mark test for models of immediate serial recall (Page & Norris, 1996b). Chapter 5 presents a new model that can simulate memory for mixtures of phonemically similar and dissimilar items, and which passes this test.

Finally, the present experiment produced transposition gradients that replicate previous findings that items are more likely to transpose to nearby positions than positions far apart, the *locality constraint* (Chapter 4). The additional transpositions between confusable items suggest phonological similarity acts on retrieval, rather than via passive decay or interference during storage (Posner & Konick, 1966). That such confusions also respect the locality constraint is a new finding, which turns out to be important for modelling phonological similarity in short-term memory (Chapter 5).

## Chapter Summary

This chapter examined the first of the three theories of serial order in Chapter 1: chaining theory. The fact that phonologically confusable items had little detectable effect on recall of surrounding nonconfusable items, either through an effect of similarity on cuing, or through an effect of errors on cuing, is difficult for current chaining models to explain. Though one might construct a very specific chaining model to fit the present data, the onus would be on the modeller to demonstrate additional evidence for such specific assumptions. Moreover, given that the next chapter demonstrates evidence for an alternative theory of serial order, there seems little point in pursuing a chaining theory of short-term memory for serial order.

# Chapter 3: Positional Theory

## *Two experiments testing Positional Theory*

The previous chapter failed to find any evidence for chaining in serial recall, throwing doubt on models in which order is stored via item-item associations. The present chapter examines evidence for positional information in serial recall, in order to address models that assume some type of position-item association. In Experiment 2, this evidence comes from transpositions between groups that maintain their position within groups (interpositions). In Experiment 3, this evidence comes from intrusions between trials that maintain their position within trials (protrusions). Both experiments also continue to examine other errors in serial recall, such as omissions and intrusions, for the purposes of modelling in Chapter 5.

### Grouping

Grouping a sequence into smaller subsequences improves retention of that sequence (Ryan, 1969a; 1969b; Wickelgren, 1967). Ryan (1969a), for example, presented lists of nine digits auditorily, which were grouped into three groups of three by a number of means: by a pause after every third digit, by a tone pip after every third digit, or simply by instruction. Grouping generally improved recall, compared to an ungrouped condition, though the advantage only proved significant when grouping was achieved by pauses (temporal grouping). Grouping of auditory stimuli can also be achieved by alternating the laterality or the voicing of groups, which is as effective as, but not additive with, temporal grouping (Frankish, 1989). In fact, even accenting the pitch of the last item in each group is sufficient to produce an advantage equal to that of temporal grouping (Frankish, 1995). Effects of temporal and spatial grouping are also found with visual presentation, though they are much smaller than with auditory presentation (Frankish, 1985).

Nevertheless, all methods of grouping have similar and striking effects on the distribution of errors. In serial position curves, these effects are sometimes revealed as mini-primacy and mini-recency effects within groups, resulting in "scalloped" shapes for each group (e.g., Figure 3-1). More generally, grouping reduces the number of transpositions between groups, except those that maintain their position within groups (i.e., interpositions).

Ryan (1969b) found that presentation rate had little effect on the size of the temporal grouping effect. Frankish (1989) showed further that, providing the pause was perceptible, auditory temporal grouping was unaffected by pause length. Frick (1989) showed that a similar grouping effect remained under concurrent articulatory suppression, which prevents rehearsal (Baddeley, 1986). These results suggest that, though extra rehearsal during pauses may play a role, particularly in the visual modality (Frankish, 1989), differences between temporally grouped and ungrouped lists mainly reflect alternative representations in memory.

One alternative representation is the recoding of groups into single units or *chunks* (Miller, 1956/1994). The advantage of grouping is then to reduce the number of chunks that must be recalled, from nine in Ryan's ungrouped case, to three in her grouped case. However, as single units, chunks are recalled in an all-or-none fashion (Johnson, 1972). While this might explain the overall reduction in transpositions between groups, it has problems explaining why interpositions between groups remain. Most interpositions occur singly, and are not an artefact of whole groups swapping (Lee & Estes, 1981; Nairne, 1991). Moreover, chunking usually implies the preexistence of long-term memory codes, which is not the case for the novel groupings of items in the above experiments. Grouping is more likely to reflect a "reordering" rather than "recoding" of items (Frankish, 1974).[1]

Grouping is not necessarily contingent on objective organisation of a list: People will often spontaneously group lists. Such *subjective grouping* was noticed by Frankish (1974): "organisational strategies are widely used even in comparatively straightforward tests of short-term memory, such as the digit-span task." (PhD abstract). For example, people will often tend to group digits into threes, as they might do for a telephone number. Surprisingly however, subjective grouping in serial recall is often overlooked. Closer inspection of serial position curves nearly always reveals some scalloping, particularly over the third and fourth positions. Though this might be dismissed as random variability in one study, the pattern is consistent

1. Though the notion of chunks and groups are distinct, they are related, as demonstrated by Bower and Springston (1970). They showed lists like *ITVRSVPBT* were recalled well when the group structure was consistent with the chunk structure (e.g., *ITV RSVP BT*), but not when it was inconsistent (e.g., *IT VRSV PBT*). Indeed, a constant grouping structure may be necessary for the development of new chunks (Bower & Winsenz, 1969). Thus chunks might be viewed as groups that have become crystallised in long-term memory.

over numerous studies (Madigan, 1980). Given the powerful effects grouping has on performance, not only in terms of lists correct, but also serial position curves and underlying error distributions, it is unwise to ignore subjective grouping. The prevalence of such grouping suggests that it plays an important role in storing serial order in short-term memory.

## Experiment 2

The first aim of Experiment 2 was to examine the detailed effects on error patterns of temporal grouping of visually presented digits. The second aim was to examine error patterns in lists of seven, eight and nine digits, without such objective grouping. Most people have considerable difficulty in serial recall of lists of seven or more items (Miller, 1956/1994) and many studies have shown how the percentage of lists recalled correctly decreases with list length in an inverse sigmoidal fashion (e.g., Crannell & Parrish, 1957). However, it is unclear from previous studies how this difficulty is manifest in error patterns. Some studies (e.g., Drewnowski & Murdock, 1980) have shown serial position curves which “stretch out” with increasing list length, maintaining characteristic primacy and recency effects, but few, if any, have examined any change in the nature of these increasing errors. Does recall of longer lists produce more omissions, more intrusions, or simply many more transpositions? Do people tend to give up after recalling the first few items, leave a series of omissions in the middle, before having a stab at the last one or two items (to give a recency effect)? This is an important consideration when recall of the middle items of long lists falls as low as 20% correct (e.g., Murdock, 1968; Madigan, 1971). Moreover, the exact effects of list length and grouping on detailed error patterns is important for testing models of serial recall (Chapter 5).

The issues of supraspan recall and grouping are closely related. It is possible that people are only able to recall lists of more than five or six items by resorting to subjective grouping of lists. This possibility has been neglected in previous studies of supraspan recall.

### Method

#### Subjects

Eighteen subjects from the APU Subject Panel were tested, seven male and six female, with a mean age of twenty-seven years.

**Materials**

Stimuli were lists of seven, eight or nine digits, drawn without replacement from vocabularies of eight, nine or ten digits respectively, in the range 0-9. The order of digits was random, except for no pairs of stepsize one (e.g., “34” or “43”) or triplets of stepsize two (e.g., “246” or “579”). Each digit appeared approximately equally often at each position.

**Procedure**

In the three ungrouped conditions of seven, eight and nine digits (conditions U7, U8 and U9 respectively), digits were presented at the rate of one every 600-ms (400-ms on; 200-ms off). The fourth condition (condition G9) had nine digits grouped as three groups of three (3-3-3 grouping). This condition had faster presentation of digits within groups, one every 450-ms (400-ms on, 50-ms off), and a pause of 450-ms between groups, so that the total presentation times for grouped and ungrouped nine-item lists was equated. Subjects were instructed to use pauses to group the digits in threes, as they might do for a telephone number.

Each digit was presented in the centre of a VDU, replacing the previous one. Subjects read the digits in silence, before attempting written recall immediately after the last digit had disappeared. Responses were written on a sheet containing the appropriate number of boxes for each list-length. Subjects were instructed to recall in a strictly forward manner, writing from left to right on the response sheet. They were encouraged to guess if they were unsure, but if no digit came to mind, to put a line through the box and proceed to the next box.

Subjects were tested in all four conditions, each as a block of 20 lists. Conditions were always attempted in the order: U7, U8, U9, G9, to reduce the chance of subjects spontaneously grouping the ungrouped lists, as might occur if the G9 condition occurred before the others. Subjects received ten practice trials. The whole experiment took about 40 minutes.

**Results**

In brief, all conditions showed evidence of grouping in error position curves, though the effects were strongest in the objectively grouped condition. Longer lists produced more transpositions and omissions, though the distribution of these errors differed: Transpositions showed a recency effect whereas omissions did not. Grouping decreased both transpositions and omissions, and also affected their distribution. In particular, most transpositions decreased, except for three- and six-apart transpositions between groups, which increased. These interpositions were most common between the middle of groups.

**Overall Performance**

The percentage of lists recalled correctly was approximately 38% in the U7 condition, 21% in the U8 condition, 9% in the U9 condition, and 15% in the G9 condition.

**Error Position Curves**

The effects of list-length and grouping were apparent in error position curves (Figure 3-1, upper and lower panel respectively). All error position curves showed some scalloping, suggesting spontaneous grouping by subjects: seven-item lists by 3-4 grouping, eight-item lists by 4-4 (or even 2-2-2-2) grouping and nine-item lists by 3-3-3 grouping. In other words, lists in conditions U7, U8 and U9, though not grouped objectively, were nevertheless grouped subjectively. Indeed, most subjects reported using some form of grouping in these conditions, even though objective grouping was only introduced in the last condition, G9. The subjective grouping in condition U9 was unfortunate, in that it no longer provided a truly ungrouped baseline with which to compare condition G9. Nevertheless, it remains unclear whether truly ungrouped recall of supraspan lists is possible.

Longer list-lengths increased errors on nearly all positions, including the first, as revealed by a one-way ANOVA on the log-odds of an error on Position 1 in conditions U7, U8 and U9, $F(2,34)=7.78$, $MSE=0.55$, $p<.005$. A two-way ANOVA on the log-odds of an error in conditions U9 and G9 showed a significant effect of objective grouping, $F(1,323)=21.47$, $MSE=0.63$, $p<.001$, and position, $F(8,323)=61.52$, $p<.001$, but no significant interaction, $F(8,323)=1.12$, $p=.35$. The lack of an interaction confirmed that the majority of subjects grouped the nine items as three groups of three in both conditions.[2] Nevertheless, the main effect of objective grouping showed that the explicit pause in condition G9 made 3-3-3 grouping more effective and/or more consistent across trials and subjects. Comparisons across conditions U9 and G9 were still useful indices of grouping therefore.

2. The presence of subjective group boundaries in the U9 condition was confirmed by a conditional analysis of first errors (Chapter 1). Though the conditional probability of a first error tended to decrease across Positions 2-3, 4-6 and 7-9 (Appendix 1), there was a significant increase in this probability across Positions 3 and 4, $Z(18)=5.14$, $p<.0001$, and Positions 6 and 7, $Z(18)=3.19$, $p<.005$, indicating the start of a new group.

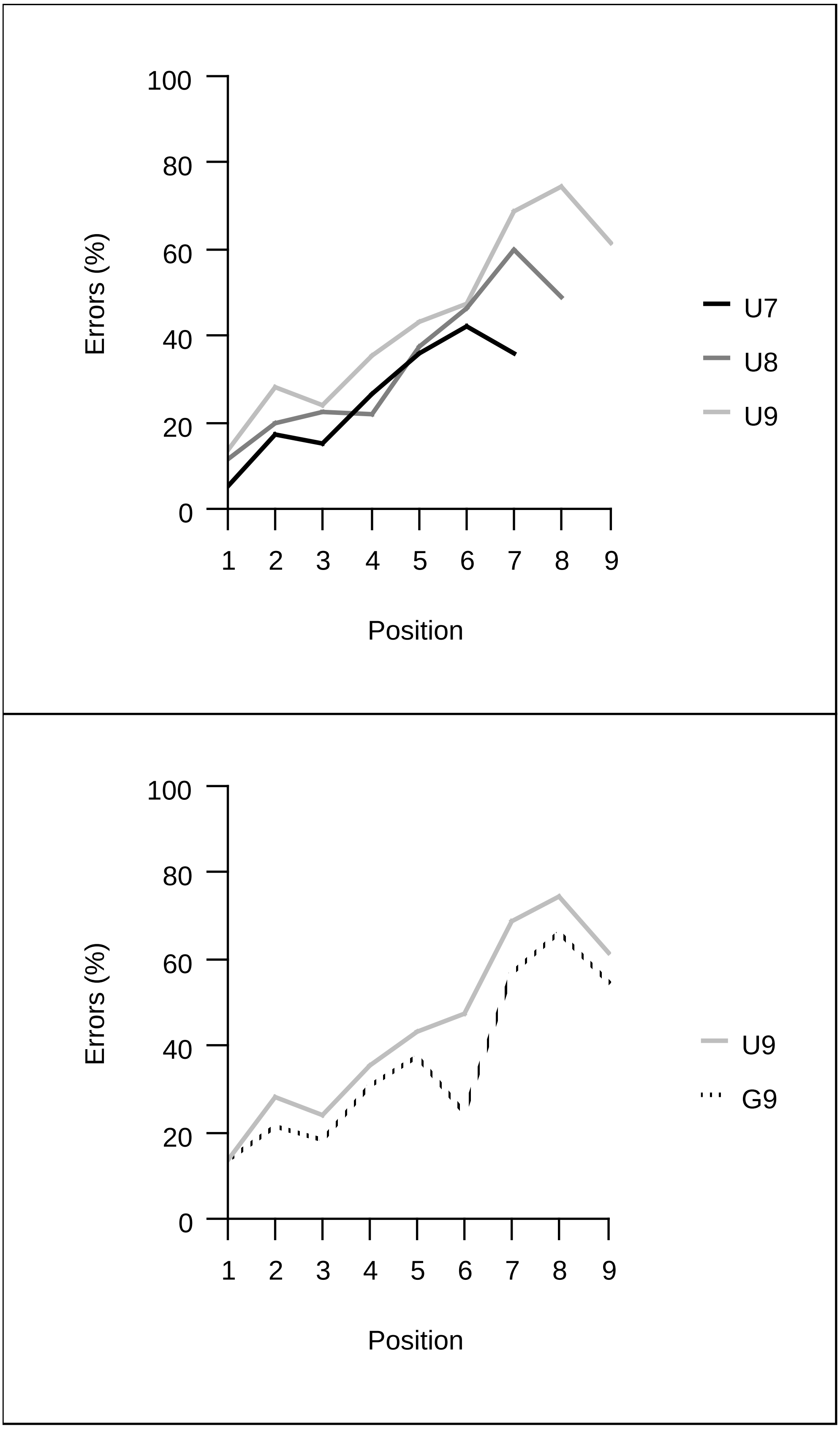



Figure 3-1: Errors by position for ungrouped lists (upper panel) and nine-item lists (lower panel) in Experiment 2.

**Error Types**

The proportion of responses that were omissions or transpositions was calculated for each condition (Table 3-1). Intrusions were rare (less than 5% of responses), as expected from the limited experimental vocabulary, as were repetitions. (Intrusions are analysed in more detail in Experiment 3; repetitions are analysed in more detail in Chapter 4.) A two-way ANOVA on the log-odds of an error for conditions U7, U8 and U9 showed a significant effect of error type (omission or transposition), $F(1,107)=48.41$, $p<.001$, list length, $F(2,136)=7.20$, $p<.005$, but no significant interaction, $F(2,107)=1.87$, $p=.16$. Transpositions were more common than omissions and both increased with list length. A two-way ANOVA on the log-odds of an error for conditions U9 and G9 also showed significantly more transpositions than omissions, $F(2,85)=11.87$, $p<.005$, but no significant effect of grouping or interaction, $F<1$ in both cases (though more specific tests below did show differences between these conditions).

| Condition | Omissions | Transpositions |
|---|---|---|
| U7 | .05<br>(.06) | .19<br>(.10) |
| U8 | .10<br>(.08) | .22<br>(.10) |
| U9 | .18<br>(.15) | .25<br>(.14) |
| G9 | .14<br>(.11) | .20<br>(.12) |

Table 3-1: Frequency of omissions and transpositions in Experiment 2.

**Omissions**

When split by output position, omissions in the ungrouped conditions showed a monotonic increase towards the end of recall (Figure 3-2). A similar increase was found for condition G9 at the level of groups, such that whole groups tended to be omitted towards the end of recall. One way of explaining this pattern of omissions is through a knock-on effect, where as soon as subjects forget an item, they “give up” on the remaining items, resulting in omissions for all subsequent positions (Experiment 1). However, this is not always the case,

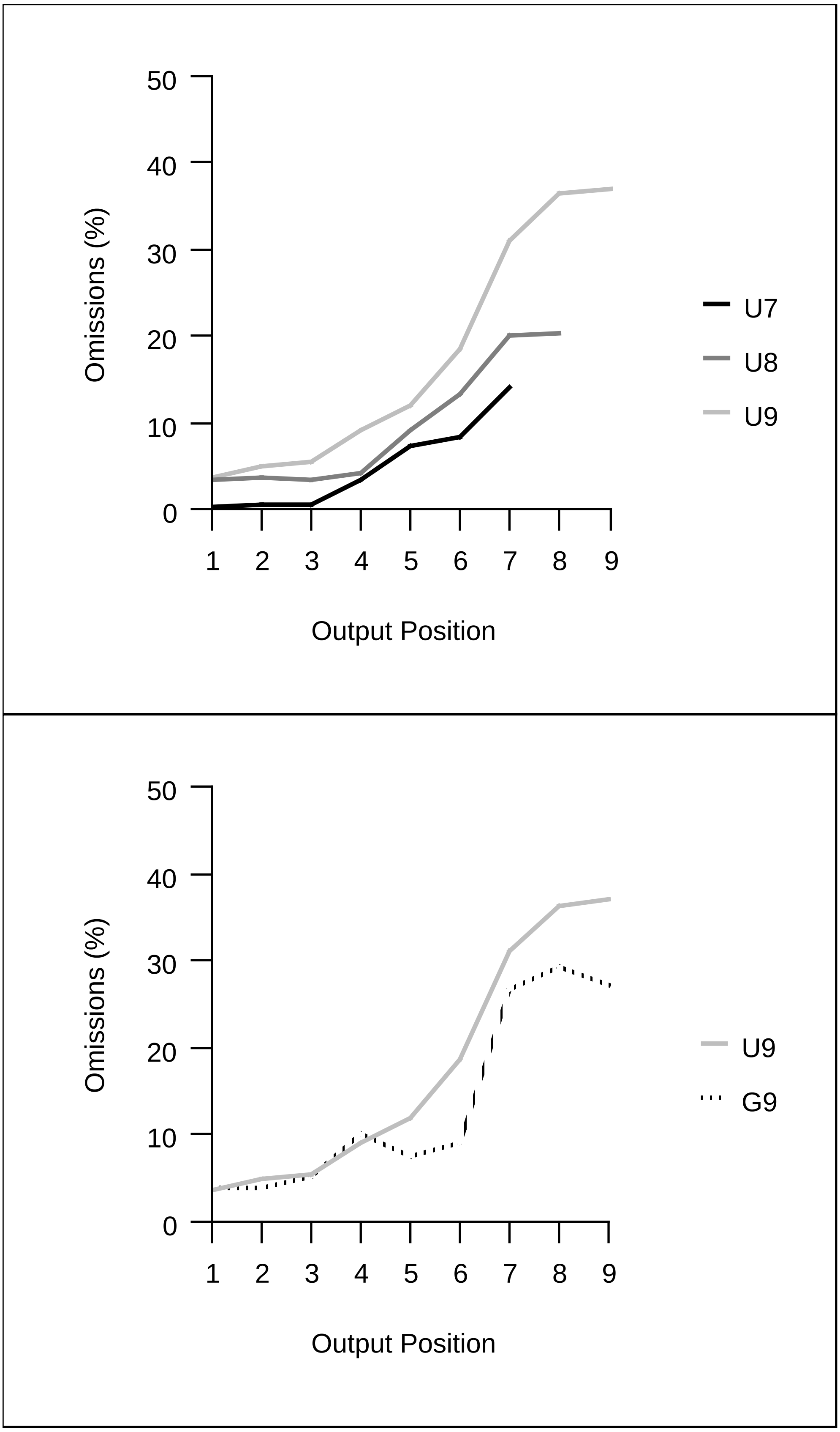



Figure 3-2: Omissions by output position for ungrouped lists (upper panel), and nine-item lists (lower panel) in Experiment 2.

because approximately 25% of the time, subjects who omitted one response (drew a line through the appropriate box) went on to make further responses (writing in subsequent boxes). Indeed, over 50% of these responses following omissions were correct. In sum, subjects could omit one item before proceeding to recall the next one correctly, but in most situations, an omission signalled that the subject could not recall the rest of the list.

**Transpositions**

The monotonic increase in omissions with output position meant that the recency in overall errors must have arisen from a much stronger recency effect in transpositions. This was confirmed, with all conditions showing a strong, last-item recency effect (Figure 3-3). There was also mini-recency at the end of groups, particularly in condition G9.

Transpositions in conditions U9 and G9 were further split by transposition distance, collapsing across subjects (upper panel of Figure 3-4). Unlike the transpositions in Figure 2-3, there was no monotonic decrease in transpositions with increasing transposition distance, particularly for the grouped condition. The decrease was punctuated by peaks for three- and six-apart transpositions: These were (necessarily) interpositions. Objective grouping in condition G9 not only increased three- and six-apart transpositions, but also decreased other transpositions between groups, as well as one- and two-apart transpositions within groups (Table 3-2). Indeed, tests of weighted log-odds showed that objective grouping significantly increased the proportion of transpositions that were interpositions, $Z(18)=4.61$, $p<.0001$, yet significantly decreased both the proportion between groups that were not interpositions, $Z(18)=2.35$, $p<.05$, and the proportion within groups, $Z(18)=2.18$, $p<.05$.

| Condition | Within Groups | Between Groups | |
|---|---|---|---|
| | | Interpositions | Other |
| U9 | .38<br>(.09) | .27<br>(.09) | .39<br>(.09) |
| G9 | .33<br>(.09) | .39<br>(.10) | .30<br>(.09) |

Table 3-2: Proportion of transpositions within and between groups in Experiment 2. (Calculated from weighted log-odds.)

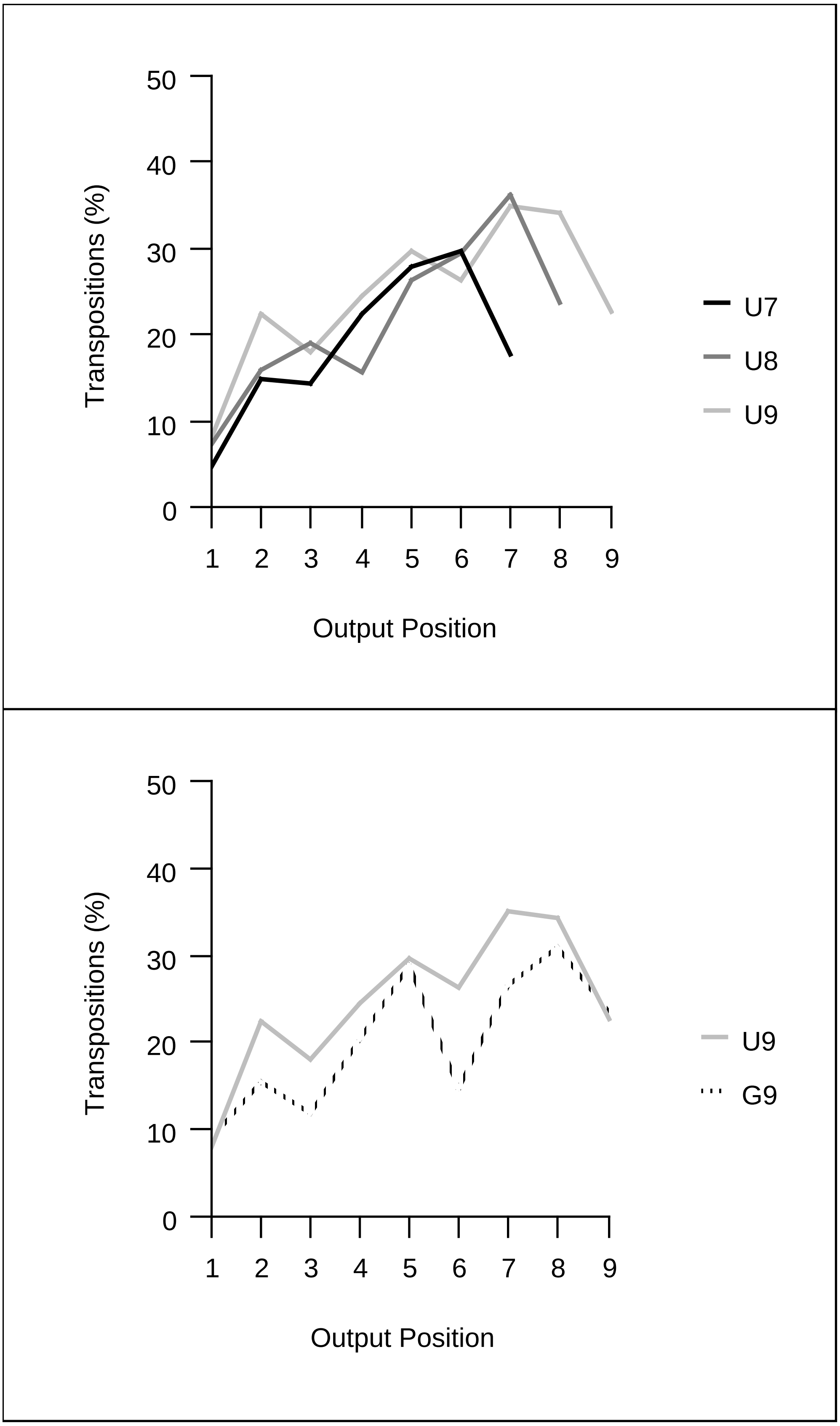



Figure 3-3: Transpositions by output position for ungrouped lists (upper panel) and nine-item lists (lower panel) in Experiment 2.

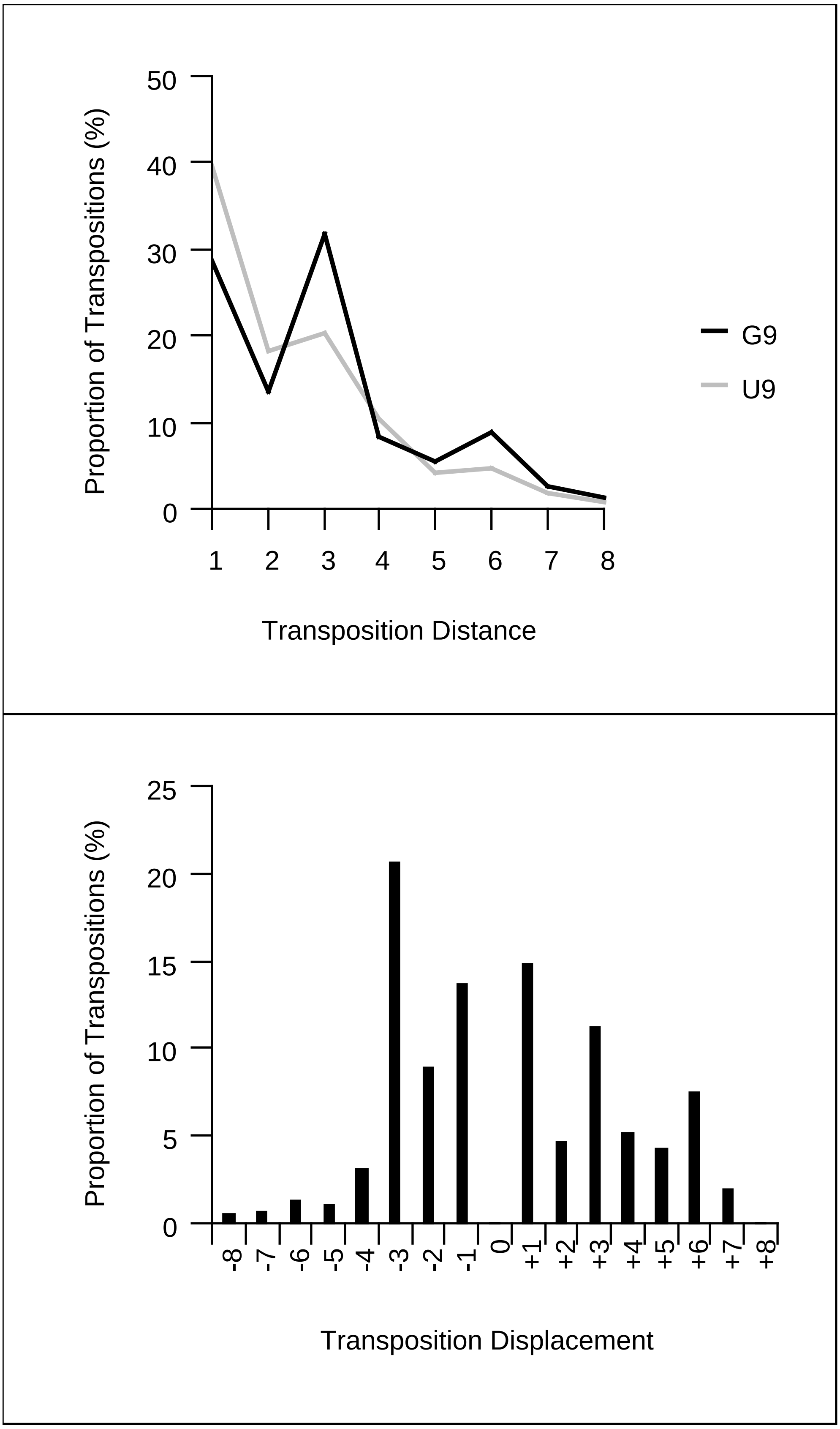



Figure 3-4: Proportion of transpositions by transposition distance in conditions G9 and U9 (upper panel), and by displacement in condition G9 (lower panel) in Experiment 2.

Further analysis of condition G9 distinguished transpositions moving forwards from transpositions moving backwards (lower panel of Figure 3-4; negative displacements represent items recalled too early, or *anticipations*; positive displacements represent items recalled too late, or *perseverations*). Most interpositions were anticipations from the immediately following group. Of the more remote six-apart interpositions between nonadjacent groups, most were perseverations (and often repetitions). The majority of interpositions were between the middle of groups ($M=.44$, $SD=.17$) rather than the start ($M=.33$, $SD=.16$) or end ($M=.28$, $SD=.16$) of groups, differences that were significant under weighted log-odds, $Z(17)>2.55$, $p<.05$ (excluding one subject who made no interpositions).

Approximately 40% of the responses following an interposition in condition G9 were correct, and 18% were associates (i.e., a further interposition). Thus interpositions were not the result of whole groups swapping (which would predict 66% of interpositions to be followed by associates). The majority of interpositions occurred singly, with the correct response often following in spite of the error. Nevertheless, some dependency between recall of groups and the items within groups was suggested by the fact that only 11% of interpositions were followed by associates in the condition U9, close to that expected by chance (and considerably greater than expected from the locality constraint; Chapter 1).

**Discussion**

The most important finding of Experiment 2 was that people will spontaneously group items presented at a constant rate without any instruction to that purpose, with a choice of group size dependent on list-length. This is hardly surprising in one sense, given that many everyday sequences such as telephone numbers are explicitly grouped, and yet it is surprising in the sense that spontaneous grouping is normally overlooked in psychological studies. Groups of three were most common, though a constant group size appeared preferable, in that lists of eight items were grouped in twos or fours. These “natural choices” of grouping strategy are those that are optimal when grouping strategy is imposed (Wickelgren, 1967).

Spontaneous grouping is rarely reported in studies of serial recall, and yet it is often evident in serial position curves (Madigan, 1980). This means that failures to find significant differences between objectively ungrouped and grouped conditions (e.g., Ryan’s tone pips) may be an artefact of subjective grouping in the "ungrouped" condition. In fact, the notion of an ungrouped, supraspan list may be a myth, and people have to resort to grouping in order to recall more than six or seven items in order (otherwise they fail completely). A need to group supraspan lists, but not subspan lists, may begin to explain some of the differences between these cases (Brooks & Watkins, 1990). The model in Chapter 5 provides a rationale for this grouping hypothesis, in terms of the limited resolution of positional codes. That spontaneous grouping is not always apparent in serial position curves may be an artefact of averaging over subjects using different grouping strategies. Moreover, serial position curves are not the best indicators of grouping, given that grouping can be evident in conditional analyses and inter-response times without necessarily being evident in serial position curves (Frankish, 1974).

The size of the grouping effect depends on factors such as method of grouping (Ryan, 1969a) and modality (Frankish, 1974). The objective, temporal grouping in the present experiment exerted effects beyond those of subjective grouping in the ungrouped condition, through either stronger or more consistent grouping across subjects. The effects of objective grouping were threefold: It 1) decreased transpositions within groups, 2) increased interpositions between groups, and 3) decreased all other transpositions between groups. The interpositions generally arose singly, without whole groups swapping (though not necessarily completely independently). Such interpositions are important because they imply that people encode the position of an item within a group. This supports positional models of serial recall.

In addition to the effects of grouping, the second aim of the present experiment was to examine the effects of list length on different error types. Intrusions and repetitions were rare. Omissions were more common, particularly for longer lists, where they were almost as common as transpositions. Omissions increased monotonically towards the end of recall, showing no recency effect. The recency effect in overall errors came from the marked reduction in transpositions on the last position. The fact that people can omit before going on to recall the next item is important for this pattern of transpositions (Chapter 4). When people are not given the option of omitting, serial position curves often fail to show any recency effects (e.g., Drewnowski & Murdock, 1980). This pattern of omissions and transpositions proves important for testing models of serial recall (Chapter 5).

## Experiment 3

The interpositions of the previous experiment demonstrated the existence of position-in-group information in recall of grouped lists. The present experiment examined the evidence for position-in-list information in ungrouped lists. This evidence came from position-specific intrusions between trials, that is, protrusions.

Conrad (1960) demonstrated protrusions in immediate serial recall of eight digits (though he called them serial order intrusions; Chapter 1). He showed erroneous responses of items that occurred at the same position in the previous list were more common than predicted by chance. He also found that increasing the intertrial interval decreased the incidence of such protrusions without affecting overall performance. He used the latter to argue that protrusions do not cause errors in recall, but arise after recall has already failed.

One way protrusions might have arisen in Conrad's experiment is through guessing strategies. When people forget an item and resort to guessing, they might tend to guess what they recalled last time. If the frequency of forgetting and hence guessing increases towards the end of recall (as might be expected from corresponding increases in omissions; Experiment 2), then later responses are likely to be guesses of the most recent items from the last trial (i.e., those from end of that trial). This would produce a greater incidence of protrusions on later positions than expected by chance, but only for short intertrial intervals (when people remember what they recalled last time). This is consistent with Drewnowski and Murdock's (1980) observation that intrusions are "overwhelmingly derived from the terminal location of the preceding list" (p. 329). Because protrusions are assumed to be guesses, this hypothesis is also consistent with Conrad's argument that protrusions do not play a causal role in forgetting.

The present experiment was an attempt to confirm Conrad's results and test the above guessing hypothesis. The present experiment had three important differences however:

1. Intertrial intervals of 2 and 20 seconds were used, rather than Conrad's intervals of 15, 25 and 40 seconds, in order to test a shorter intertrial interval. Also, Conrad did not report giving any instruction to subjects for the unfilled intertrial interval, whereas subjects in the present experiment were required to shadow a random sequence of digits between trials. This was to prevent subjects dwelling on (or even "rehearsing") previous lists.

2. Lists of five words were used, rather than Conrad's lists of eight digits. The use of such short lists was to reduce the need for subjects to group the lists subjectively (Experiment 2), a factor overlooked by Conrad. In order to produce significant numbers of errors for such short lists however, recall was delayed slightly: Subjects were required to shadow three further digits during the retention interval.

3. Lists were constructed such that no word appeared in two successive trials, unlike Conrad, who reused the same items on each trial. This was to ensure that any protrusions from previous lists were, by necessity, intrusions. The proportion of intrusions that were protrusions can therefore be compared to that expected by chance (one fifth), without needing to control for any artefactual correlations between item positions across lists.

It was hoped that these procedural differences would allow a clearer demonstration of proactive interference of position-in-list information.

## Method

### Subjects

Eighteen students from Cambridge University were tested, nine male and nine female, with a mean age of twenty-five years.

### Materials

Stimuli were lists of five phonologically nonconfusable, single-syllable, low-frequency words. All words contained five letters, had a Kucera-Francis frequency between four and five, and possessed both a unique vowel sound and a unique first letter. Half the lists had words drawn from the set: *yacht*, *goose*, *verve*, *psalm*, *wedge*, *haunt* and *clump*. The other half had words drawn from the set: *kneel*, *midge*, *latch*, *shine*, *bathe*, *flown* and *trout*. These two sets were alternated across trials, so that no word appeared in two successive trials. The order of words within lists was randomised with the constraint that, over all trials, each word appeared equally often at each of the five positions.

### Procedure

Each word was presented in the centre of a VDU, replacing the previous one, at a rate of one every second (500-ms on, 500-ms off). The fifth word was followed by a further sequence of three digits (drawn randomly without replacement from the set 1-9), presented at the same rate as the words. Subjects vocalised each word and digit as it appeared, but recalled only the words, by speaking them aloud in the same order that they saw them. Their responses were written down by the experimenter. Subjects were encouraged to guess if they were unsure, or to say “blank” if no word came to mind. After finishing recall, subjects vocalised a further two random digits (the Short condition) or twenty random digits (the Long condition), presented at the same rate of one a second, before pressing a key to commence the next trial.

Subjects received six practice trials, before being tested on a block of 28 lists for each condition (though the first two trials per block were not analysed in order to allow appreciable proactive interference to emerge). The order of conditions was counterbalanced across subjects. The whole experiment took about 40 minutes.

**Scoring Protrusions**

Two types of protrusions can be identified: intrusions of items at the same position in the previous list (*input protrusions*) and intrusions of items at the same position in the previous report (*output protrusions*). Given that recall on the previous trial may not be veridical, distributions of input and output protrusions can differ. These two types are compared below.

**Results**

In brief, overall performance was significantly worse in the Short than Long condition, mainly owing to an increase in intrusions, and the incidence of protrusions was significantly above chance in both conditions. These results are contrary to Conrad’s (1960). Furthermore, output protrusions were the most common intrusion over all positions. This result contradicts the guessing account of protrusions outlined above.

**Overall Performance**

The proportion of lists correct in the Short condition ($M=.31$, $SD=.25$) was less than in the Long condition ($M=.42$, $SD=.27$), a difference that was significant, $Z(18)=3.60$, $p<.001$.

**Error Types**

The main difference between the two conditions was a greater incidence of omissions and intrusions in the Short condition than Long condition (Table 3-3). Tests of weighted, log-

| Condition | Omissions | Intrusions | Transpositions |
|---|---|---|---|
| Short | .06<br>(.06) | .09<br>(.07) | .22<br>(.13) |
| Long | .04<br>(.05) | .05<br>(.07) | .22<br>(.12) |

Table 3-3: Frequency of omissions, intrusions and transpositions in Experiment 3.

odds showed these differences were significant in both cases, $Z(18)=2.67$, $p<.01$, and $Z(18)=4.76$, $p<.0001$, respectively. There was no significant difference in the incidence of transpositions, $Z(18)=0.06$, $p=.95$, and repetitions were negligible.

The greater error rate in the Short condition was spread mainly over middle positions (upper panel of Figure 3-5). There was no evidence of spontaneous grouping in these curves, or in transposition gradients. The distribution of omissions and transpositions (lower panel of Figure 3-5) was similar to that in Experiment 2. The distribution of intrusions showed a small recency effect, with most intrusions occurring on the penultimate position.

**Intrusions**

Extravocabulary intrusions were rare, and those that did occur were normally phonologically related words (e.g., *verge* for *verve, or shown* for *flown*). Of the nine possible intravocabulary intrusions on each trial, the majority were immediate intrusions of one of the five items in the previous list (*input intrusions*) or previous report (*output intrusions*).[3]

The frequency of immediate intrusions, and the proportion that were protrusions, was calculated for the 15 subjects who made at least one immediate intrusion (Table 3-4). A two-way ANOVA on the log-odds of an immediate intrusion showed a significant effect of condition, $F(1,42)=142.97$, $p<.001$, though no significant effect of scoring by input or output, or interaction, $F<1$. The effect of condition reflected a greater incidence of immediate intrusions in the Short than Long Condition. There was a trend for a greater proportion of output than input intrusions, but the lack of significance in this case was not surprising, because the two measures were highly correlated (given that most responses were correct).

3. A small proportion (24%) of output intrusions were also transpositions with respect to the current trial.

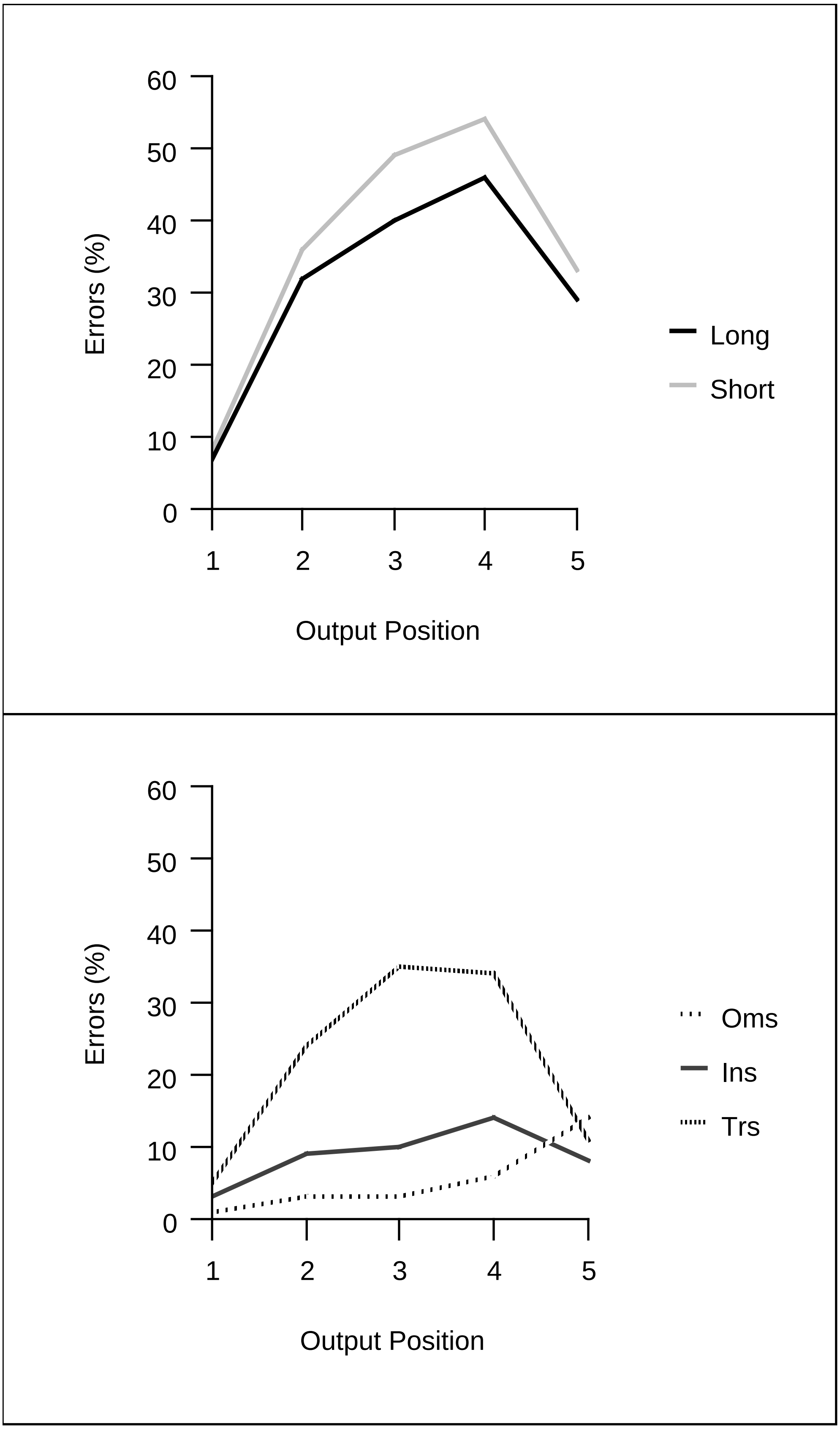



Figure 3-5: Errors by position for Long and Short conditions (upper panel), and error types by output position for the Short condition (lower panel) in Experiment 3. (Oms=omissions, Ins=intrusions, Trs=transpositions.)

| Condition | Immediate intrusions Input | Immediate intrusions Output | Protrusions Input | Protrusions Output |
|---|---|---|---|---|
| Short | .10<br>(.04) | .15<br>(.05) | .31<br>(.19) | .37<br>(.17) |
| Long | .07<br>(.04) | .09<br>(.05) | .32<br>(.28) | .38<br>(.30) |

Table 3-4: Frequency of immediate intrusions and proportion that were protrusions in Experiment 3.
(Calculated from weighted log-odds, *n=15*.)

A two-way ANOVA on the log-odds that an immediate intrusion was a protrusion showed no significant effects of condition, scoring, or interaction, $F(1,42)<1.97$, $p>.17$. Given that there were more immediate intrusions in the Short than Long condition, this implies that there were also more protrusions in the Short than Long condition, contrary to Conrad (as confirmed by an ANOVA on the proportion of responses that were protrusions, which showed a significant effect of condition, $F(1,51)=30.93$, $MSE=.052$, $p<.001$). Most importantly, the proportion of immediate intrusions that were protrusions was significantly above chance (.*20*) for both input, $Z(18)>2.82$, $p<.005$, and output protrusions, $Z(15)>4.38$, $p<.0001$.

Output intrusions were examined by position, collapsing across subjects and condition. The resulting intrusion gradients showed peaks for output positions that corresponded to the same output position in the previous report (i.e., protrusions; upper panel of Figure 3-6). In other words, an intrusion was likely to have come from the same or nearby position in the previous report. Though the number of output intrusions was greatest for middle positions, the proportion that were protrusions was greatest for the first position (lower panel of Figure 3-6).

Protrusions could have arisen because subjects mistakenly repeated their entire previous report in recall of the current list. This would result in a protrusion being scored on every position. However, further analysis shows such perseveration of whole sequences was rare: Only 7% of the output intrusions on Positions 1 to 4 were followed by a further output protrusion, whereas 34% were followed by a correct response. Thus protrusions, like interpositions, occur singly, and not simply from recalling the wrong list on the wrong trial.

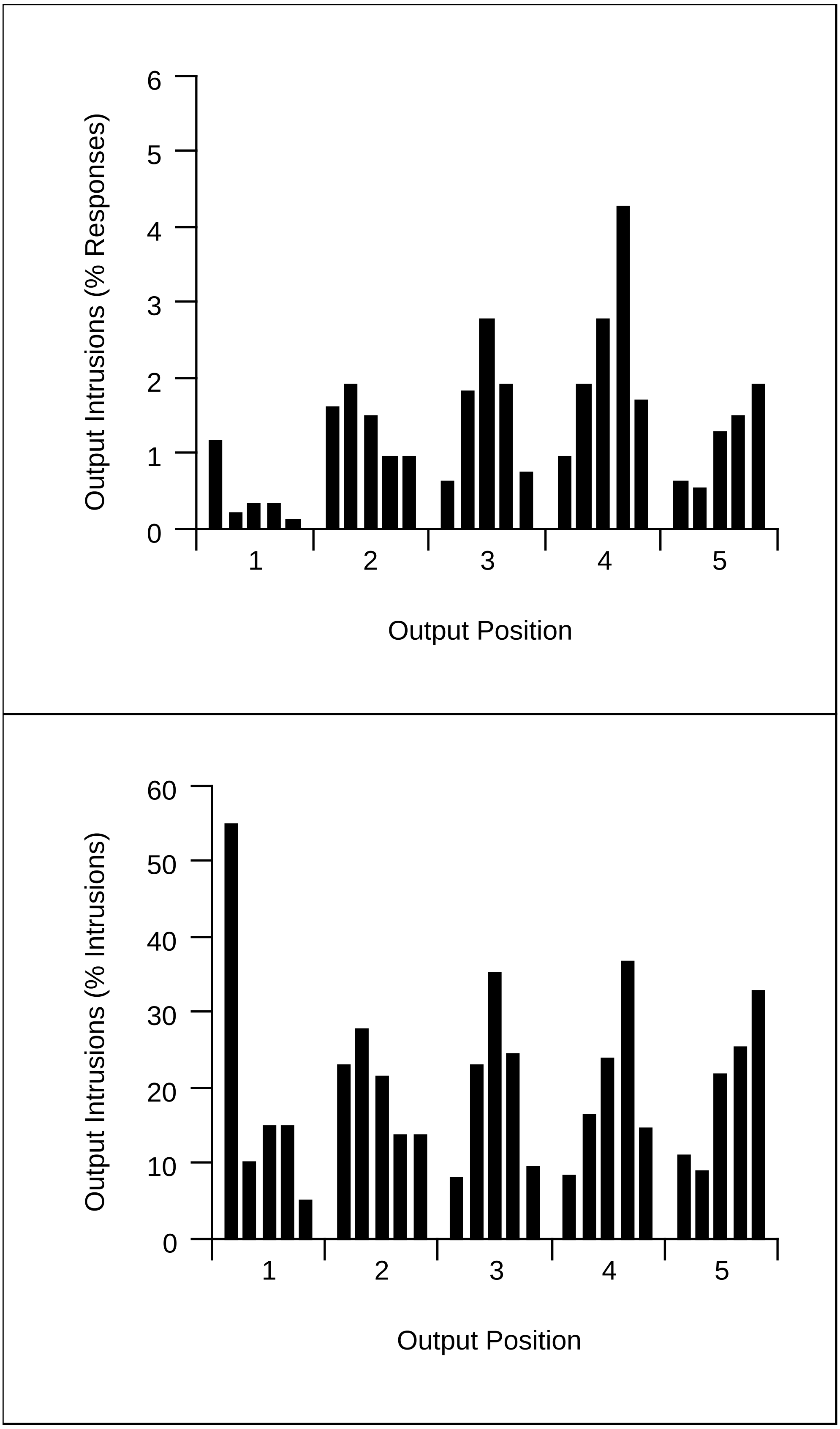



Figure 3-6: Output intrusions as a proportion of responses (upper panel) and as a proportion of intrusions per output position (lower panel) in Experiment 3.

**Discussion**

In agreement with Conrad (1960), the present experiment showed a significant incidence of protrusions, which decreased as the intertrial interval increased. Furthermore, protrusions were the most common intrusions across all positions in recall of ungrouped, five-item lists. The triangular-shaped intrusion gradients were therefore a graphic illustration of proactive interference of position-in-list information.

Unlike Conrad however, the greater number of protrusions with a short intertrial interval was mirrored by significantly worse overall performance. The most likely reason why Conrad failed to find such a difference in overall performance is that his shortest intertrial interval was 15 seconds, rather than the 2 seconds of the present experiment. If the amount of proactive interference is inversely related to the intertrial interval (e.g., Turvey, Brick & Osborn, 1970), the present experiment is a more powerful test of the effects of intertrial interval, by virtue of a greater range of proactive interference.

Given that the poorer performance in the short intertrial interval owed mainly to a greater number of intrusions (with a small increase in omissions, but hardly any change in transpositions), it seems reasonable to conclude, contrary to Conrad, that proactive interference does play a causal role in forgetting in short-term serial recall. A similar conclusion was reached by Sanders and Willemsen (1978b). In particular, one noncausal explanation for protrusions was ruled out by present results: the hypothesis that protrusions are an artefact of a guessing strategy. This hypothesis predicts that protrusions should be confined towards later positions, when recall falters, and yet protrusions were found across all positions in the present experiment (and contrary to Drewnowski & Murdock, 1980, intrusions were not overwhelmingly from the last position of the previous trial). Further evidence that protrusions are not simply guesses is given in Experiment 5. More likely, protrusions result from response competition at each position of recall (Chapter 5).

Also unlike Conrad's data, the proportion of immediate intrusions that were protrusions was still significantly above chance after 20 seconds between trials. One reason may be that Conrad employed immediate serial recall, rather than delaying recall by 3 seconds as in the present experiment. Longer retention intervals will tend to increase proactive interference (e.g., Crowder, 1993). Another reason may be because Conrad's design meant he could not distinguish intrusions from transpositions, making classification of protrusions uncertain. Even longer intertrial intervals therefore, such as Conrad's 40 second delay, may be required before protrusions fall to chance levels. As such, the proactive interference in the present experiment demonstrates a surprising longevity of short-term memory for positional information. Indeed, Nairne found evidence for positional information after two minutes of distraction following incidental learning (Nairne, 1991).

Conrad only measured input protrusions. However, the present experiment showed that output protrusions were more common than input protrusions (a trend confirmed in Chapter 4). In other words, if an item is recalled in the wrong position on one trial, it appears more likely to protrude on the next trial in its position of recall rather than its position of presentation. This suggests that recall is itself a learning episode, such that, in the case of errors, an item is relearned in a different position. The previous report may be a greater source of proactive interference because it represents a more recent learning episode than the previous list. Alternatively, recall may be a stronger learning episode than presentation, as suggested by the fact that the Hebb effect requires multiple recall attempts; multiple vocalised presentations are insufficient (Cunningham, Healy & Williams, 1984). Another possibility is that responses enter a separate rehearsal store, which becomes a secondary source of interference between trials (Estes, 1991). The questions of proactive interference and recoding during recall are covered more fully in Chapter 6.

Finally, the intrusion gradients in Figure 3-6 demonstrated that intrusions were more common in the middle than the start or end of reports. However, the proportion that were protrusions was greatest at the start, suggesting that the first position is coded more precisely. In other words, proportional intrusion gradients give an idea of the *positional uncertainty* associated with each position: Shallower gradients indicate greater positional uncertainty. Nevertheless, intrusion gradients are not a perfect reflection of positional information in short-term memory. There are several reasons why the positional information used in serial recall may be considerably more precise (giving the sharper transposition gradients in Figure 2-2). Firstly, intrusion gradients necessarily index positional information from the previous trial, which is likely to become less accurate over time. Secondly, there may be several sources of proactive interference, such as that from even earlier trials, which will introduce additional noise to the extent that the sources are uncorrelated. Thirdly, there are extraneous reasons for intrusions, such as people's predisposition to guess certain words. (One subject for example recalled the word "shine" on nearly every trial.) This additional noise will blur intrusion gradients even further. These points are relevant to the question of whether positional information is sufficient to underlie serial recall (below).

## General Discussion

The present experiments demonstrated two types of positional errors: interpositions between groups that maintain their position within group (Experiment 2), and protrusions between trials that maintain their position within trials (Experiment 3). These errors are evidence of positional information in short-term memory, supporting positional theory. They cannot be explained by ordinal or chaining theory.

The intrusion gradients in Experiment 3 resemble the transposition gradients in Experiment 1. As Page and Norris (1996b) demonstrate however, transposition gradients do not, on their own, necessitate a positional model. Their ordinal model produced similar transposition gradients for each position. This is because errors in the relative order of nearby items also produce peaked transposition gradients. However, the fact that intrusions show peaked gradients does necessitate a positional model. This is because intrusions that maintain positions between trials cannot be attributed to errors of relative order within a trial (Chapter 4). Protrusions indicate that items are coded for position independently of surrounding items. A similar argument applies to interpositions between groups.

Page and Norris (1996b) made several arguments against interpreting positional errors as evidence for positional models of immediate serial recall. Firstly, they argued that positional information might be limited and therefore insufficient to support serial recall. The limitation of positional information was based on the argument that group sizes of three are optimal, in which position can be characterised as *start*, *middle*, and *end* (Wickelgren, 1967). These codes only require specification of the first and last item of each group, since the middle item can be defined by exclusion. Such codes are sufficient to explain the interpositions in

Experiment 2. For larger groups of items however, the codes *start*, *middle* and *end* would not be sufficient to order nonterminal items, suggesting that positional coding is limited to three positions at the most. This suggestion is refuted by the intrusion gradients in Experiment 3. These five peaked gradients demonstrate that positional information extends beyond start, middle and end. Subjects in Experiment 3 must have possessed at least five positional codes, perhaps even *first*, *second*, *third*, *fourth*, and *fifth* (though Chapter 6 argues for a somewhat different representation of position).

This prompts the question of whether subjects in Experiment 2 possessed codes for “first” through to “ninth” in the nine-item, ungrouped condition. This is possible, but the fact that they preferred to spontaneously group such lists suggests that there may well be a limit to the number of positions people can distinguish, as Page and Norris suggest. By breaking a list into groups, a smaller range of positional codes may suffice, by using *first*, *second* and *third* to code both an item’s position within a group, and the group’s position within the list. This is illustrated in Chapter 5. The model developed in that chapter has a limited range of positional coding, yet one that is more fine-grained than the *start*, *middle* and *end* of Page and Norris. Moreover, the model demonstrates that this information is sufficient to support serial recall.

Another argument offered by Page and Norris is that positional errors have typically been demonstrated with delayed rather than immediate serial recall. The phonological similarity effect, which Page and Norris use as a signature of their model of the phonological loop, decreases as the recall delay increases (Baddeley, 1986). A corresponding increase in positional errors would produce a double dissociation that might suggest two different sources underlying serial recall, an ordinal (phonological) one and a positional (nonphonological) one. However, while it is true that delayed recall was employed in Experiment 3 (to increase overall error rates), the meta-analyses in Chapter 4 reveal that positional errors also arise in immediate serial recall of span-length lists. Moreover, the model developed in Chapter 5 explains the trade-off between positional and phonological errors without appealing to two different theories of serial order. The increase in positional errors with delay is attributed to a ratio-rule of proactive interference (e.g., Crowder, 1993), applying to positional information, and the decrease in phonological errors is attributed to rapidly-decaying, phonological traces (e.g., Tehan & Humphreys, 1995), constituting only supplementary item information.

A third argument by Page and Norris is that positional errors are epiphenomenal rather than causal. This might be suggested by the rarity of positional errors like protrusions (Experiment 3). There are several counterarguments. Firstly, interpositions are a far more common example of positional error. Indeed, they were more common than adjacent transpositions in Experiment 2. Secondly, Conrad's belief in a noncausal role of protrusions was contradicted by Experiment 3, which did suggest a causal role. Finally, positional errors are not restricted to guesses (Experiments 4 and 5). Indeed, anecdotal evidence suggests that people often make interpositions without even being aware of having made an error.

A final argument offered by Page and Norris was whether positional errors are an artefact of subsidiary processes in serial recall. For example, are they an artefact of subjects copying down responses from previous trials during written recall? This cannot be true of the spoken recall in Experiment 3. Are they an artefact of output processes operating in a speech buffer, like those assumed to underlie speech errors (Chapter 8)? This seems unlikely in the case of protrusions in the Long condition of Experiment 3, which remained above chance even after at least 28 seconds of shadowing digits and letters between the termination of recall in one trial and the initiation of recall in the next. (The speech buffer account also has difficulty explaining why interpositions remain under articulatory suppression, Page & Norris, 1996a, and why protrusions do not increase with degree of vocalisation, Murray, 1965.) Are they an artefact of visuospatial strategies such as imagery? This also seems unlikely, given that there was no objective spatial information in the sequential presentation of Experiments 2 and 3 (or in the spoken recall of Experiment 3), and given that imagery was rarely reported during debriefing. Moreover, others have argued that the role of visuospatial information in such experiments is minimal (Hitch & Morton, 1975).

Thus none of the arguments offered by Page and Norris (1996b) appears to hold in the light of Experiments 2 and 3. These experiments demonstrate positional information that extends over at least five positions and plays an important, causal role in serial recall from short-term memory. The exact nature of this information is deferred to Chapters 5, 6 and 8.

## Chapter Summary

This chapter described two experiments providing evidence for positional errors in serial recall from short-term memory. These errors are explicable by positional theory, but not chaining or ordinal theories. Nevertheless, Chapter 4 continues to entertain all three theories and tests their predictions in a more detailed analysis of error distributions. This takes the form of meta-analyses over a number of experiments, complementing and confirming the analyses in Experiments 1-3, and providing a fuller set of constraints for the model in Chapter 5.

# Chapter 4: Meta-analyses of Errors

## *Empirical Constraints on Error Distributions*

This chapter describes three meta-analyses of a number of recent, unpublished studies on short-term serial recall conducted at the Applied Psychology Unit. These analyses were to test the generality of the results of Experiments 1-3, and also to test more detailed predictions of three specific models of serial recall, a chaining model (Murdock, 1995), a positional model (Burgess & Hitch, 1992) and an ordinal model (Page & Norris, 1996b). Meta-analyses over several different experiments were necessary because some types of errors are rare (e.g., repetitions and protrusions), hampering statistical tests within any one experiment. The results are summarised in a set of *empirical constraints* on error distributions in serial recall.

### Three Models of Serial Recall

In order to guide some aspects of the meta-analyses, three specific models of serial recall were considered, each of which exemplified one of the general theories of serial order in Chapter 1. The first was a closed-loop, compound chaining model, based on the Power Set Model of Murdock (1995). Though Murdock did not specify the precise nature of the closed-loop chaining (i.e., whether or not errors are fed back as cues), such a closed-loop model has been analysed independently by Henson (1994). The second model was a positional model based on the Articulatory Loop Model of Burgess and Hitch (1992). This model uses a context signal to cue each position, such that cues for nearby positions overlap in symmetrical manner (Chapter 1; decay processes in this model were ignored for simplicity). The third model was an ordinal model based on the Primacy Model of Page and Norris (1996b). This assumes a primacy gradient of activations, invariant across positions (decay was again ignored).

#### Competition Space

Though the above models differ in many respects, they can be compared using the abstract notion of a *competition space*. Competition space indicates the strength with which each item competes for each response during serial recall. The competition space for the first three responses in serial recall of five items is shown in Figure 4-1. The filled bars represent the strength with which each item (from left to right) competes for the first, second and third response (in each column). Assuming some random noise in these strengths, the height of each bar relative to the others relates to the probability of recalling that item at that position (i.e., the random noise can sometimes cause errors).

The broken bars represent items that have been recalled. Thus Figure 4-1 illustrates competition space at the start of recall (leftmost column), after Item 2 has been recalled erroneously in Position 1 (middle column), and after Item 3 has been recalled erroneously in Position 2 (rightmost column). Once recalled, an item is suppressed. This suppression reduces the probability of recalling it again, explaining why repetitions are rare (below). Suppression also explains the interdependency between responses (Henson et al., 1996), such that the probability of recalling an item depends on what has been recalled previously (i.e., items are selected without replacement). All three models above assume a process of suppression (implicitly in Murdock, 1995; explicitly in Burgess & Hitch, 1992, and Page & Norris, 1995).

In competition for the first response (leftmost column), it is assumed all three models are equivalent.[1] In other words, all models predict the first item as the most likely response, the second item as the next most likely, etc. The first difference between the three models arises when the first error occurs, where Item 2 is recalled in Position 1 (middle column). The models differ in their predictions as to what should follow this error. The Power Set Model predicts the most likely next response is Item 3, because it will be cued strongly by previous recall of its associate, Item 2. The Articulatory Loop Model predicts that Item 1 and Item 3 will be equally likely to follow, because the cue for Position 2 overlaps equally with those for Position 1 and Position 3 (e.g., Figure 1-2 in Chapter 1). The Primacy Model predicts that Item 1 will be most likely to follow, because it remains the strongest competitor.

The prediction of the Primacy Model, that Item 1 will follow the error on Position 1, was termed *fill-in* by Page and Norris (1996b). In more general terms, fill-in is a property such that “when an item is missed out in recall, due to a transposition, it is liable to be recalled in the next position” (Page & Norris, 1996b, p. 8). Fill-in is important in preventing a cascade of further errors. This is evident by considering the situation where Item 2 is followed by a

1. An additional start-of-list context is assumed in order to cue the first item in the Power Set Model.

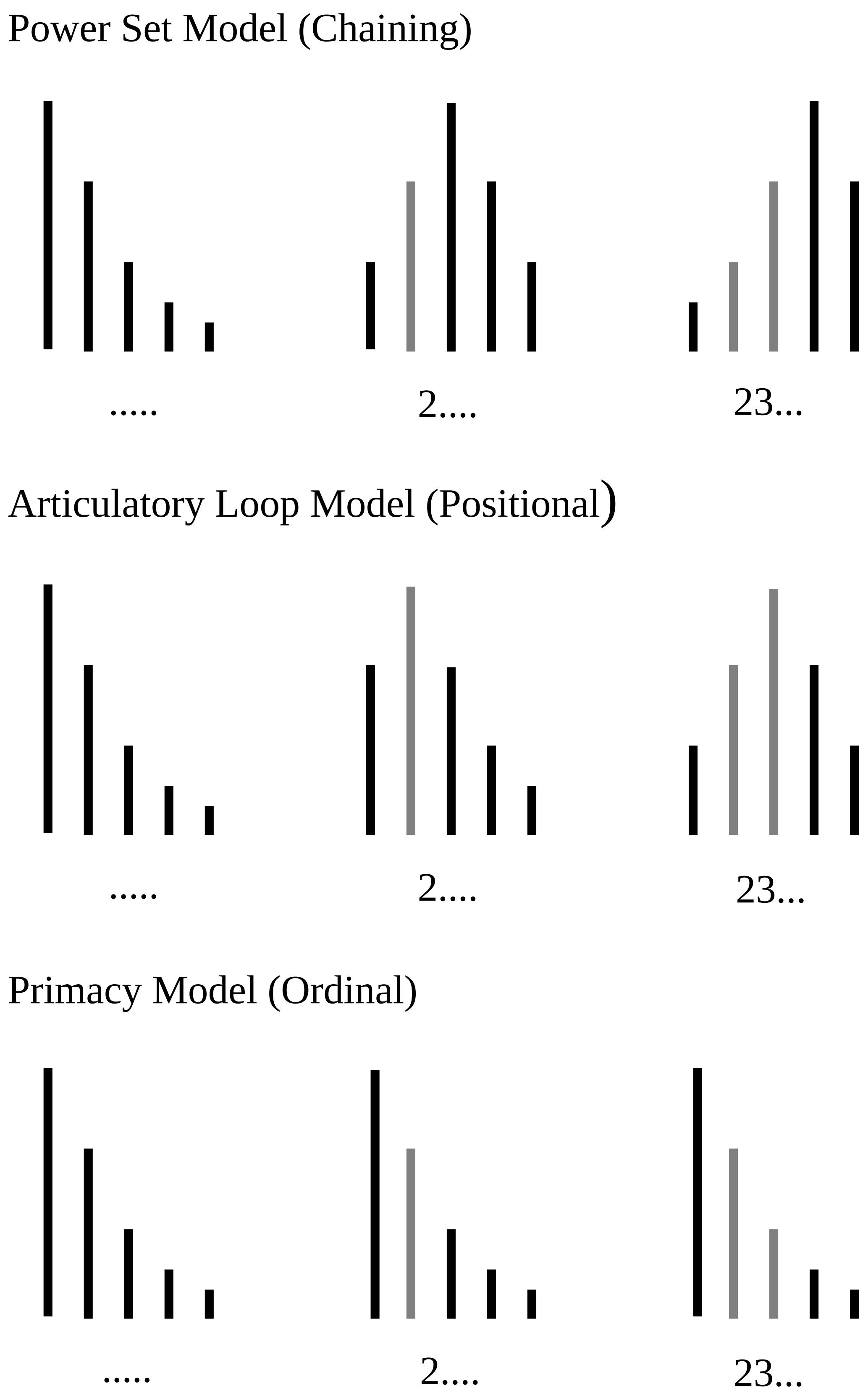



Figure 4-1: Competition space within each model for the first three responses to a list *12345* recalled as *23...*, illustrating absence of weak fill-in.

further error of Item 3 (rightmost column). The Power Set Model predicts the most likely next response is Item 4, because it will be cued strongly by previous recall of its associates Item 2 and Item 3. In other words, the Power Set Model predicts a further slippage of items. The Articulatory Loop Model also predicts Item 4 will follow, because the cue for Position 3 overlaps more with the cue for Position 4 than the cue for Position 1. Only the Primacy Model predicts that the most likely next response remains Item 1, to “fill-in” the gap and prevent further slippage. In other words, only the Primacy Model predicts that the probability of fill-in increases with further errors; the other models predict that the probability of fill-in decreases, such that the last item is unlikely to be recalled until the end, when all others have been recalled and suppressed. Further consideration reveals that the lack of fill-in in the Power Set and Articulatory Loop Models is why neither produce sufficient recency (Henson et al., 1996).

On the other hand, consider the situation in Figure 4-2, where the first two items have been recalled correctly (leftmost column). The middle column then shows the competition space after Item 5 is recalled erroneously in Position 3. The Power Set and Articulatory Loop Models predict the next most likely response is the correct response, Item 4, whereas the Primacy Model predicts fill-in of Item 3.

This example illustrates the distinction between strong fill-in and weak fill-in. The Primacy Model shows strong fill-in, in that the earliest unrecalled item will always be the most likely response following an error. The Power Set and Articulatory Loop Models do not predict strong fill-in. The Articulatory Loop Model in particular predicts that the correct response is always most likely following an error (providing the correct item has not already been recalled and suppressed, as in Figure 4-1). Only in situations where the correct item has already been recalled can (weak) fill-in can occur, as in the rightmost column of Figure 4-2. In this case, Item 4 is recalled correctly in the fourth position, and all three models predict that Item 3 will finally fill-in.

The distinction between strong and weak fill-in is important because a model can show weak fill-in without showing strong fill-in. Though this is not true of the Articulatory Loop Model, because of the symmetrical nature of its positional cue (a situation actually made worse once decay is added, Henson et al., 1996), it is true of the new positional model

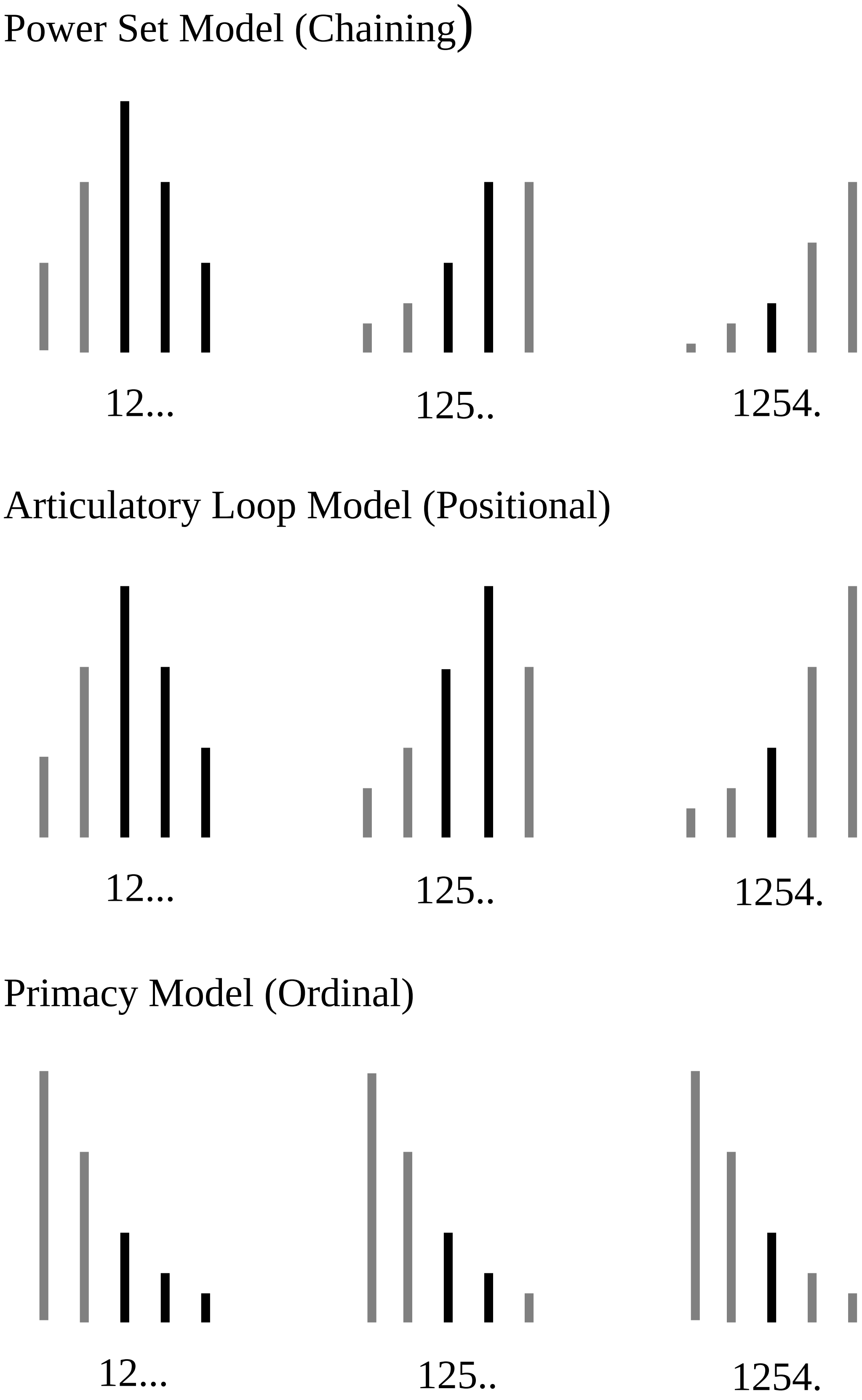



Figure 4-2: Competition space within each model for the last three responses to a list *12345* recalled as *1254.*, illustrating absence of strong fill-in.

developed in Chapter 5. This model assumes asymmetrical positional cues, biased towards earlier items. The strength of fill-in was one of the questions asked in the meta-analysis below.

## Meta-analysis 1

In total, 37 conditions from 14 different experiments were analysed, using a computer program developed by the author. These experiments all employed immediate serial recall of objectively ungrouped lists of phonologically dissimilar, nonrepeated items. The conditions differed in list length (from five to nine items), nature of items (digits, letters or words), presentation rates (between one and two items per second), presentation modality (visual, vocalised, or auditory) and recall method (written or spoken). Further details of the conditions are given in Appendix 2. A number of pairwise, binomial sign-tests were performed across the conditions, accompanied by 95% confidence intervals (*CI*) for the median value.

### Primacy Constraint

Primacy was evident in all error position curves in Experiments 1-3. Its prevalence was tested further by comparing the frequency of errors on the first two positions across all 37 conditions. Errors on the first position were less frequent than on the second position in all cases, $N=37$, $p<.001$, $CI=(.11,.15)$, reinforcing the ubiquity of primacy in serial recall.

### Recency Constraint

Recency was also evident in Experiments 1-3, though it was weaker than primacy, and confined to the last one or two positions. Last-item recency was tested by comparing the frequency of errors on the last two positions. The frequency on the last position was less than on the penultimate position in only 20 conditions (and equal in 4 conditions), suggesting that recency is not a reliable effect in serial recall, $N=33$, $p=.15$, $CI=(-.02,.04)$.

However, of the 13 conditions with no last-item recency, 12 employed lists of words, and 7 of these used five-syllable words. These conditions showed large increases in omissions towards the end of recall (Experiment 2; below). When the 10 conditions with words of more than one syllable were excluded from analysis, the presence of last-item recency was reliable, arising in 19 conditions (and equal in 2 conditions), $N=25$, $p<.01$, $CI=(.01,.10)$. This suggests that recency is normally found, except when there are large numbers of omissions.

### Locality Constraint

The locality constraint, that items transpose small distances about their correct positions, was introduced in Experiment 1. The generality of this constraint was tested by comparing the frequency of one-and two-apart transpositions, weighted by the opportunity for such transpositions.[2] One-aparts were more frequent than two-aparts in all conditions, $N=37$, $p<.001$, $CI=(.03,.04)$, demonstrating the fundamental nature of the locality constraint.

### Fill-in Constraint

It is possible for data (and models) to meet the locality constraint without meeting the fill-in constraint (above). For example, a sequence *12345* recalled as *13452* contains three one-apart transpositions, and one three-apart transposition. Though the ratio of these transpositions would meet the locality constraint, this example violates the fill-in constraint because Item 2 was not recalled immediately after it was replaced by Item 3.

To measure fill-in, analysis was confined to responses following the first error in a report. To illustrate the nature of such responses, data from the ungrouped conditions of Experiment 2 were collapsed across subjects (Table 4-1). Of the 207 responses following a first error of Item $i+1$ on Position $i$ (as in Figure 4-1), the majority were the fill-in errors of Item $i$ predicted by the Primacy Model, and only half as many were the associate errors of Item $i+2$ predicted by the Power Set Model (top row of Table 4-1). In other words, when $i=1$, an incorrect report of *12345* is more likely to be *21345* than *23145* (contrary to Figure 4-1). Thus, there was evidence for fill-in. There were hardly any immediate repetitions of the correct Item $i+1$, but this is attributable to the suppression of items already recalled.

| | Following Response | | | |
|---|---|---|---|---|
| First Error | Fill-in (Item $i$) | Correct (Item $i+1$) | Associate (Item $i+2$) | Other |
| Item $i+1$ | .53 | .01 | .21 | .25 |
| Item $j>i+1$ | .25 | .48 | .08 | .19 |

Table 4-1: Proportion of responses following a first error on Position $i$ in Experiment 2.

2. Given transposition distances of $i$ and $j$ $(i<j)$, this weighting means scaling the number of $j$-apart transpositions by a factor $(n-i)/(n-j)$, where $n$ is the list length, reflecting the fewer opportunities for transpositions further apart.

To measure the strength of fill-in, analysis was confined to responses following a first error of Item *j* on Position *i*, where Item *j* was an item other than Item *i* or Item *i+1* (as in Figure 4-2). Of the 336 such responses, the majority were the correct responses of Item *i+1* predicted by the Articulatory Loop Model, and only half as many were the fill-in errors predicted by the Primacy Model (bottom row of Table 4-1). In other words, when $i=3$, incorrect report of *12345* is more likely to be *12543* than *12534* (as in Figure 4-2). Thus, there was no evidence for strong fill-in.

To test the generality of this conclusion, similar calculations were performed in the meta-analysis. With a first error of Item *i+1*, the proportion of following responses that were fill-in errors was greater than the proportion that were associate errors in 35 conditions (equal in 2 conditions), demonstrating highly reliable weak fill-in, $N=35$, $p<.001$, $CI=(.20,.32)$. With a first error of Item *j*, where $j>i+1$, the proportion of following responses that were correct was greater than the proportion that were fill-in errors in 33 conditions (equal in 1 condition), $N=36$, $p<.001$, $CI=(.16,.22)$, demonstrating that strong fill-in is the exception rather than the rule. Taken together, these analyses confirm that the fill-in is stronger than predicted by the Articulatory Loop Model, but not as strong as predicted by the Primacy Model.

One caveat applies to the above analysis. Many of the lists are likely to be spontaneously grouped (Chapter 3). The influence of grouping may confound the analysis, perhaps reducing the strength of fill-in, given that interpositions tend to be followed by correct responses (Experiment 2). The fact remains however that the interpositions themselves, or indeed any type of positional error, cannot be explained by models with strong fill-in (below).

**Omission Constraint**

The omissions in Experiments 2 and 3 increased towards the end of recall. To test the reliability of this finding, the frequency of omissions on the last two output positions was compared. The frequency was greater on the last position than penultimate position in 32 conditions (equal in 3), confirming the reliability of the finding, $N=34$, $p<.001$, $CI=(.04,.17)$.

The fact that omissions increase towards the end of recall might suggest that the last item is omitted more often than any other. Indeed, this is what is predicted by the Primacy Model (Page & Norris, 1996b). To illustrate this, the upper panel of Figure 4-3 shows the

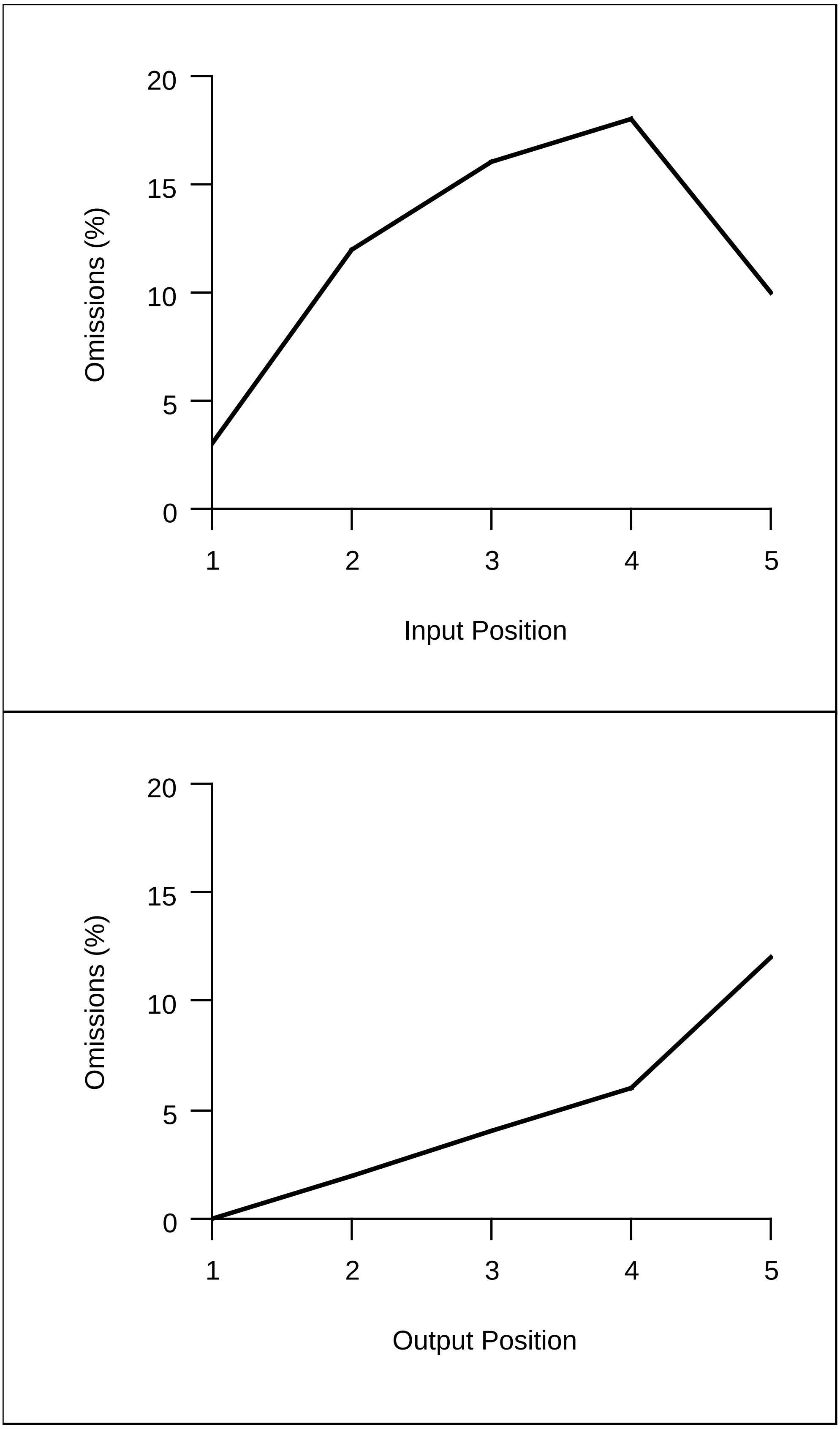



Figure 4-3: Omissions by input position (upper panel) and by output position (lower panel) averaged across both conditions of Experiment 3.

frequency of omissions against input position averaged over both conditions of Experiment 3. The increase in omissions with output position (lower panel) was not paralleled by a similar increase with input position: The last item was more often recalled somewhere than the penultimate item. This pattern of results can be explained if the last item is sometimes recalled too early, replacing the penultimate item, and followed by an omission. To test whether the pattern was an exception rather than the rule, the meta-analysis compared the frequency of omissions on the last input position with the frequency of omissions on the penultimate input position. The frequency on the last position was greater than on the penultimate position in 17 conditions (and equal in 3 conditions), $N=34$, $p=.57$, $CI=(-.02,.03)$. This unreliable difference indicates that the increase in omission towards the end of recall does not always reflect failure to recall the last item. As well as being troublesome for the Primacy Model, this pattern of item errors contrasts with the flat distribution assumed by Lee and Estes (1977, 1981). This is probably because they, like Healy (1974), did not consider lists of more than four items.

**Repetition Constraint**

Repetitions in Experiments 1-3 were rare. However, their distribution was highly constrained: They were always widely separated in reports, with the majority being items recalled at the start of recall that were recalled again towards the end of recall. In condition PN of Experiment 1 for example, repetitions comprised approximately 2% of responses (11% of errors) and the two occurrences were, on average, 3.34 positions apart in reports. The most common repetition was of the first item, recalled correctly on Position 1 and again incorrectly on Position 6 (hence the exception to the locality constraint for Position 6 of this condition in Experiment 1). This pattern is shown in Figure 4-4 (the peak on the fourth input position probably reflects the effect of the 3-3, subjective grouping in Experiment 1).

The significance of this distribution of repetitions can be illustrated by a simple guessing model. According to this model, subjects who fail to recall an item correctly guess at random from the set of list items. Simulations of such a simple model, fitted to overall error rates in condition PN, produced repetitions that comprised 16% of responses (84% of errors), far in excess of the data. Simulations also gave a mean distance between two occurrences of an item of 2.21 positions, considerably smaller than in the data. Though a different frequency of

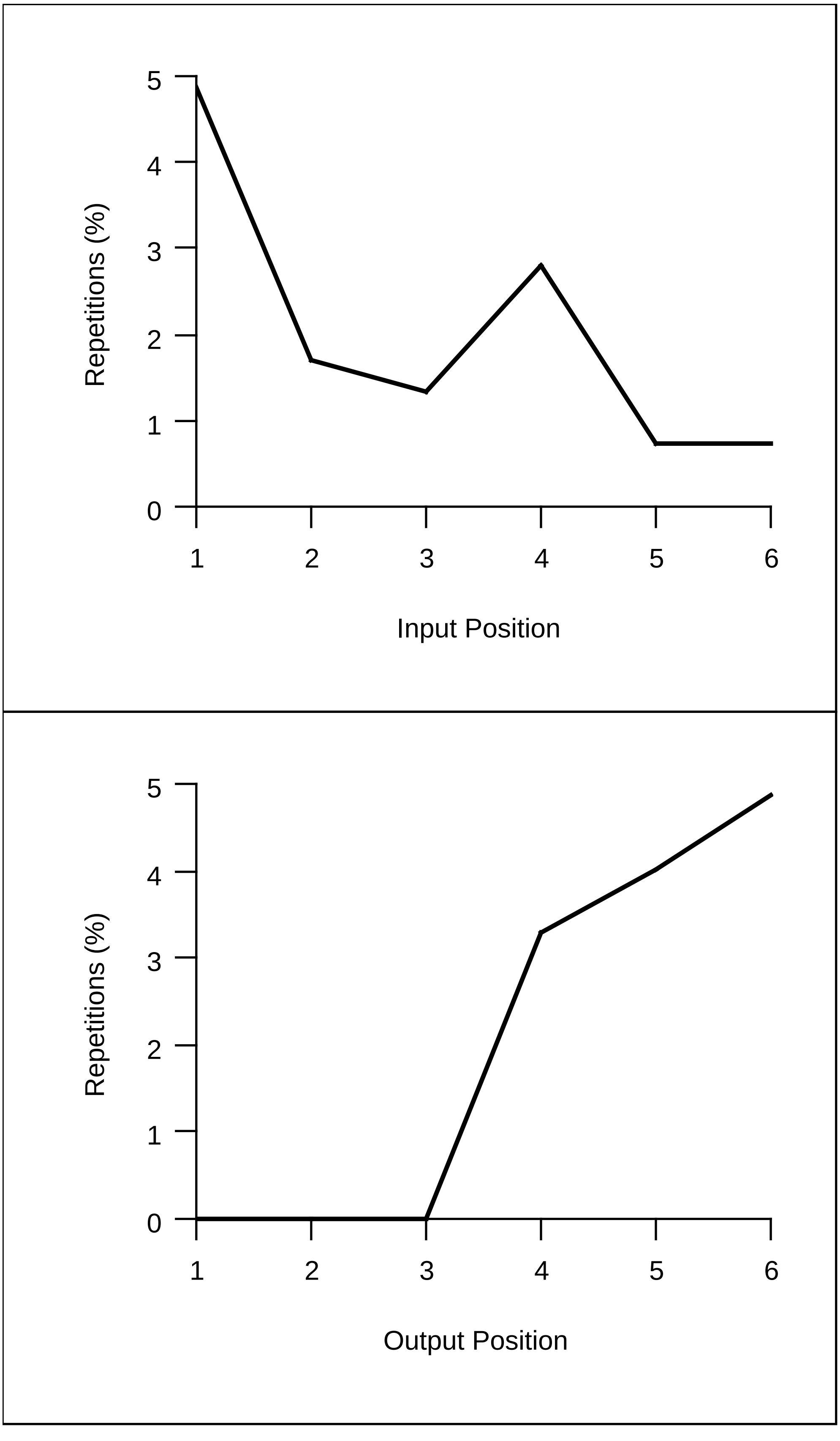



Figure 4-4: Repetitions by input position (upper panel) and by output position (lower panel) for PN condition of Experiment 1.

repetitions would result if subjects' guesses were biased towards neighbouring list items, as required by the locality constraint, this would produce an even smaller mean distance between the two occurrences of a repeated item. This example suggests that responses in serial recall are normally chosen without replacement, supporting the idea of response suppression described above. Nonetheless, the fact that repetitions do sometimes occur suggests that response suppression is not perfect; it probably wears off over time (Chapter 5).

The hypothesis that most repetitions involve early list items was examined by testing whether over 50% of repetitions were from the first two input positions. This was true of 24 conditions in the meta-analysis, confirming the hypothesis, *N=37, p<.05, CI=(.01,.19)*. The hypothesis that most repetitions occur towards the end of recall was examined by testing whether over 50% of repetitions occurred on the last two output positions. This was true of 26 conditions (equal in 1 condition), confirming this hypothesis too, *N=36, p<.01, CI=(.02,.17)*.

**Protrusion Constraint**

The protrusions measured in Experiment 3 were also rare. Nonetheless, they represented a significant proportion of immediate intrusions; a proportion greater than expected by chance. To test whether this was true more generally, the proportion of erroneous items that occurred at the same position in the previous report, given that they occurred somewhere in that report, was compared with that expected by chance (which is $1/n$, where $n$ is the list length). This proportion was above chance in 35 conditions (and equal in 1 condition), demonstrating that output protrusions are a reliable finding, *N=36, p<.001, CI=(.06,.10)*. Moreover, this proportion was greater than the corresponding proportion for input protrusions in 28 conditions (and equal in 3 conditions), *N=34, p<.001, CI=(.01,.03)* supporting the suggestion in Experiment 3 that output protrusions are a better index of positional information. Finally, the proportion of output protrusions followed by a correct response was greater than the proportion followed by a further protrusion in 35 conditions, *N=37, p<.0001, CI=(.11,.20)*, supporting the conclusion of Experiment 3 that protrusions normally arise singly, without intrusion of whole subsequences.

The theoretical importance of positional errors like protrusions can also be illustrated in competition space. Figure 4-5 shows the competition space in recall of the second of five

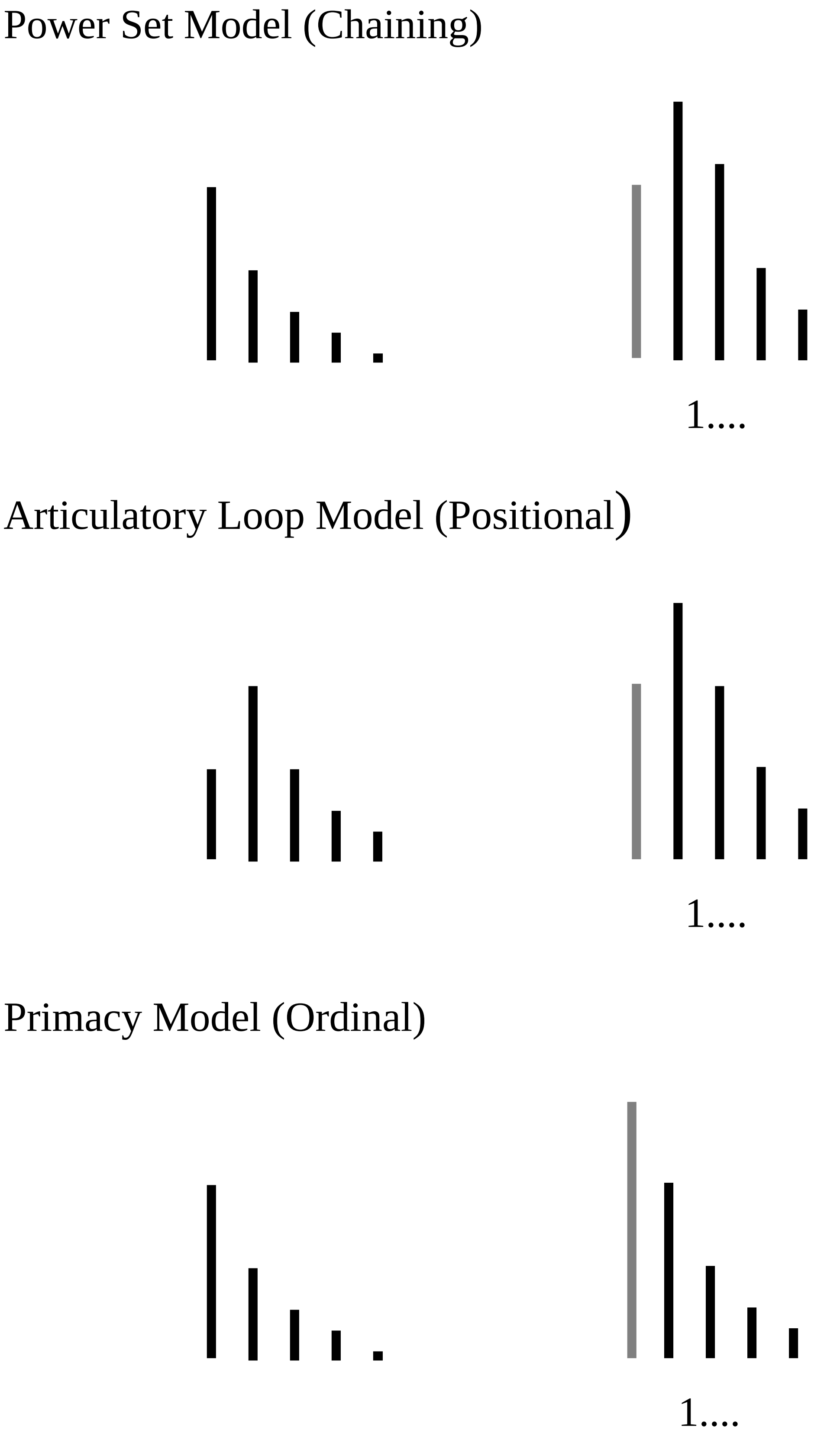



Figure 4-5: Competition space within each model for the second response to a list *12345* recalled as *1....*, illustrating competition from items in the previous trial.

items (rightmost column), including competition from items in the previous trial (leftmost column), assuming suppression for the previous trial has worn off. In the Power Set Model, the cue for the second item bears no necessary resemblence to the cue for the second item of the previous trial (unless the first item happened to be the same in both trials). Thus, if there is to be an intrusion from the previous trial, there is no reason for it to be an protrusion of the second item from that trial. (In Figure 4-5, the most likely intrusion is the first item of the previous trial, assuming that the two lists share remote associations with the same start-of-list context). A similar argument applies to ordinal models like the Primacy Model, because a start-of-list cue (Page & Norris, 1996b) would mean that the most likely intrusion is always the first item from the previous trial. Only the Articulatory Loop Model predicts that the most likely intrusion is a protrusion, as in the data. This is because only a positional model assumes separate cues for each position, and, assuming the same cues are reused on each trial, any proactive interference will be of a positional kind. This illustrates the point made in Chapters 1 and 3, that positional errors necessitate a positional theory.

## Meta-analysis 2

This meta-analysis examined 9 conditions from 9 different experiments with grouped lists of phonologically dissimilar, nonrepeated items, to test the reliability of the results of grouping in Experiment 2. Further details of the conditions are given in Appendix 2.

### Interposition Constraint

The grouped condition of Experiment 2 showed a greater proportion of three-apart interpositions than two-apart transpositions. To test the reliability of this finding, the frequency of transpositions *n* positions apart (with groups of size *n*) was compared to the frequency of transpositions *n-1* positions apart, weighted by the opportunity for such transpositions (Footnote 2). The proportion of interpositions was greater in all 9 conditions, $N=9$, $p<.005$, $CI=(.01,.02)$. This confirms that interpositions in grouped lists override the locality constraint.

The grouped condition of Experiment 2 also demonstrated that more interpositions arose between the middle of groups than the start or end of groups. This finding was confirmed by comparing the proportion of interpositions between the middle of groups with the proportion between the start and end of groups. The proportion on middle positions was greater in all conditions, *N=9, p<.005, CI=(.02,.07)*. Finally, it was also confirmed that the proportion of interpositions followed by a correct response was greater than the proportion followed by a further interposition in all conditions, *N=9, p<.005, CI=(.28,.45),* supporting the conclusion that interpositions, like protrusions, arise singly.

## Meta-analysis 3

This meta-analysis examined 10 conditions from 3 different experiments that employed ungrouped lists in which phonologically similar and phonologically dissimilar items alternated. This was to test the reliability of the findings of Experiment 1. Further details of the conditions are given in Appendix 2.

### Confusion Constraint

Experiment 1 demonstrated that phonologically confusable items tend to transpose with one another, causing more errors for confusable items than nonconfusable items in lists where they alternate. All 9 conditions in the meta-analysis also showed a higher frequency of errors for confusable than nonconfusable items, *N=10, p<.005, CI=(.12,.22)*. However, Experiment 1 failed to find a consistent effect of confusable items on the recall of alternated nonconfusable items. This failure prompted two conclusions: 1) there is no effect of phonological similarity on cuing, and 2) there is no effect of errors on cuing (Chapter 2).

To test this finding, the frequency of errors on nonconfusable positions in alternating curves was compared with that in nonconfusable curves. There was a higher frequency of errors on nonconfusable positions in alternating curves in 8 of the 10 conditions, a result that was almost reliable, *N=10, p=.05, CI=(.02,.07).* This suggests the first finding in Experiment 1 may not generalise, particularly for lower-span subjects. One possible reason for this is the general knock-on effects of errors (Experiment 1). To test this notion, the above errors were conditionalised on correct recall of preceding items (Henson et al., 1996). In this case, the conditional probability of errors on nonconfusable positions in alternating curves was greater than in nonconfusable curves in only 4 conditions (and equal in 1 condition); a result that was not reliable, *N=9, p=.75, CI=(-.01,.01)*. This is consistent with the knock-on effects of an error (though also consistent with an effect of errors on cuing). More importantly, it is inconsistent with an effect of similarity on cuing, ruling out most chaining models (Chapter 1).

Finally, Experiment 1 also reported that phonological confusions were weighted by the distance between the two confusable items. This was confirmed by comparing weighted proportions of two-apart and four-apart confusions (Footnote 2), with all 10 conditions showing a greater proportion of the former, *N=10, p<.005, CI=(.01,.08)*.

## Summary of Empirical Constraints

The three meta-analyses revealed a rich set of empirical constraints on serial recall from short-term memory. In summary, the nine constraints were:

1. The primacy constraint: Recall of the first item is better than the second.

2. The recency constraint: Recall of the last item is better than the penultimate item, providing there are not too many omissions towards the end of recall.

3. The locality constraint: Items transpose small distances about their correct position.

4. The (weak) fill-in constraint: If an item is not recalled up to, or on, its correct position, it is the most likely error, other than an omission, on the following position.

5. The omission constraint: Omissions increase towards the end of recall, but not necessarily through failure to recall the last item anywhere.

6. The repetition constraint: Repetitions are literally few and far between, most often representing items recalled near the start and the end of a report.

7. The protrusion constraint: An erroneous item is more likely to occur at the same position as it appeared in the previous report than is expected by chance; intrusion of the whole report is rare.

8. The interposition constraint: Interpositions between groups are more common than expected by the locality constraint, most often between middle positions of groups, and without transposition of whole groups.

9. The confusion constraint: Phonologically similarity causes confusion in retrieval of items, but not in cuing of subsequent items (though the additional errors caused by confusions may have a small effect on retrieval of subsequent items).

### Comparison of Models

Without going into the full details of the three models considered above, it is worth noting how many of the empirical constraints are met by each model.

The Power Set Model meets the primacy and locality constraints. However, it has no specified mechanism to produce omissions or repetitions. It also fails to produce sufficient recency (Murdock, 1995), probably because it does not have enough fill-in (above), and it cannot meet the confusion constraint (Henson et al., 1996). Most importantly, being a chaining model, it offers no account of the positional errors required by the protrusion and interposition constraints. These failures remain true of other variations of serial order in TODAM, such as the nesting or chunking model (Murdock, 1983, 1993, 1995).

The Articulatory Loop Model meets the primacy, locality, omission and recency constraints, though its recency is often insufficient (Burgess & Hitch, 1992). However, it does not meet the fill-in, repetition or confusion constraints (Henson et al., 1996). Being a positional model, it has the potential to meet the protrusion and interposition constraints, as demonstrated by more recent developments of the model (Burgess & Hitch, 1996a). Further revisions of the model also address the fill-in and confusion constraints (Burgess & Hitch, 1996b), though not necessarily at a quantitative level (Chapter 5).

The Primacy Model meets the primacy, recency and locality constraints (Page & Norris, 1996b), though its fill-in property is too strong (above). It also meets the omission, repetition and confusion constraints (Henson et al., 1996), though not completely satisfactorily in the case of the omission constraint (Chapter 5). Being an ordinal model however, it cannot meet the protrusion and interposition constraints.

## Chapter Summary

This chapter described three meta-analyses of a number of experiments on serial recall from short-term memory. These analyses were driven by consideration of three specific models of serial recall, which make different predictions about the exact distribution of errors. They also served to confirm the generality of results in Experiments 1-3. The results of the meta-analyses were summarised in nine empirical constraints and none of the three models is able to meet all these constraints. In the next chapter, a new model is developed that can.

# 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 $i = 1,2...N$ in a sequence of $N$ items, $x(i)$ and $y(i)$ respectively, can be parameterised as:

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

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.

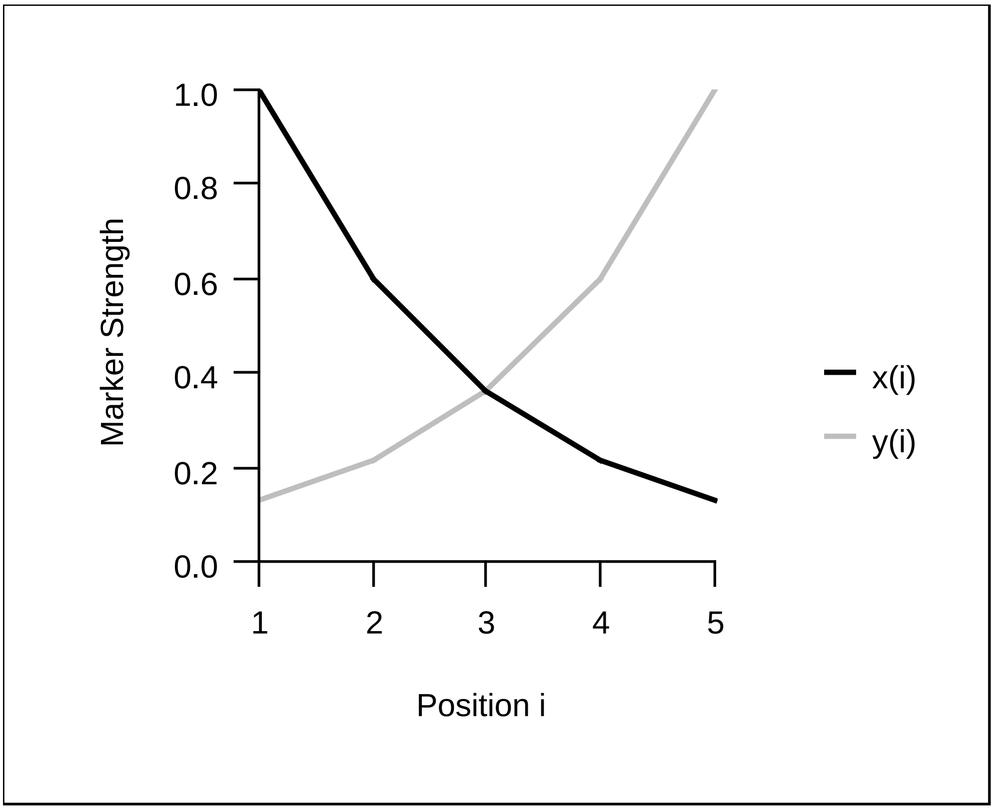



Figure 5-1: Start and end marker strengths, $x(i)$ and $y(i)$, on Positions $i=1..5$ of a five-item list. ($N=5$, $S_0=E_0=1.00$, $S=E=0.60$).

The positional code for Position *i* can be represented by the vector ***p****(i)=<x(i) y(i)>*. For example, the first position in Figure 5-1 has the code *<1.00 0.13>*, whereas the middle position has the code *<0.36 0.36>*. 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(**p**,**q**)*, between positional codes ***p****(i)* and ***q****(j)* for positions *i* and *j*:

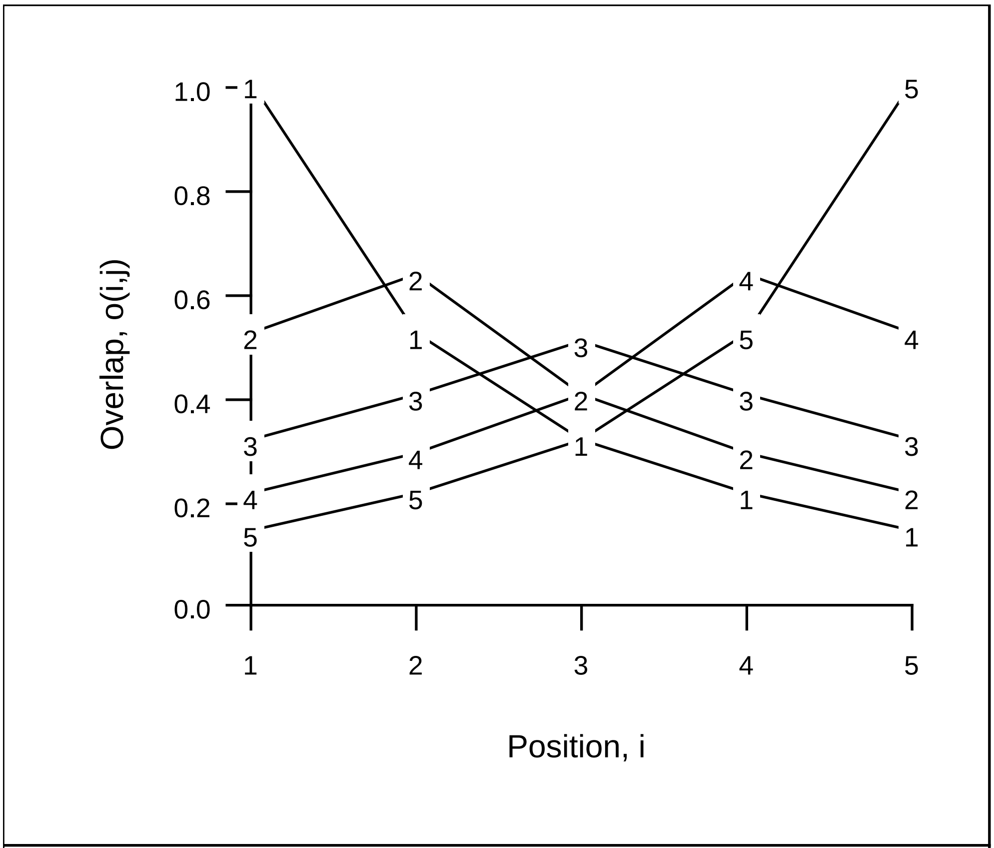
$$o(\boldsymbol{p}, \boldsymbol{q}) = \sqrt{\boldsymbol{p} \cdot \boldsymbol{q}} \times exp\left(-\sqrt{\sum_k (p_k - q_k)^2}\right) \qquad \text{Equation 5-2}$$

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).



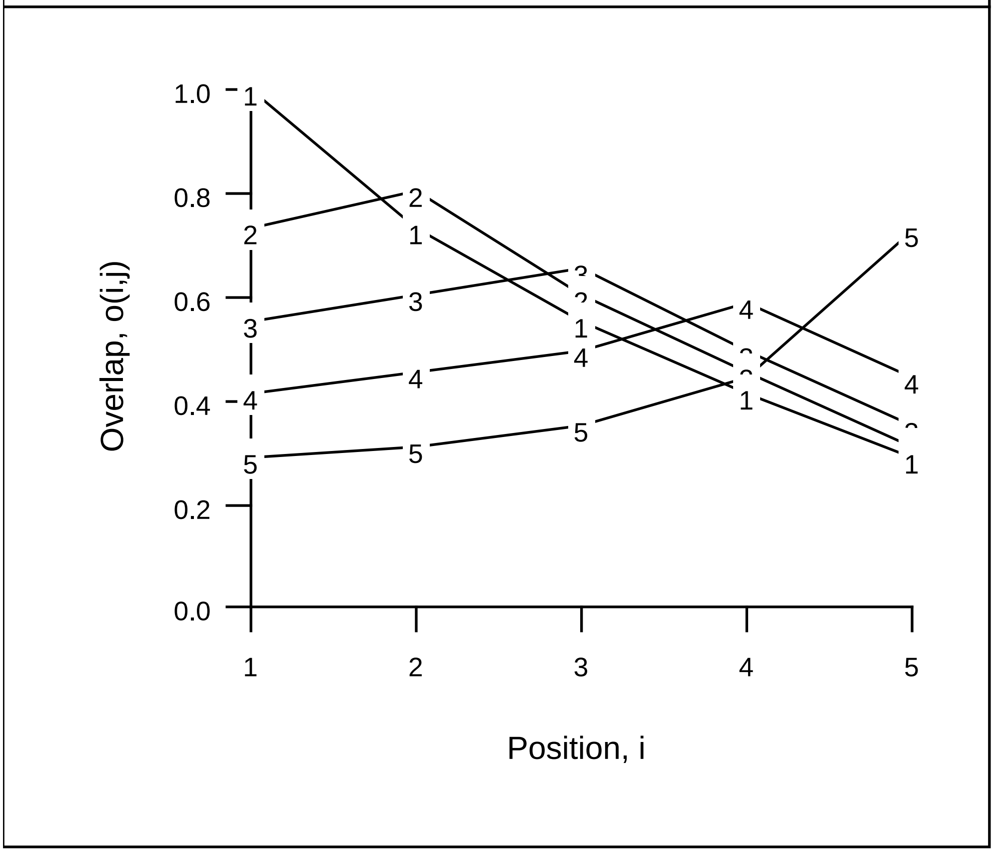




Figure 5-2: Positional uncertainty functions for each Position $j=1..5$ of a five-item list. (Upper panel, $S_0=E_0=1.00$, $S=E=0.60$; lower panel, $S_0=1.00$, $E_0=0.60$, $S=0.80$, $E=0.48$.)

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:

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

where $o(i,j)$ is the overlap between positional codes for Positions *i* and *j* (equivalent to $o(\mathbf{p},\mathbf{q})$ 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 $F$:

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

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

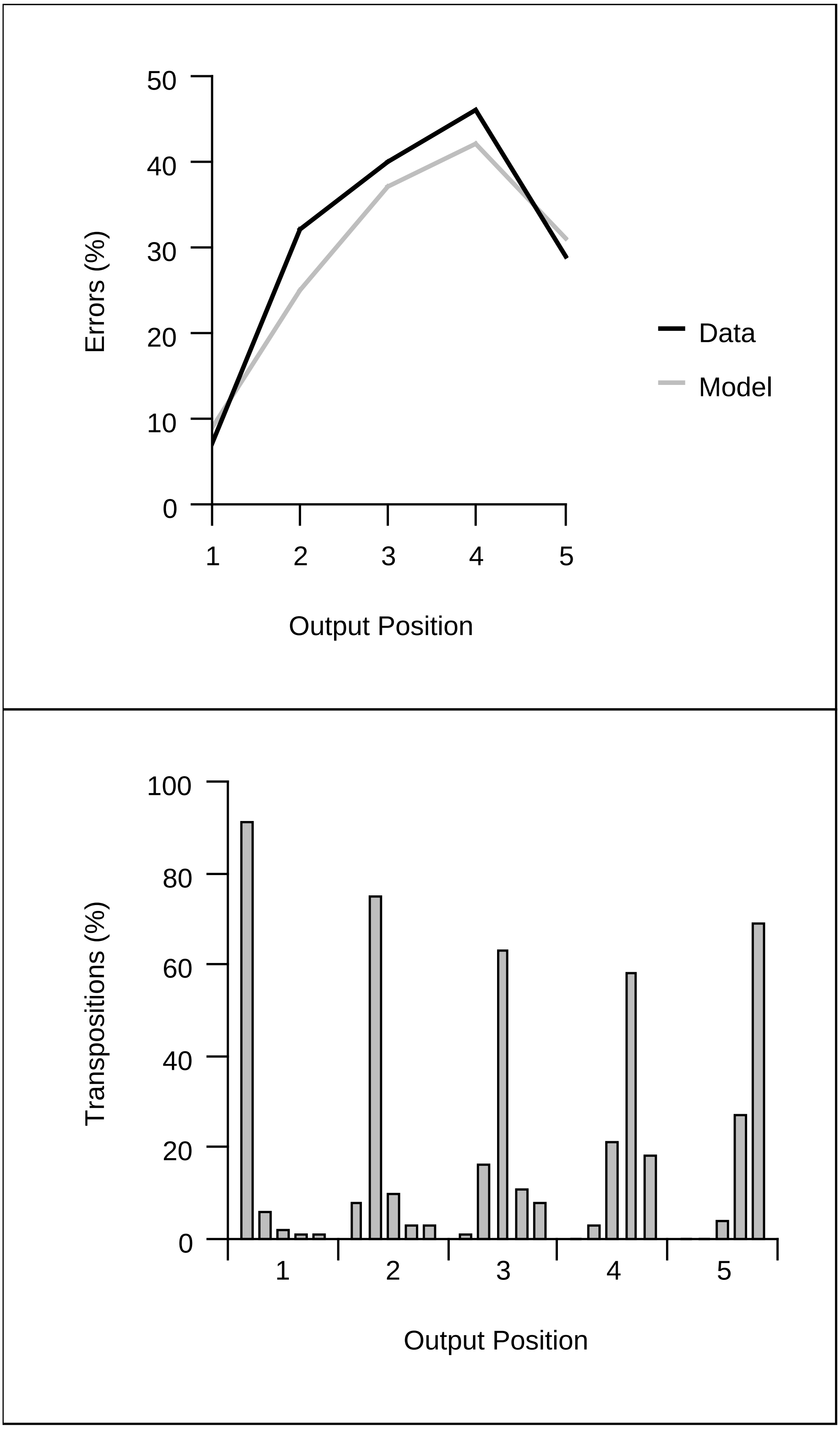



Figure 5-3: Errors by position from data and from SEM (upper panel) and transposition gradients from SEM (lower panel) in Fit 1.

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

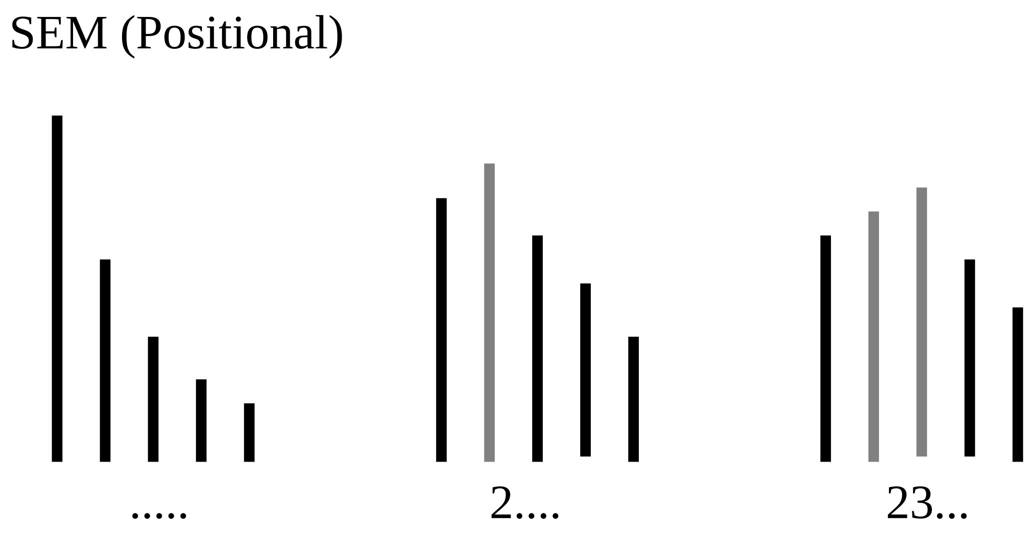



Figure 5-4: Competition space within SEM for the first three responses to a list *12345* recalled as *23...*, illustrating weak fill-in.

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, $F_0=F=0.60$ 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, $T^2=7.07$, $F(10,8)=0.33$, $p=.95$.

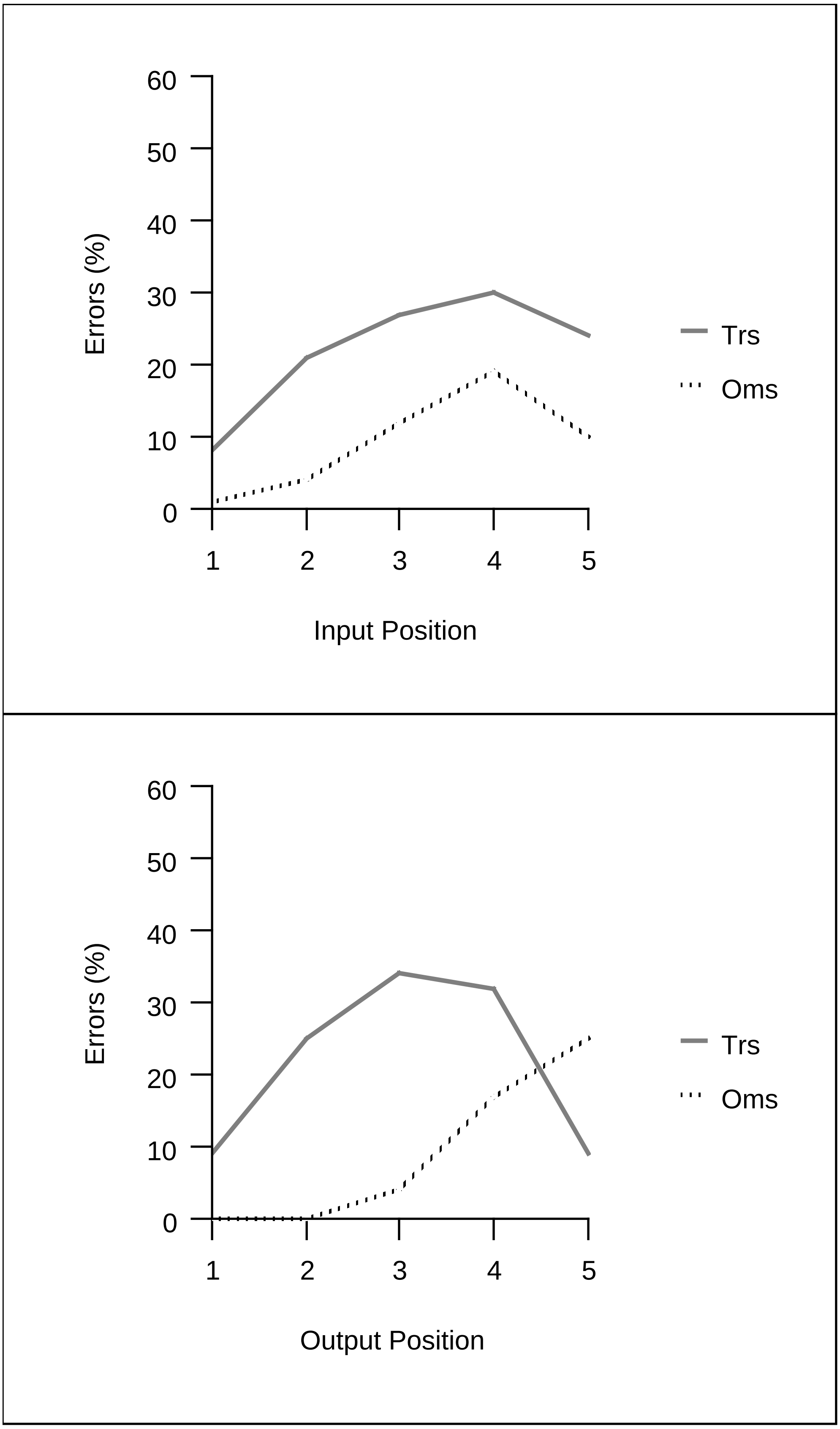



Figure 5-5: Omissions and transpositions by input position (upper panel) and output position (lower panel) from SEM in Fit 2.
(Trs=Transpositions; Oms=Omissions.)

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.

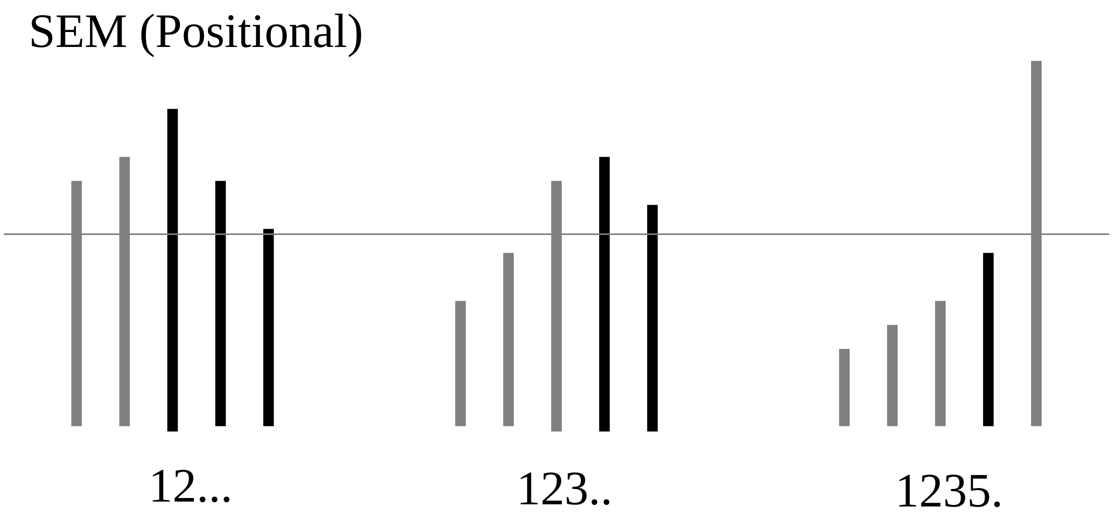



Figure 5-6: Competition space within SEM for last three responses to a list *12345* recalled as *1235-*, illustrating possibility of omissions.

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:

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

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.

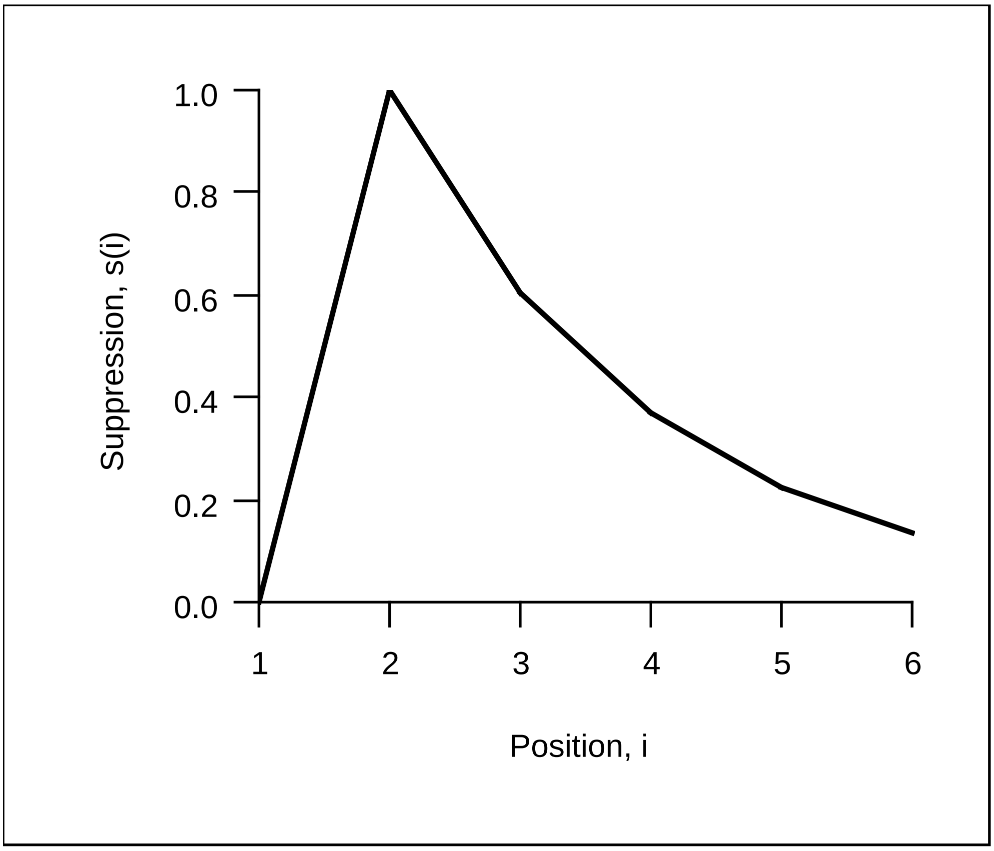



Figure 5-7: Suppression profile for an item recalled at Position 1.

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

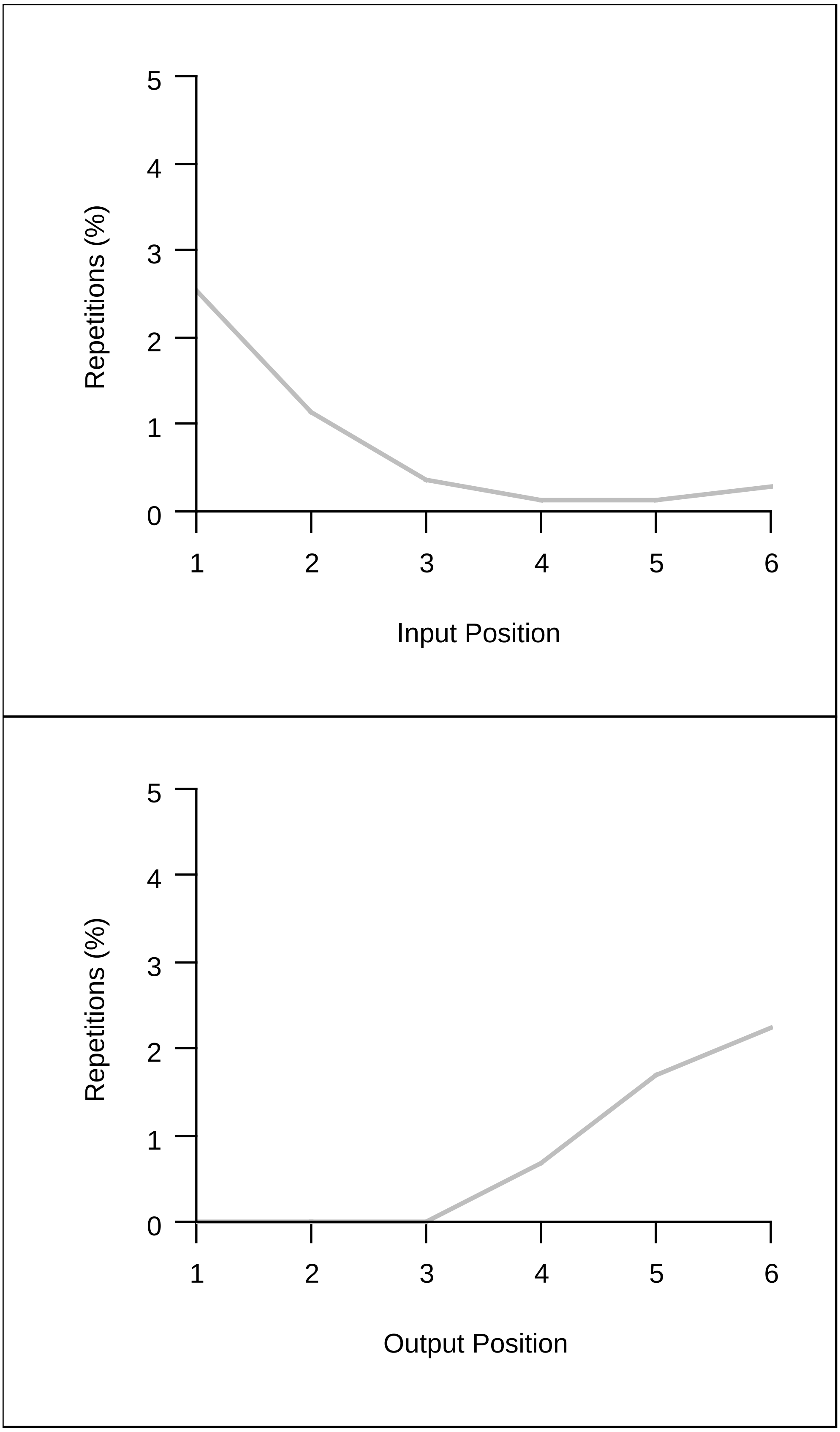



Figure 5-8: Repetitions by input position (upper panel) and output position (lower panel) from SEM in Fit 3.

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.

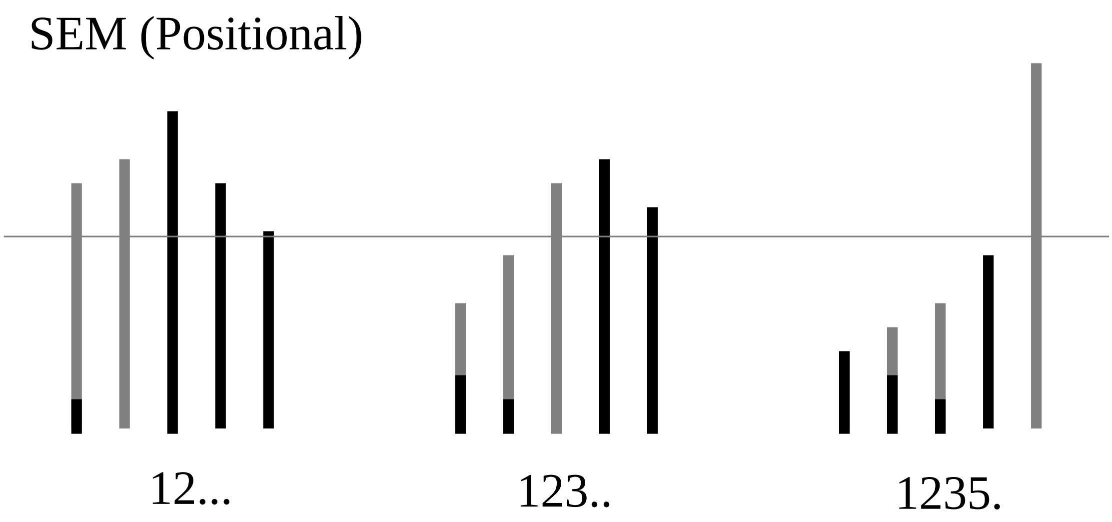



Figure 5-9: Competition space within SEM for the last three responses to a list *12345* recalled as *12351*, illustrating possibility of repetitions.

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).

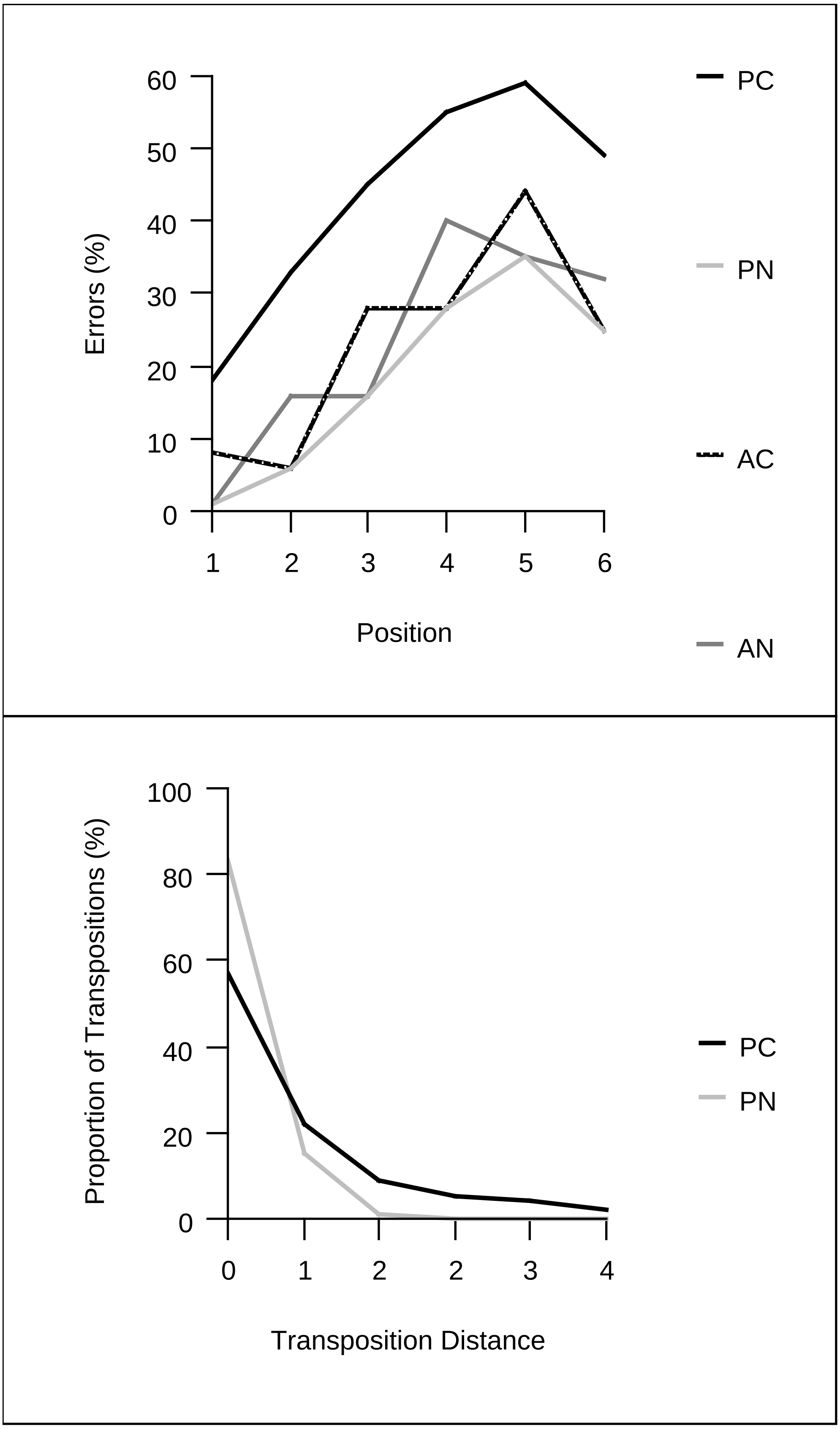



Figure 5-10: Errors by position (upper panel) and proportion of transpositions (including correct responses) by transposition distance (lower panel) from SEM in Fit 4.

**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 *<1 0>* and *<0 1>*, and therefore have only a finite resolution. As the number of positions coded within this range

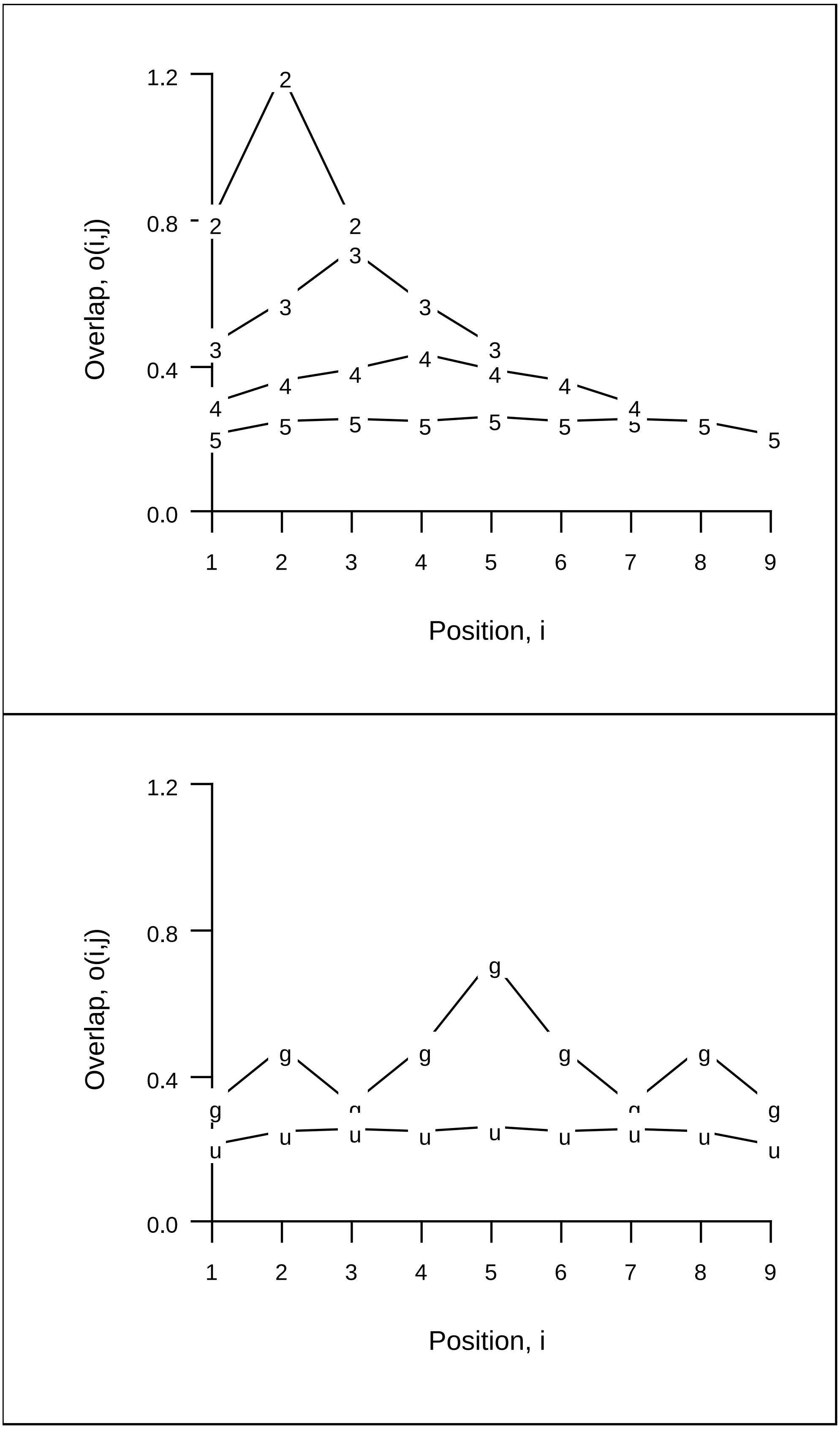



Figure 5-11: Positional uncertainty functions for the middle Position *j=2,3,4,5* of three-, five-, seven- and nine-item ungrouped lists (upper panel), and the middle Position *j=5* of ungrouped (line of 'u's) and 3-3-3 grouped (line of 'g's) nine-item lists (lower panel).

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.

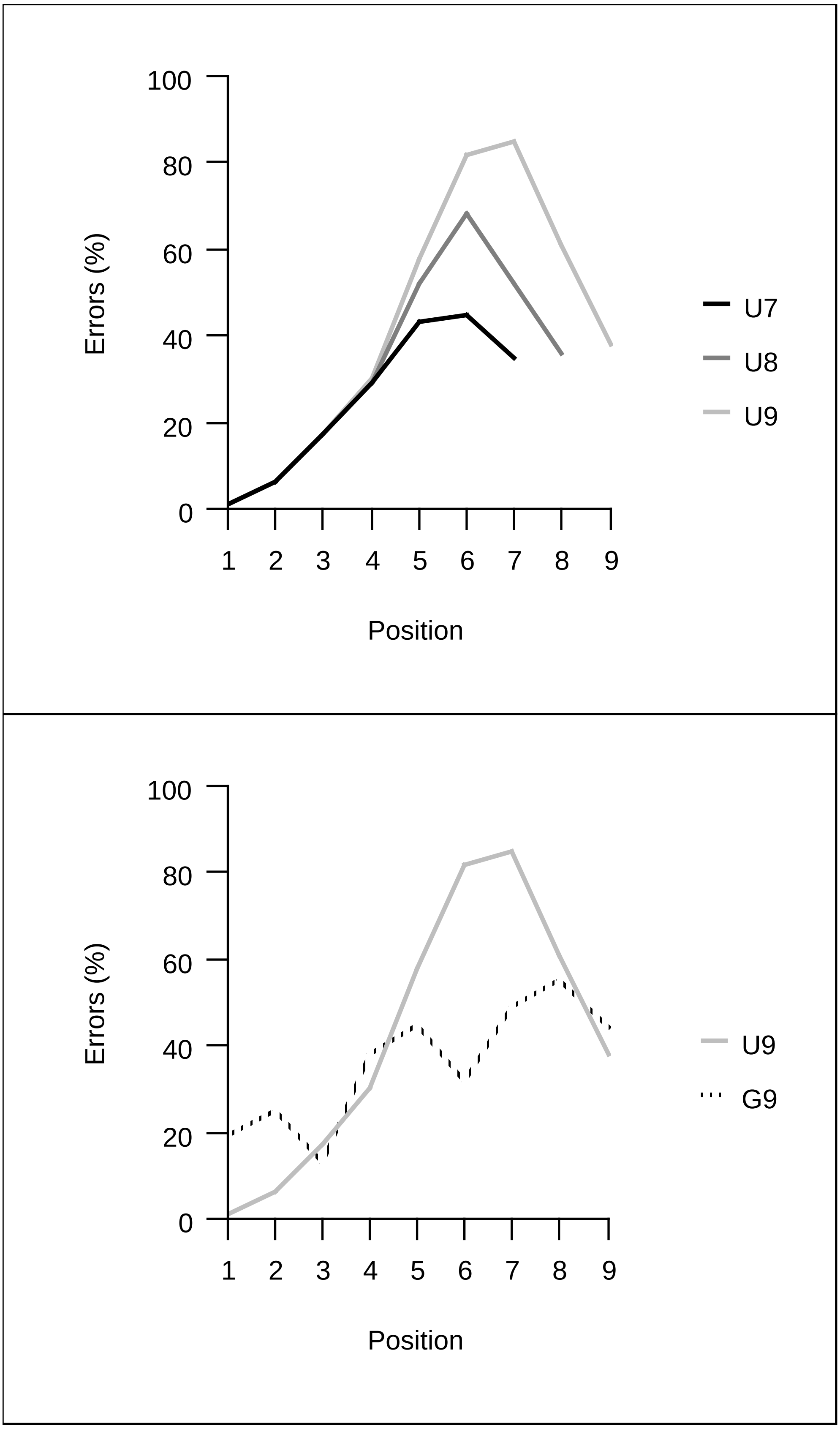



Figure 5-12: Errors by position for ungrouped lists (upper panel) and nine-item lists (lower panel) from SEM in Fit 5.

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

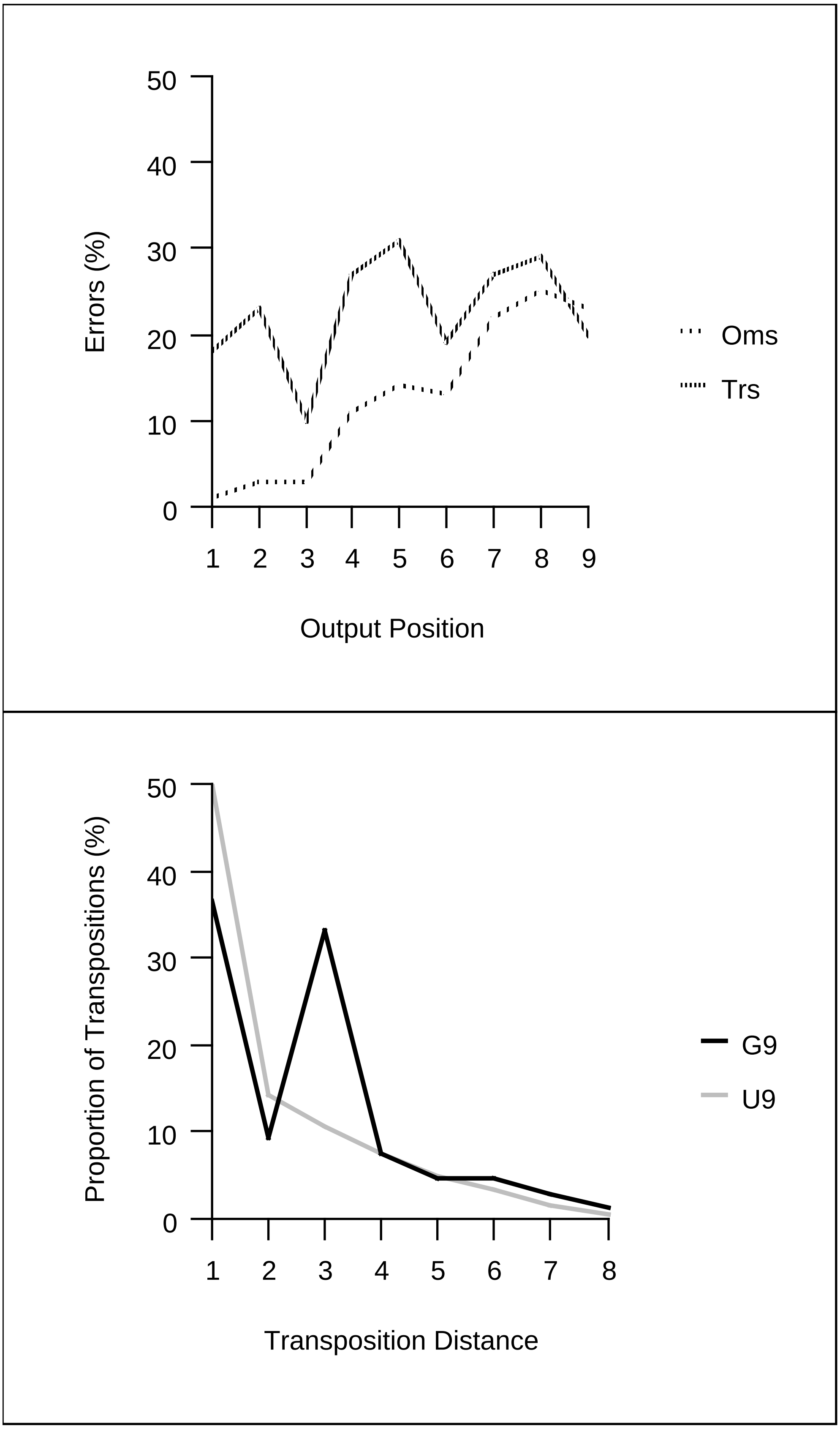



Figure 5-13: Omissions and transpositions by output position (upper panel) and proportion of transpositions by transposition distance (lower panel) from SEM in Fit 2. (Trs=Transpositions; Oms=Omissions.)

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

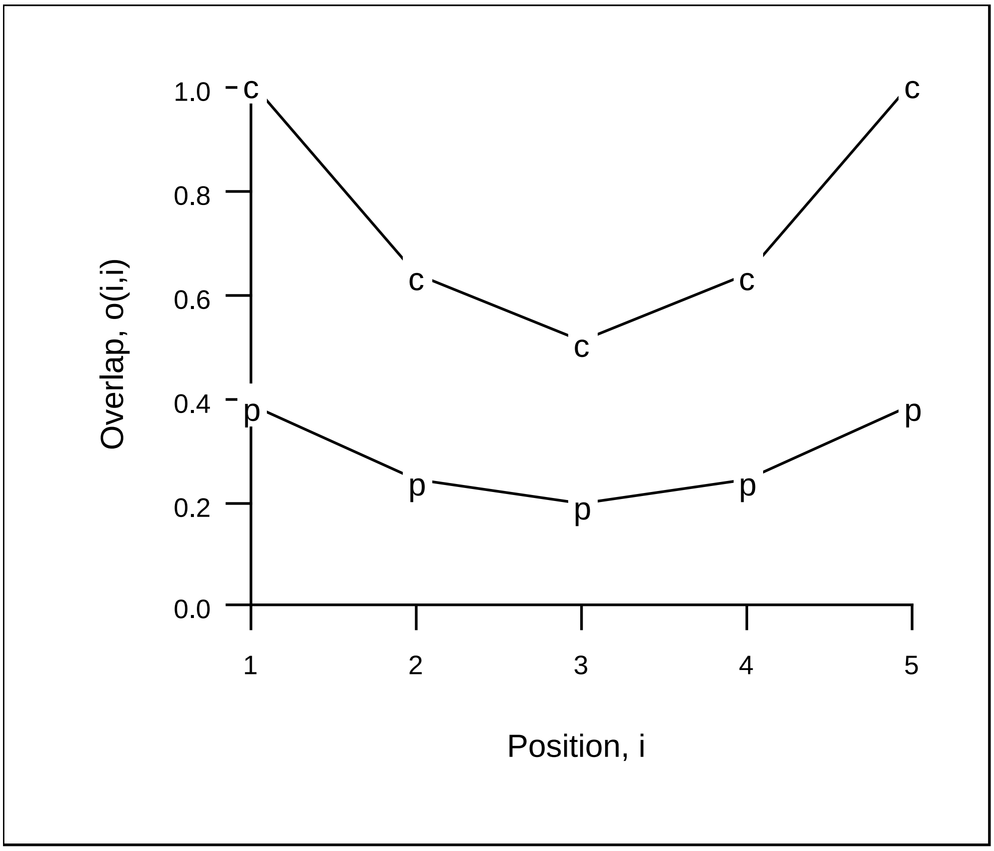



Figure 5-14: Example peaks of positional uncertainty functions for the current trial (line of ‘c’s) and previous trial (line of ‘p’s).

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

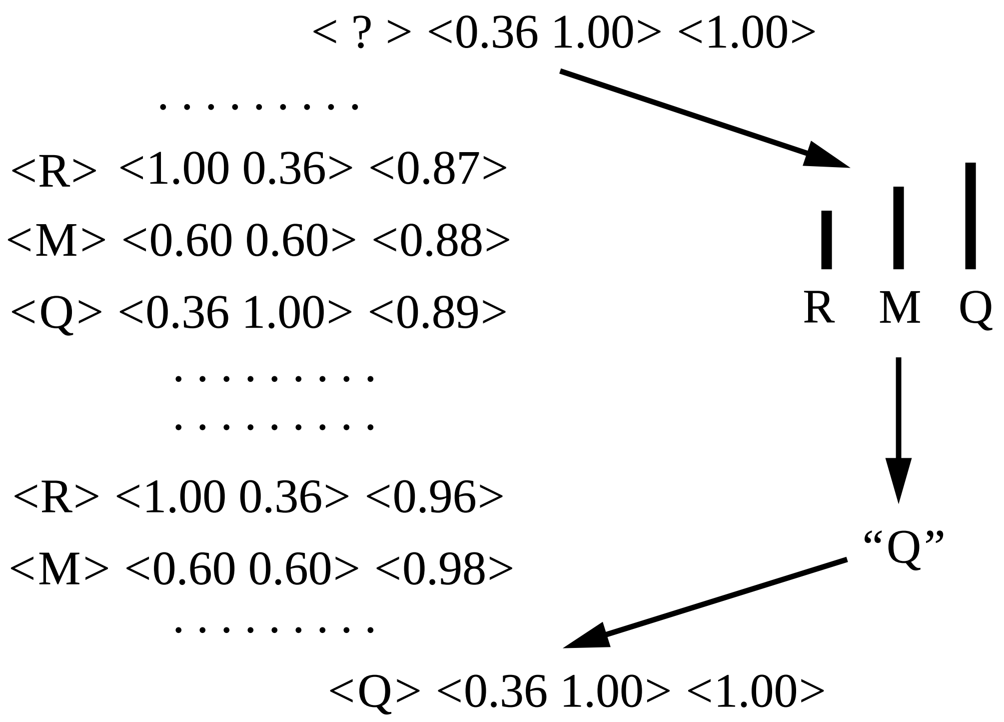



Figure 5-15: Schematic representation of SEM’s rehearsal process.
(Group-level codes not shown.)

(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

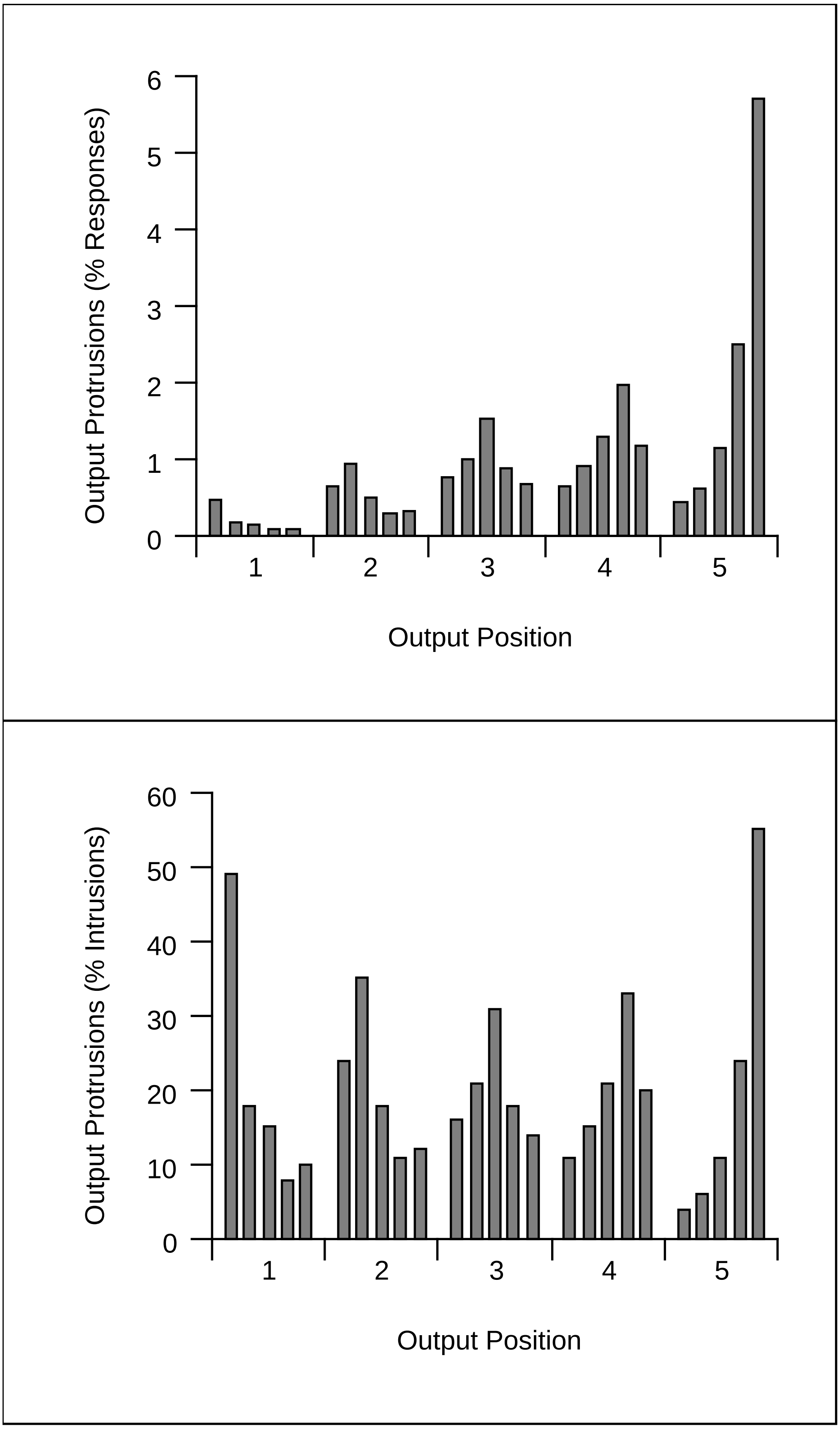



Figure 5-16: Output intrusions as a proportion of responses (upper panel) and as a proportion of intrusions per output position (lower panel) from SEM in Fit 6.

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 be 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 & Hennelly, 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 (Metcalfe & 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 (Hendrikx, 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 & Dosher, 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 (Hendrikx, 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 nonalphanumerical 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 reinstateable (positional) and non-reinstateable (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 reinstateable 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.

# Chapter 6: Absolute or relative position?

## *The nature of positional errors*

Experiment 2 showed that transpositions between groups tend to maintain the same position within a group. Such interpositions were as common as adjacent transpositions. Experiment 3 showed that intrusions between trials tend to maintain the same position within a trial. Such protrusions were significantly more common than expected by chance. However, both these types of positional error were demonstrated with groups or lists of equal length. The present chapter describes two experiments examining positional errors between sequences of different length. The results of these experiments are predicted (and fitted) by SEM, but not predicted by other models of short-term memory.

### Absolute and Relative Position

What happens to errors between sequences of different lengths? Do substitutions between such sequences maintain *absolute* position from the start of sequences (e.g., first, second, third, etc.), or do they maintain position *relative* to the end as well as the start? For example, consider transpositions between a group of three items followed by a group of four: Does the third and final item of the first group tend to swap with the third item of the second group, or with the fourth and final item of the second group (Figure 6-1)? The *absolute interpositions* in the former case are transpositions respecting absolute position within a group (i.e., third to third); the *terminal interpositions* in the latter case are transpositions respecting relative position within a group (i.e., end to end). Or consider recall of a list of seven items on one trial followed by recall of a list of five items on the next: Are most intrusions on the fifth and final position of the report of five items from the fifth position of the previous report, or from the seventh and final position of the previous report? In other words, which are more common: *absolute protrusions* or *terminal protrusions*?

The nature of such positional errors is important in light of SEM (Chapter 5). Because the end marker maintains the same maximum strength and rate of change, irrespective of sequence length, the model predicts terminal positional errors will dominate over absolute

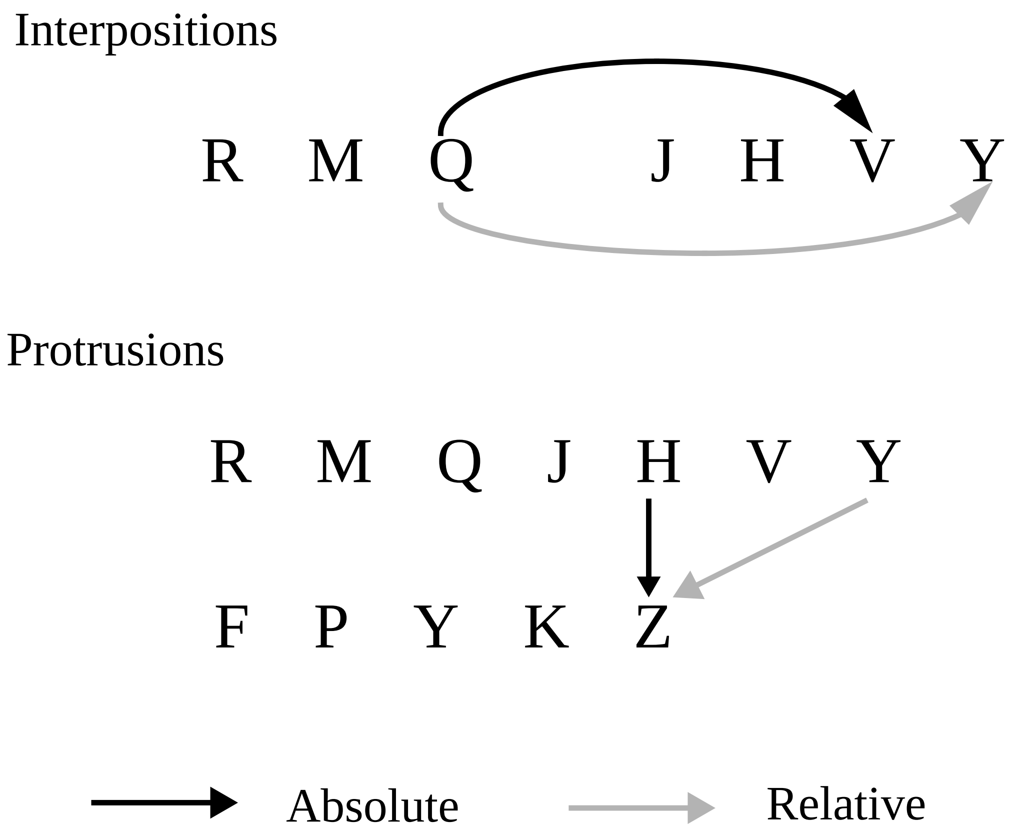



Figure 6-1: Illustration of absolute and relative positional errors.

positional errors. To take the above example of different list lengths, terminal protrusions are predicted because the cue for the last position in recall of a five item list is identical to the cue for the last position in recall of a seven item list. Thus, the positional overlap between the positional cues for the ends of the two lists will always have greater overlap than the positional cues for the fifth positions of each list.

The Articulatory Loop Model (Burgess & Hitch, 1992, 1996a, 1996b) predicts the opposite: that absolute positional errors will dominate over terminal positional errors. This is because their context window (Chapter 1) moves in the same, constant manner, irrespective of the length of the sequence. Thus the cue for the fifth position will be the same, whether there are five, six, seven, or more items, and maximum positional overlap will always arise for identical absolute positions. Indeed, this prediction would appear to follow from any model where the positional cue is derived from some regular or real-time temporal oscillation (e.g., Brown et al., 1996). Though the abstract formulation of the Lee and Estes (1981) model does not specify the exact nature of the positional cues it employs, viewing perturbations as the results of cyclic reactivations would also seem to imply coding of absolute position.

To distinguish these models, two experiments below directly compare the incidence of absolute and terminal positional errors. In the first experiment, these errors are interpositions between groups of different sizes. In the second experiment, these errors are protrusions between lists of different lengths. These experiments are the first tests of predictions of SEM, that position is coded relative to both the start and the end of sequences, and so terminal positional errors should exceed absolute positional errors in both cases. In addition, both experiments allowed subjects to indicate the confidence of each response. Page and Norris (1996b) have suggested that positional errors, particularly protrusions, might be the result of guessing strategies, implying that subjects are less confident of positional errors than other responses. If so, positional errors should disappear once guesses are removed from analysis. If positional errors remain however, there will be further support for the integral role of positional information in serial recall.

## Experiment 4

The present experiment tested whether transpositions between groups of different size maintain absolute or terminal position within groups. In the Grouped 3-4 condition, lists of seven items were split into a group of three followed by a group of four. The *critical positions* were Positions 3 and 7 (final positions within groups), and Positions 3 and 6 (third positions within groups). Terminal interpositions between the ends of groups were errors when Item 7 was recalled in Position 3, or Item 3 was recalled in Position 7. Absolute interpositions between the third position of groups were errors when Item 3 was recalled in Position 6, or Item 6 was recalled in Position 3. In the Grouped 4-3 condition, the lists were split into a group of four followed by a group of three. In this case, the critical positions were Position 3 and 7, and Positions 4 and 7. Terminal interpositions were then errors when Item 7 was recalled in Position 4, or Item 4 was recalled in Position 7, and absolute interpositions were errors when Item 3 was recalled in Position 7, or Item 7 was recalled in Position 3. Given the interdependencies between responses in a report (Chapters 4, 5), the chance probability of terminal and absolute interpositions cannot be determined in any simple manner. Therefore, an Ungrouped condition was included to check that differences in terminal and absolute interpositions were not simply an artefact of different baseline probabilities. Finally, all conditions allowed subjects to distinguish between confident responses and less confident responses (guesses), to test whether interpositions were the result of guessing strategies.

## Method

### Subjects

Twenty-four students from Cambridge University were tested, twelve male and twelve female, with a mean age of twenty years.

### Materials

Three blocks of thirty lists were constructed. Lists were permutations of seven single-syllable, low-frequency, phonologically nonconfusable words, drawn from a subset of those in Experiment 3: *goose, verve, latch, bathe, flown, clump* and *trout*. The order of words within lists was randomised except for the constraint that, over a block of trials, each word appeared approximately equally often at each position.

### Procedure

Blocks were assigned to three conditions for each subject. In the Ungrouped condition, the seven words were presented in the centre of a VDU at a rate of just over one a second (600-ms on, 200-ms off), each word replacing its predecessor. In the Grouped 3-4 condition, there was an additional 800-ms pause between the third and fourth words; in the Grouped 4-3 condition, there was an additional 800-ms pause between the fourth and fifth words. Subjects read the words in silence and were told to use the pause to group the lists appropriately. As soon as the last item had disappeared, a cue followed for immediate, serial recall.

Subjects recalled the list by writing the first letter of each word in two rows of seven boxes provided on a response sheet. Subjects were told to write responses they were sure about in the top row, and responses they were not sure about, or which were guesses, in the bottom row. They could go up and down the rows as much as they liked, as long as they gave one and only one response in each column (i.e., gave exactly seven responses in total). All seven words were permanently on display, from which subjects could guess if necessary.

Subjects were asked to write from left to right on the response sheet, recalling the lists in a forward order. Though they only had to write the first letter of each word, subjects were told to remember the lists as lists of whole words (and all reported obeying this instruction).

Subjects received six, ungrouped practice trials, followed by the three blocks of lists. The order of blocks was counterbalanced across subjects. The order of conditions was constrained by the fact that the Ungrouped condition was always first, followed by the two grouped conditions, which alternated across subjects. This was to reduce the chance of subjects subjectively grouping the ungrouped lists, as might happen if a grouped condition preceded the ungrouped one. The whole experiment took about 40 minutes.

**Results**

In brief, terminal interpositions were more common than absolute interpositions, irrespective of whether guesses were included or excluded. Many terminal interpositions were repetitions of an item at the end of both groups. Confidence and accuracy of responses were highly correlated, as expected, though guesses were far from random and a considerable number of errors were not indicated as guesses.

**Overall Performance**

The proportion of lists correct was greater in the Grouped 3-4 ($M=.39$, $SD=.29$) and Grouped 4-3 ($M=.39$, $SD=.27$) conditions than the Ungrouped condition ($M=.22$, $SD=.22$). Tests of weighted log-odds showed the difference was significant in both cases, $Z(24)>7.25$. $p<.0001$, but no significant difference between the two styles of grouping, $Z(24)=0.09$, $p=.93$. Error position curves (Figure 6-2) suggested some spontaneous 4-3 grouping in the Ungrouped condition, though several grouping strategies were reported (e.g., 3-4, 2-2-3 and 3-2-2). Removing guesses reduced the number of correct responses on most positions.

**Errors on Critical Positions**

With the same seven items per trial, the only errors made were transpositions. With guesses removed, the frequency of errors was calculated from the number of transpositions remaining. In order to compare error frequencies on critical positions with baseline measures in ungrouped lists, the ungrouped lists were scored as if they were grouped 3-4 or 4-3.

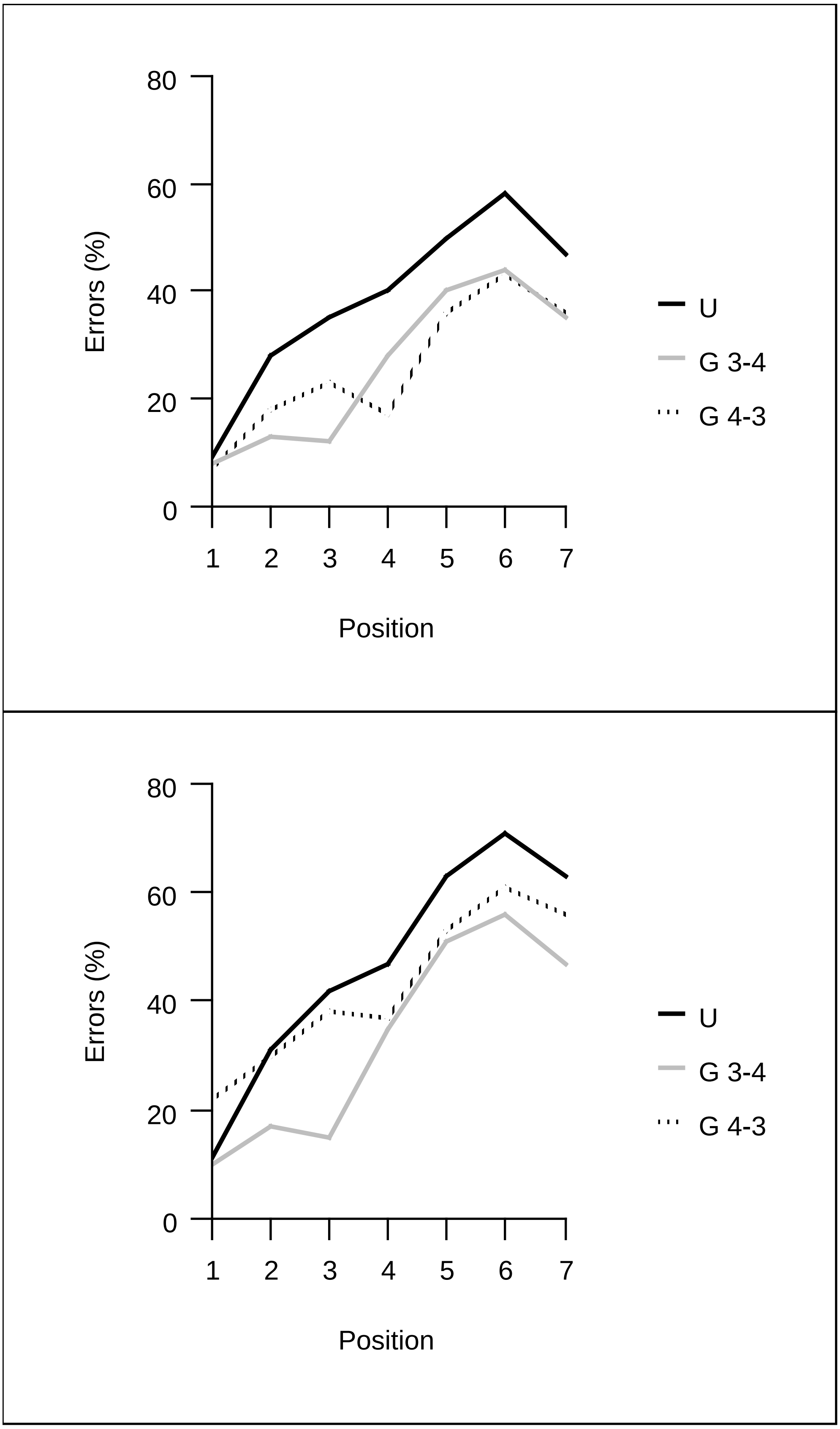



Figure 6-2: Errors by position with guesses (upper panel) and without guesses (lower panel) in Experiment 4.
(U=Ungrouped; G=Grouped.)

Under 3-4 grouping, a three-way ANOVA on the log-odds of an error on either critical position showed effects of condition (Grouped 3-4 vs. Ungrouped), $F(1,161)=128.46$, $p<.001$, guesses, $F(1,161)=218.90$, $p<.001$, and position, $F(1,161)=4.62$, $p<.05$, but no significant interactions, $F(1,161)<1.34$, $p>.25$. As expected, explicit grouping reduced error frequencies (Experiment 2), as did excluding guesses (Table 6-1). The effect of position reflected fewer errors on final positions than third positions of groups, also as expected from Experiment 2.

Under 4-3 grouping, a three-way ANOVA on the log-odds of an error on either critical position showed effects of condition (Grouped 4-3 vs. Ungrouped), $F(1,161)=83.42$, $p<.001$, and guesses, $F(1,161)=188.41$, $p<.001$, but not position, $F(1,161)=1.74$, $p=.19$. The interaction between condition and position approached significance, $F(1,161)=3.06$, $p=.08$, but no other interactions did, $F(1,161)<1.31$, $p>.25$. Apart from the slightly different pattern in the Ungrouped condition, the results resembled those under 3-4 grouping (Table 6-1).

| | Grouped | | Ungrouped | |
|---|---|---|---|---|
| | Third | Final | (Third) | (Final) |
| Grouping 3-4 | | | | |
| Guesses Included | .28<br>(.16) | .24<br>(.16) | .47<br>(.20) | .41<br>(.17) |
| Guesses Excluded | .12<br>(.09) | .10<br>(.10) | .21<br>(.16) | .21<br>(.14) |
| Grouping 4-3 | | | | |
| Guesses Included | .30<br>(.17) | .27<br>(.16) | .41<br>(.17) | .44<br>(.18) |
| Guesses Excluded | .15<br>(.12) | .12<br>(.11) | .21<br>(.14) | .21<br>(.14) |

Table 6-1: Frequency of errors on critical positions in Experiment 4.

The proportion of errors on a critical position that were interpositions from the other critical position was calculated for the 22 subjects who made at least one error per critical position with guesses excluded. These proportions were small; the majority of errors on critical positions were from adjacent, within-group positions. Nevertheless, under 3-4 grouping, a three-way ANOVA on the log-odds showed a significant effect of position, *F(1,147)=9.49, p<.005*, and a significant interaction between position and condition, *F(1,147)=9.42, p<.005*. No other effects were significant, *F(1,147)<1.66, p>.20*. The interaction between position and condition reflected a greater proportion of interpositions between final positions than third positions in the Grouped 3-4 condition, but not the Ungrouped condition (Table 6-2). In other words, interpositions respected relative rather than absolute position, and this did not appear to be an artefact of different baseline probabilities.[1]

Under 4-3 grouping, a three-way ANOVA on log-odds did not show any significant effects, *F(1,147)<1.83, p>.18*, though the effect of condition, *F(1,147)=3.20, p=.08*, and interaction between condition and position, *F(1,147)=2.75, p=.10*, approached significance. Despite the lack of statistical significance, the pattern of results was very similar to that under 3-4 grouping (Table 6-2). A reason for the difference in significance of results in the Grouped 3-4 and Grouped 4-3 conditions is given in the Discussion.

| | Grouped | | Ungrouped | |
|---|---|---|---|---|
| | Third | Final | (Third) | (Final) |
| Grouping 3-4 | | | | |
| Guesses Included | .12<br>(.09) | .18<br>(.13) | .15<br>(.07) | .17<br>(.09) |
| Guesses Excluded | .12<br>(.09) | .20<br>(.18) | .18<br>(.13) | .18<br>(.13) |
| Grouping 4-3 | | | | |
| Guesses Included | .13<br>(.09) | .15<br>(.10) | .17<br>(.09) | .18<br>(.09) |
| Guesses Excluded | .13<br>(.12) | .16<br>(.15) | .18<br>(.13) | .18<br>(.13) |

Table 6-2: Proportion of errors that were interpositions in Experiment 4.
(Calculated from weighted log-odds, *n*=22.)

1. It remains possible that the differences in proportions are an artefact of differences in overall numbers of errors on critical positions, with fewer errors on final than third positions within groups (Table 6-1). This possibility is discounted in the analysis below, where errors are confined to repetitions in the second group.

To confirm the possible interactions between critical position and grouping condition, pairwise comparisons were performed on the weighted log-odds of interpositions on third versus final positions in groups. The proportion was significantly greater between final positions of groups in the Grouped 3-4 condition, whether or not guesses were included, $Z(22)>2.00$, $p<.05$. No such differences were significant in the Ungrouped condition, $Z(21)<0.98$, $p>.33$. The proportion was also greater in the Grouped 4-3 condition, but this difference was not significant either with, $Z(21)=0.88$, $p=.40$, or without, $Z(21)=1.09$, $p=.28$, guesses. This was also true of the Ungrouped condition, $Z(22)<0.41$, $p>.78$.

**Between-group Repetitions**

The previous analyses demonstrated that a transposition at the end of one group was more likely to come from the end of the other group than from the third position of the other group, at least in the Grouped 3-4 condition. Closer inspection of the data revealed that many of these errors occurred at the end of the second group, and were repetitions of an item recalled at the end of the first group. In the Grouped 3-4 condition for example, Item 3 was sometimes recalled on both Position 3 and Position 7. Even if a different item was recalled on Position 3, that item was likely to be recalled again on Position 7 (e.g., Item 2 might be recalled on both Position 3 and Position 7). This suggested that many interpositions might be perseverations resulting from proactive interference from recall of the first group on recall of the second, much like the proactive interference between reports in Experiment 3. Consequently, responses in the second group were examined in more detail (in a manner parallel to immediate intrusions in Experiment 3 and Experiment 5, and which allowed comparison with chance levels). Specifically, analysis was restricted to the third and the final position within the second group.

Given a response on a critical position in the second group that was a repetition of an item recalled somewhere in the first group (a *between-group repetition*), the interest was whether that item came from the same critical position of the first group. For this analysis, the two grouped conditions were collapsed together, and, as before, the ungrouped lists were treated as if they were grouped in the corresponding manners. Given the small numbers involved, guesses were included. The proportion of between-group repetitions that were absolute or terminal interpositions was then calculated for the 19 subjects that made at least one between-group repetition per critical position (Table 6-3). Pairwise comparisons of weighted log-odds showed a significantly greater proportion of interpositions between final positions of groups than third positions of groups in grouped lists, *Z(19)=2.68, p<.01*, but not in ungrouped lists, *Z(19)=0.08, p=.94*. This was not due to different overall incidence of between-group repetitions, which did not differ significantly in either case, *Z(19)<0.45 p>.6*5.

| | Grouped | | Ungrouped | |
|---|---|---|---|---|
| | Third | Final | (Third) | (Final) |
| Between-group Repetitions | .16<br>(.07) | .17<br>(.08) | .16<br>(.07) | .17<br>(.08) |
| Interpositions | .24<br>(.19) | .40<br>(.24) | .31<br>(.19) | .31<br>(.19) |

Table 6-3: Frequency of between-group repetitions, including guesses, and proportion that were interpositions in Experiment 4.
(Calculated from weighted log-odds, *n=19.*)

The chance probability that between-group repetitions maintain absolute or terminal position is difficult to determine exactly, because there are three responses in the first group in the Grouped 3-4 condition and four in the Grouped 4-3 condition. This means the chance probability lies somewhere between *.25* and *.33*. Taking an average value of *.29* (a value close to that in the ungrouped lists), repetitions between the final positions of groups were more frequent than expected by chance in the grouped lists, *Z(19)=2.82, p<.005*, but not the ungrouped lists, *Z(19)=0.65, p=.52*.

Finally, repetitions on the first position of the second group were still predominantly from the first position of the previous group. Indeed, the proportion of between-group repetitions on the first position of groups in the grouped list (*M=.48, SD=.27*) was significantly greater than the baseline figure in ungrouped lists (*M=.37, SD=.28*) and the figure of *.29* expected by chance, *Z(16)>2.03, p<.05* in both cases. In other words, repetitions between groups respected both terminal positions, the start and the end, of groups.

**Guesses**

The proportion of responses that were guesses was greater in the Ungrouped condition (*M=.29, SD=.15*) than the Grouped 3-4 (*M=.23, SD=.16*) or Grouped 4-3 (*M=.22, SD=.14*) conditions. A two-way ANOVA on the log-odds of a guess showed a significant effect of condition, $F(2,460)=14.69$, $p<.001$, output position, $F(6,460)=94.17$, $p<.001$, and an interaction that almost reached significance, $F(12,460)=1.68$, $p=.07$. Guesses increased towards then end of recall, with a particularly large increase across group boundaries (i.e., the whole of the second group was often "guessed", like the omission of groups in Experiment 2).

Responses in the Ungrouped condition were split by whether or not they were correct and whether or not they were guesses, forming contingency tables for each subject. A combined test of significance of these tables showed an extremely high correlation between accuracy and confidence of responses, $Z(24)=27.97$, $p<.0001$, mainly owing to the large number of correct responses that were not guesses (Table 6-4). Nevertheless, almost half of the errors were not indicated as guesses (*M=.48, SD=.07*). Though some of these may have reflected a failure or reluctance to indicate guesses, such a large proportion suggests that subjects were often unaware of having made an error. At the same time, a considerable proportion of guesses were correct (*M=.31, SD=.08*), suggesting that guesses were more than random choices of list items (of which only *.14* would be correct).

| | Not Guess | Guess |
|---|---|---|
| Correct | 2673 | 438 |
| Error | 926 | 1003 |

Table 6-4: Number of guesses and errors collapsed across subjects in the Ungrouped condition of Experiment 4.

**Discussion**

The present experiment showed that interpositions between groups respect the terminal positions of groups rather than absolute position within groups. Though differences were small, the proportion of errors that were interpositions was significantly greater between the ends of groups than between the third positions of groups, particularly in the Grouped 3-4 condition. Similarly, the proportion of repetitions that were interpositions was greater at the end of the second group than the third position of that group, and significantly above chance levels. These data suggest that position within a group is coded relative to both the start and end of that group, confirming the prediction of SEM and questioning the coding of absolute position in other models, such as the Articulatory Loop Model (Burgess & Hitch, 1992) and its extension to grouping (Burgess & Hitch, 1996a, 1996b).

The proportion of errors that were interpositions was unchanged by the removal of guesses, which showed no interaction with grouping or critical position. This implies that interpositions are not simply the result of guessing strategies. The difference in proportions of terminal and absolute interpositions in grouped lists seemed to reflect a depression of absolute interpositions relative to ungrouped lists, rather than an elevation of terminal interpositions. The depression of transpositions between the same absolute position within groups resembles the depression of transpositions between groups that were not interpositions in Experiment 2. The lack of significant elevation of terminal interpositions was surprising, given that grouping increased this proportion in Experiment 2. When analysis was confined to between-group repetitions however, grouping did increase the proportion of repetitions that were terminal interpositions, as well as decreasing the proportion that were absolute interpositions.

The demonstration that interpositions respect position relative to the end as well as start of groups can be explained simply by SEM. Because grouping conditions were blocked in the present experiment, subjects knew the size of both groups in advance. Thus, it is feasible that the strength of the end marker could represent expectation for the end of a group (Chapter 5). Then the strength of the end marker at the third position in a group of three, where the final item is expected, will differ to its strength at the third position in a group of four, where the final item is not yet expected; equivalent strength of the end marker will only occur at the fourth position in the group of four. The capability of SEM to fit present data is confirmed in the General Discussion.

It is not apparent how the present results can be explained by other accounts of grouping. For example, Hitch, Burgess, Shapiro, Culpin and Malloch (1995) suggested that grouping is a rhythmic process driven by internal oscillators. This is based on the fact that the grouping advantage for visual material may be removed by irrelevant, background tones, which they suggest entrain the internal oscillators to a different rhythm (e.g., Treisman, Cook, Naish & MacCrone, 1994). However, Henson (1996a) failed to replicate their results in two experiments with a fixed list length procedure, rather than span procedure, and a more sensitive index of grouping. In any case, internal oscillators would seem to predict absolute rather than terminal interpositions, contrary to present results. This suggests that the grouping advantage is not solely due to internal oscillators.

This is not to deny that rhythm contributes to grouping effects in other situations, such as when the group sizes are equal. The 3-3-3 temporal grouping in Experiment 2 for example conforms to a natural 4/4 rhythm in each metrical segment, whereas the 3-4 and 4-3 groupings in the present experiment have different rhythms in each segment. This may be one reason why interpositions were much more frequent in Experiment 2 than the present experiment. Thus, a rhythmic account may be necessary to explain why regular group sizes are more effective than irregular group sizes (Wickelgren, 1967), a result not necessarily predicted by SEM. Finally, there are other aspects of the possible interaction between grouping and articulatory suppression or finger tapping, such as differences between internal and external pacing (Hitch et al., 1995), which clearly warrant further investigation.

The observation that many interpositions are repetitions of an item at the end of both groups suggests that interpositions may be the result of proactive interference from recall of the first group on recall of the second. Indeed, around half of the interpositions measured on critical positions were repetitions, which is probably sufficient to explain the differences between grouped and ungrouped conditions in Table 6-2. In other words, interpositions between groups may result from the same output effects that cause protrusions between reports (Experiment 3). The only difference is that repetitions between groups must contend with the additional effect of suppression. Suppression of previous responses reduces repetition within reports, but has little effect on repetition between reports, given that it has normally worn off between trials (Chapter 5). The refractory nature of suppression also explains why terminal interpositions appeared more frequent in the Grouped 3-4 condition than Grouped 4-3 condition: When an item is recalled at the end of the first group, there is more time for suppression to wear off before the end of the second group in the Grouped 3-4 condition, with three intervening responses, than the Grouped 4-3 condition, with only two intervening responses (Chapter 5). The issue of output effects is resumed in the General Discussion.

## Experiment 5

The previous experiment demonstrated that transpositions between groups of different size tend to maintain terminal rather than absolute position. The present experiment tests whether the same is true of intrusions between lists of different length. Specifically, in the Variable condition, subjects saw either five, six or seven words on a given trial. In the Fixed condition, subjects always saw six words on each trial. Given that Chapter 4 demonstrated that output protrusions are more common than input protrusions, the former measure was used in the present experiment. The critical positions were therefore the fifth position in reports and the final position in reports. In the Variable condition, absolute protrusions were intrusions on the fifth position of a report that also occurred on the fifth position of the previous report; terminal protrusions were those on the final position of a report that also occurred on the final position of the previous report. In the Fixed condition, measurements of absolute and terminal protrusions are of course confounded, but the frequency of protrusions on the fifth position and final position were also examined, to give a comparative baseline measure of protrusions.

The Variable condition was interesting for a further reason. Precautions were taken to ensure that subjects in this condition did not know in advance the length of the list on each trial. This raises questions about interpretation of the end marker in SEM. If the end of a list is unpredictable, it is hard to see how the strength of the end marker during presentation of a list could represent the degree of expectation for the end of that list (Chapter 5). Interpretation would be particularly difficult if protrusions were found between the ends of reports, rather than between the same absolute positions. Such errors would require the end marker to grow in strength towards the end of a list in the same manner, irrespective of the list length, even when the end of the list is not known until it occurs.

Finally, both Variable and Fixed conditions used the same method of distinguishing confident responses from guesses as in Experiment 4. There is evidence to suspect that protrusions might disappear when guesses are removed from reports. This evidence comes from research suggesting that all intrusions are guesses. For example, Dillon and Thomas (1975) showed that instructing subjects not to guess dramatically reduced the proportion of errors that were intrusions. Indeed, subjects were less confident of intrusions than other errors. Bjork and Healy (1974) found a similar result, provided intrusions were not phonological confusions. Dillon and Thomas, like Conrad (1960), used their results to argue that proactive interference results from correct items being inaccessible, and therefore being replaced by guesses, which are often items from previous trials. They argue against the notion (in SEM) that proactive interference is due to response competition. If Dillon and Thomas are correct, protrusions should be affected more by the removal of guesses than other types of error.

**Method**

**Subjects**

Thirty students from Cambridge University were tested, twenty male and ten female, with a mean age of twenty years.

**Materials**

Stimuli were lists of five, six, or seven single-syllable, low-frequency words, drawn from the same set as Experiment 3. The words were split into two subsets that were alternated across trials, such that no word appeared in two consecutive trials. The order of words within lists was randomised except for the constraint that, over all trials, each word appeared approximately equally often at each position.

**Procedure**

Two blocks of 46 lists were constructed. The first list of six items in each block was not analysed. In the Fixed block, the remaining 45 lists also contained six words. In the Variable block, there were 15 lists of five words, 15 of six words and 15 of seven words. The order of lists in the Variable block was such that no two consecutive trials had lists of the same length.

Each word was presented in the centre of a VDU, replacing the previous word, at a rate of just over one a second (600-ms on, 200-ms off). Subjects were instructed to read the words in silence. Some time after the last word had disappeared, a cue appeared to signal immediate serial recall. The pause before this cue appeared was such that the amount of time elapsing between the onset of the first word and the onset of the cue was identical, no matter how many words in the list (i.e., the cue appeared 200-ms after the offset of the last word in seven-item lists, 1000-ms after in six-item lists, and 1800-ms after in five-item lists).

The instructions for recall were exactly the same as in Experiment 4, with the confidence of each response being indicated via two rows of boxes on the response sheet. The number of boxes in each row always equalled the number of words that were presented in that trial and all 14 words in the experimental vocabulary were permanently on display. Unlike Experiment 4 however, there was a separate response sheet for each trial. Initially, all response sheets were face down in a pile on the left of the subject. When subjects saw the cue for recall, they turned over the top response sheet from the pile, wrote their responses, and then put the response sheet face down in a pile on their right. In this way, subjects did not know in advance the length of the list on a given trial in the Variable condition until starting recall, and could not see the responses they gave in the previous trial.

Subjects received eight practice trials, two of five words, two of six words, two of seven words, and two of six words, in that order. The order of the Fixed and Variable conditions that followed was alternated across subjects. The experiment took 45 minutes.

### Results

In brief, terminal protrusions were more frequent than absolute protrusions in both the Variable and Fixed conditions, irrespective of whether guesses were included or excluded. In fact, protrusions were the most common intrusion on all six positions in the Fixed condition, and were less likely to be guesses than other types of intrusion. The results were generally a close analog to those of Experiment 4.

## Overall Performance

A greater proportion of six-item lists were correct in the Fixed condition ($M=.58$, $SD=.28$) than Variable condition ($M=.45$, $SD=.31$), a difference that was significant under weighted log-odds, $Z(30)=4.65$, $p<.0001$. As expected, the corresponding proportion was higher for five-item lists in the Variable condition ($M=.80$, $SD=.20$) and lower for seven-item lists in the Variable condition ($M=.20$, $SD=.20$). The advantage of fixed-length lists was apparent over all positions, except perhaps the last (Figure 6-3; guesses included). These

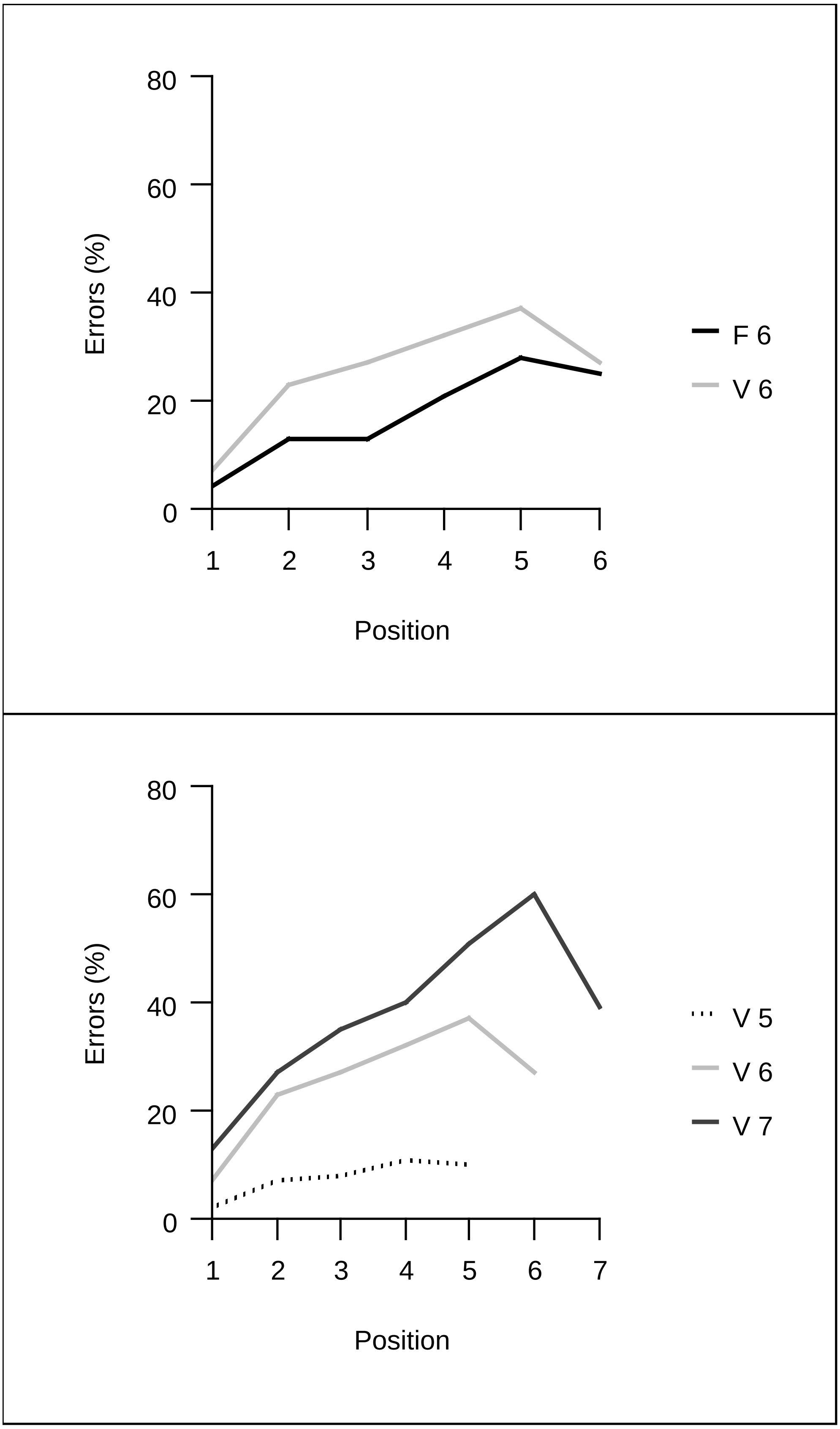



Figure 6-3: Errors by position for six-item lists in Fixed and Variable conditions (upper panel) and five-, six-, and seven-item lists in the Variable condition (lower panel) of Experiment 5. (F=Fixed; V=Variable.)

differences could be attributable to spontaneous grouping in threes, for which there was more evidence in the Fixed than Variable condition. Interestingly, recency remained strong in the Variable condition, even though the last item was not known in advance.[2]

**Errors on Critical Positions**

A three-way ANOVA on the log-odds of an error on either critical position showed a significant effect of condition, *F(1,203)=4.54, p<.05*, guesses, *F(1,203)=174.55, p<.001*, and position, *F(1,203)=7.35, p<.01*. There were no significant interactions, *F(1,203)<1.10, p>.30*. Errors were more frequent on the fifth position than final position, indicating a recency effect, and more frequent in the Variable than Fixed condition. Not surprisingly, excluding guesses reduced the frequency of errors (i.e., remaining transpositions and intrusions, Table 6-5).

| | Variable | | Fixed | |
|---|---|---|---|---|
| | Fifth | Final | Fifth | Final |
| Guesses Included | .32<br>(.18) | .25<br>(.17) | .27<br>(.23) | .25<br>(.22) |
| Guesses Excluded | .09<br>(.08) | .07<br>(.05) | .08<br>(.08) | .07<br>(.07) |

Table 6-5: Frequency of errors on critical positions in Experiment 5.

The proportion of errors on critical positions that were protrusions from the same critical position in the previous report was calculated for the 17 subjects who made at least one error per critical position with guesses excluded. A three-way ANOVA on log-odds showed significant effects of guesses, *F(1,112)=6.48, p<.05,* and position, *F(1,112)=35.68, p<.05,* but neither an effect of condition, *F(1,112)=3.41, p=.07,* nor any interactions, *F(1,112)<2.66, p>.11*, quite reached significance. Protrusions were more frequent on final than fifth positions, and the effect of excluding guesses was to increase the frequency of protrusions, mainly on the final position (Table 6-6). Four pairwise comparisons on weighted log-odds confirmed that a significantly greater proportion of errors were terminal rather than absolute protrusions in both conditions, whether or not guesses were included, *Z(17)>2.14, family-wise p<.05.*

2. This is in contrast to Bunt (1976), who located the advantage of fixed length lists mainly on later positions. (Crowder, 1969, located the advantage on early positions, but with free rather than serial recall.)

| | Variable | | Fixed | |
|---|---|---|---|---|
| | Fifth | Final | Fifth | Final |
| Guesses Included | .13 (.09) | .20 (.14) | .16 (.13) | .22 (.14) |
| Guesses Excluded | .12 (.13) | .26 (.25) | .16 (.17) | .32 (.24) |

Table 6-6: Proportion of errors that were protrusions in Experiment 5.
(Calculated from weighted log-odds, *n=17*.)

**Immediate Intrusions**

The previous analyses demonstrated that the proportion of errors that were protrusions was greater between final positions than between fifth positions. However, there were also fewer errors on final positions than fifth positions, potentially affecting the proportion that were protrusions. To overcome this problem, and compare the proportion of protrusions with chance levels, the following analyses restricted errors to immediate intrusions from the preceding report (with guesses included). The frequency of immediate intrusions, and the proportion that were protrusions, was calculated for the 22 subjects who made at least one immediate intrusion per critical position (Table 6-7).

A two-way ANOVA on the log-odds of a protrusion showed an effect of position, $F(1,63)=8.53$, $p<.005$, but no effect of condition, or interaction, $F(1,63)<1$, $p>.38$ in both cases. The effect of position was confirmed by two pairwise comparisons on weighted log-odds, which showed a significantly greater proportion of intrusions were protrusions between final than fifth positions in both conditions, $Z(22)>2.44$, $p<.05$. These differences did not owe to differences in the overall incidence of immediate intrusions, for which an ANOVA showed no significant effects of position, condition or interaction, $F(1,63)<1$, $p>.40$ in all cases.

The proportion of immediate intrusions that were protrusions was also compared to that expected by chance. In the Fixed condition, the chance proportion was .17 (given that an intruding item could come from one of six positions in the previous report); the proportion of protrusions was significantly greater than this on both critical positions, $Z(22)>4.75$, $p<.0001$. In the Variable condition, the chance proportion was not so clear (given that the previous

| | Variable | | Fixed | |
|---|---|---|---|---|
| | Fifth | Final | Fifth | Final |
| Immediate Intrusions | .20 (.07) | .18 (.08) | .20 (.09) | .20 (.09) |
| Protrusions | .27 (.19) | .37 (.23) | .30 (.23) | .42 (.24) |

Table 6-7: Frequency of immediate intrusions, including guesses, and proportion that were protrusions in Experiment 5.
(Calculated from weighted log-odd*s*, *n=22*.)

report could contain five, six or seven items). Using the same figure of .17 as for the Fixed condition, the proportion of protrusions was significantly greater than chance on both critical positions, *Z(22)>3.55, p<.0005*. Even taking a more conservative estimate of chance of .20 (as if the previous report contained only five items), proportions of protrusions on both the fifth position and the final position were still greater than chance, *Z(22)=2.15, p<.05*, and *Z(22)=5.03, p<.0001*, respectively.

Finally, protrusions on the first position of a report in the Variable condition were predominantly from the first position of the previous report. Indeed, as a proportion of immediate intrusions (*M=.57, SD=.34*), they were significantly more frequent than the chance level of *.20, Z(23)=5.51, p<.0001*. Protrusions respected both terminal positions of reports.

**Immediate Intrusions in the Fixed Condition**

Immediate intrusions on all six positions in the Fixed condition, including guesses, were collapsed over subjects (Figure 6-4). Unlike Experiment 3, intrusions increased towards the end of reports (probably because subjects had to guess rather than omit in the present experiment). Otherwise, the data replicated those of Experiment 3, with protrusions being the most common intrusion for all six positions, and the proportion of immediate intrusions that were protrusions being greatest at the start and end of reports. There was some evidence for spontaneous grouping of the six items into two groups of three (e.g., many protrusions on Position 6 came from Position 3 of the previous report). This probably explains why the intrusion gradients are not as smooth as in Experiment 3. In fact, several subjects reported it

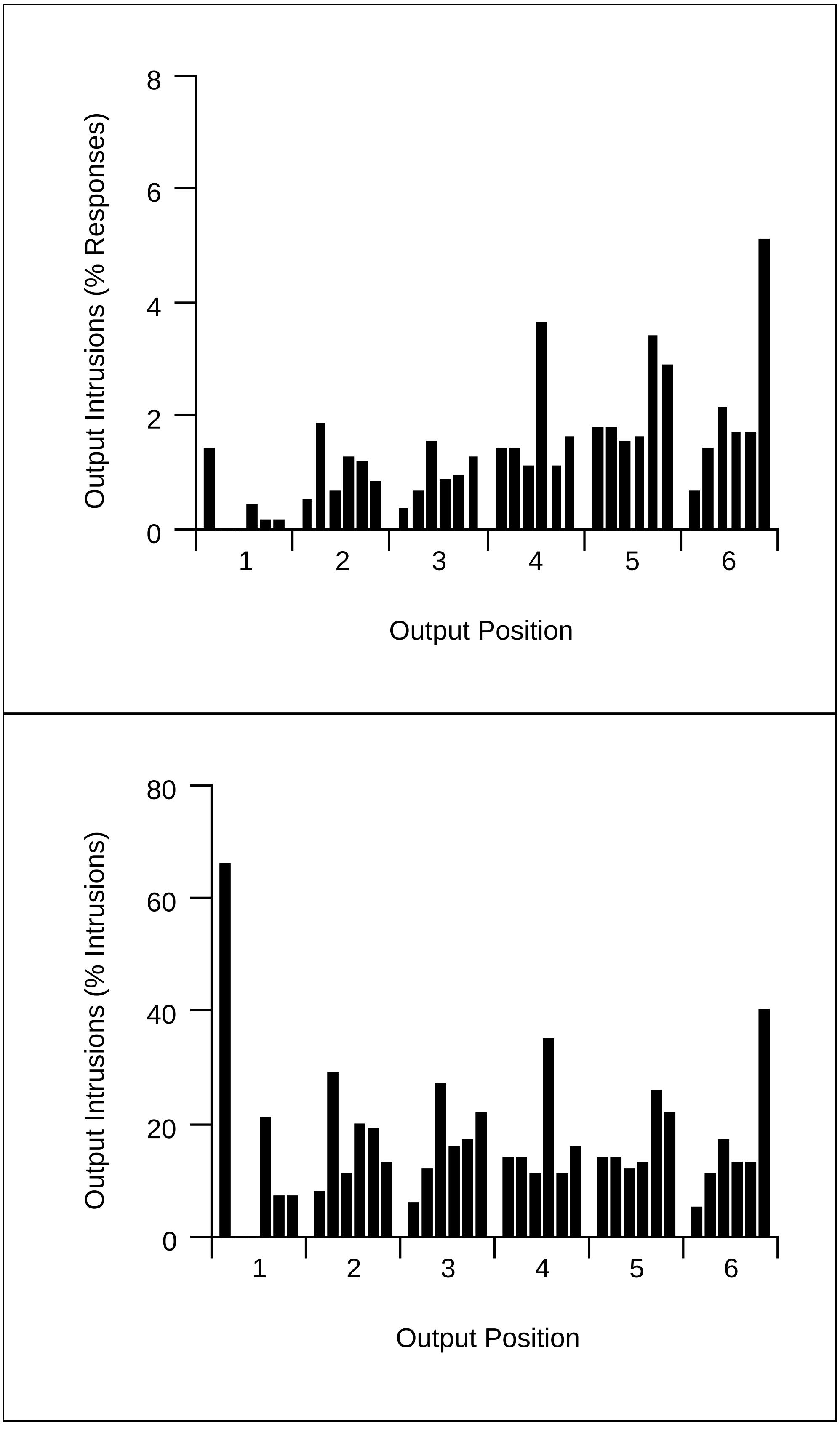



Figure 6-4: Output intrusions as a proportion of responses (upper panel) and as a proportion of intrusions per output position (lower panel) in Experiment 5.

"easier to find a rhythm" in the Fixed than Variable condition. Grouping was therefore less likely to affect the pattern of protrusions in the Variable condition.

The proportion of errors that were immediate intrusions with guesses (*M=.48, SD=.12*) was greater than without guesses (*M=.34, SD=.19*), a difference that was significant under weighted log-odds, *Z(30)=6.14, p<.0001*. This was in contrast to the proportion of errors that were protrusions with (*M=.17, SD=.09*) and without (*M=.16, SD=.13*) guesses, which did not differ significantly, *Z(30)=0.72, p=.47*. (Indeed, the proportion on final positions actually increased; Table 6-6). These results imply that subjects were more confident of an intrusion from the same position in the previous report than other types of intrusion.

**Guesses**

Not surprisingly, the proportion of responses that were guesses was greater in the Variable condition (*M=.22, SD=.20*) than Fixed condition (*M=.17, SD=.17*), *Z(30)=1.98, p<.05*. Again, the frequency of guesses increased towards the end of recall, paralleling the similar increase in omissions in Experiment 3, with a one-way ANOVA in the Fixed condition showing a significant effect of output position *F(5,179)=24.60, p<.001*.

As in Experiment 4, responses in the Fixed condition were split by whether or not they were correct and whether or not they were guesses. A combined test of significance of contingency tables showed an extremely high correlation between accuracy and confidence of responses, *Z(30)=39.46, p<.0001* (Table 6-8). Nevertheless, a considerable proportion of errors were not indicated as guesses (*M=.37, SD=.11*), reinforcing the conclusion of Experiment 4, that subjects are often unaware of errors, and a similar proportion were correct (*M=.37, SD=.15*), reinforcing the conclusion that guesses were more than random choices of list items (of which only .17 would be correct).

| | Not Guess | Guess |
|---|---|---|
| Not Error | 6347 | 505 |
| Error | 502 | 926 |

Table 6-8: Number of guesses and errors collapsed across subjects in the Fixed condition of Experiment 5.

**Discussion**

The present experiment showed that protrusions between reports respect terminal position rather than absolute position. This was demonstrated in the Variable condition, where the proportion of errors that were protrusions was significantly greater between the final positions of reports than between the fifth positions of reports, even when the length of lists was varied from trial to trial in an unpredictable manner. This suggests that position in a report is coded relative to both the start and end of that report, again confirming the prediction of SEM, and questioning the absolute positional information assumed in other models (e.g., Brown et al., 1996; Burgess & Hitch, 1996b).

Protrusions were also more probable between the final position of reports than the fifth position of reports in the Fixed condition. SEM predicts this because the positional uncertainty is smaller for end items, where positional coding is particularly sharp (Chapter 5). SEM also explains why the proportion of absolute protrusions, though less than that of terminal protrusions, was still greater than chance in the Variable condition: There is still considerable positional overlap between the cue for the fifth position in, say, a list of seven items and the fifth position in a list of five items (as shown in Fit 8 below).

The present results also showed that protrusions are not simply the result of guessing strategies. Removing guesses had little effect on the proportion of errors that were protrusions. Removing guesses did reduce the proportion of errors that were intrusions however. In other words, intrusions were particularly likely to be guesses, in agreement with Dillon and Thomas (1975). One reason why intrusions were not removed completely might be that some subjects were not bothering to indicate all their guesses. Alternatively, guesses may be more likely to be intrusions than transpositions simply because there is a greater chance of a guess being an intrusion than a transposition, particularly if subjects tend not guess an item they have already recalled (Chapter 7). In other words, guesses might have a higher baseline chance of being intrusions than other types of error. Both these possibilities are consistent with the distribution of guesses and errors in Table 6-8, and, unfortunately, the present experiment provides no way of clarifying this situation. What is clear is that intrusions that maintain relative position were less affected by the removal of guesses than other intrusions. This corresponds to sharpening intrusion gradients by reducing the noise from completely random guesses (Experiment 3).

Present results represent an important replication of the output protrusions found in Experiment 3 (and Page & Norris, 1996a). Output protrusions have now been shown to occur both with and without vocalisation, with both spoken and written recall, and with immediate as well as delayed recall, confirming that positional information is ubiquitous in serial recall (Chapter 4). Moreover, the results from the Fixed condition suggest that positional coding can extend over six positions, as well as the five in Experiment 3 (though there was also evidence for spontaneous grouping of the six items in the Fixed condition). In any case, the fact that protrusions were found in immediate, serial recall of visually presented lists (a task assumed to rely predominantly on the phonological loop; Baddeley, 1986) further questions the sufficiency of the Primacy Model (Page & Norris, 1996b) as a model of immediate, serial recall. Moreover, in as far as recall of six words in the Fixed condition was within most students' spans (58% of lists being recalled correctly), this finding also questions the assumption of Tehan and Humphreys (1995). According to these authors, immediate recall is immune to proactive interference. On the contrary, proactive interference (of a positional kind) acts even on immediate serial recall of phonologically-coded, span-length lists (i.e., proactive interference is a matter of degree, rather than all-or-none).

Somewhat ironically however, the confirmation of terminal protrusions in a situation where the end of the list is unpredictable does not help interpretation of the end marker of SEM. As in Experiment 4, present results rule out any interpretation where the positional cue changes constantly over time or position, as with the internal oscillators of Brown et al. (1996) and Hitch et al. (1995). However, the present results are also problematic for an interpretation in terms of expectation for the end of a list (Chapter 5). Though an expectancy interpretation might explain better performance on six-item lists in the Fixed condition than in the Variable condition, it seems incompatible with terminal protrusions, which require an end marker that grows towards the end of list in a manner independent of list length, even when the list length is not known in advance. This issue is resumed in the General Discussion.

## General Discussion

The present experiments demonstrated that substitutions between sequences of different lengths tend to maintain terminal positions rather than absolute positions, whether those sequences are reports on different trials, or groups within the same trial. These findings suggest that position within a sequence is coded with respect to both the start and the end of that sequence, confirming the prediction of the Start-End Model (Chapter 5).

The present experiments also demonstrated that positional errors are not simply the result of guessing strategies. Neither protrusions nor interpositions were any more likely to be guesses than other errors. This supports SEM's assumption that positional errors result from competition amongst responses for a particular position, rather than guesses after the correct item has been forgotten (c.f., Conrad, 1960; Dillon & Thomas, 1975). The fact that significant numbers of errors were not indicated as guesses suggests that this response competition may operate at an unconscious level, supporting the observation that people are often unaware of errors (Chapter 1). Nevertheless, it is clear that people do sometimes resort to conscious guessing when no response comes to mind, particularly in the present experiments where they had to give a response for every position. Such guesses were more accurate than would be predicted by random choices from the experimental vocabulary. The guesses in Experiment 5 were also likely to be intrusions, in agreement with Dillon and Thomas (1975), and Bjork and Healy (1974). Incorporating a role for guessing, in addition to response competition, allows SEM to explain these somewhat paradoxical results (below).

The present results support the assumption that position is coded by markers at the start and end of sequences. The start marker can explain the positional errors between the start of groups (Experiment 4) and the start of reports (Experiment 5); the end marker can explain the positional errors between the end of groups and the end of reports. Nevertheless, the important question remains: What are the psychological correlates of these markers? The start marker needs only be triggered by the first item in a sequence and could depend on any psychological variable that decreases monotonically during subsequent items, such as attention. The end marker on the other hand needs to grow steadily towards the end of a sequence. When the length of a sequence is known in advance, as for the different size groups in Experiment 4, the end marker could quite plausibly represent expectation for the end of the group. When the length of a sequence is unknown however, as in the Variable condition of Experiment 5, such an expectancy interpretation becomes less plausible, particularly when protrusions remain between the ends of those sequences. Two possible solutions to this problem are given below.

**Positional Codes Generated during Rehearsal**

Positional errors between sequences are clearest when measured with respect to output position (Experiment 3; Chapter 4). In other words, they are clearest when the previous sequence reflects a recall episode rather than a presentation episode (e.g., recall of the previous group in Experiment 4, or recall of the previous list in Experiment 5). Because recall episodes are normally more recent than presentation episodes, this finding is not on its own a problem for SEM, which assumes less general contextual change for more recent episodes (Chapter 5). However, because presentation and recall are confounded with recency in this way, and because output position and input position are normally highly correlated (given that responses are usually correct), it is difficult to determine the relative influence of previous presentation episodes and previous recall episodes. Indeed, it remains possible that only recall episodes are the source of positional errors. In other words, the interpositions in Experiment 4 and the protrusions in Experiment 5 may be explained solely by proactive interference from positional codes generated during recall of the previous group or previous list. By extending the notion of recall to any form of overt or covert rehearsal, this hypothesis can even explain anticipations from later groups during recall of earlier groups (Experiment 2): Any rehearsal of the later groups before recall begins may be sufficient to generate positional codes for items in those groups, and cause interpositions during recall of earlier groups.

This *rehearsal hypothesis*, that positional codes are only generated during rehearsal, and not during presentation, has the advantage of making interpretation of the end marker easier. Because the length of a sequence is known at recall, expectation remains a possible psychological correlate. The disadvantage of this hypothesis is that it begs the question of how the order of items is stored before rehearsal begins: If positional codes are only established during rehearsal, they can not be used to order the very first rehearsal. An alternative means of ordering items is required. Thus, if the rehearsal hypothesis were confirmed, SEM would no longer be sufficient as a model of serial recall.

There is circumstantial evidence against the rehearsal hypothesis. For example, several studies have demonstrated that people extract positional information even under incidental learning (e.g., Hintzman, Block & Summers, 1973; Toglia & Kimble, 1976; Nairne, 1991; though not as well as under intentional learning, Navey-Benjamin, 1990; Tzeng, Lee & Wetzel, 1979). However, these demonstrations used long lists and considerable delays before recall. More relevant to serial recall from short-term memory is an experiment by Estes (1991). In a condition where subjects rehearsed lists overtly during the retention interval, Estes showed that about 70% of intrusions in a rehearsal protocol were likely to be recalled at the same position in recall, in agreement with the rehearsal hypothesis. However, he also showed that about 57% of items that did not occur at the correct position in a rehearsal protocol did occur at the correct position in recall (i.e., at the same position as in the original presentation). This led Estes to propose that there are two sources of positional information, a “direct” one from the presentation episode and an “indirect” one from rehearsal episodes. (SEM can also explain this data with its assumption that every rehearsal of an item creates a new token, without the need to postulate different sources per se.) In other words, Estes assumed positional information can be generated during presentation as well as rehearsal, contrary to the rehearsal hypothesis.

However, though Estes’s data suggest that two sources of information influence recall, they do not actually require both to be positional. Estes’s indirect source may be positional, but his direct source need only be ordinal in order to explain why correct responses can occur in spite of incorrect positional information from the indirect source. M. P. A. Page (personal communication, 1995) observed that Nairne’s (1991) data may similarly be explained by use of ordinal rather than positional information. Better evidence would come from positional errors in situations where input and output position are not positively correlated. One possibility is to use backwards recall of lists or groups (Henson, 1995), where input and output position are negatively correlated, though the processes underlying backwards recall remain unclear (Chapter 5). Another possibility is to use a part-list recall paradigm. Lee and Estes

(1981) for example showed that interpositions occurred even when recall of only one of three groups was required. However, subjects did not know in advance which group was to be recalled and so were likely to rehearse all three groups, perhaps allowing positional codes to be generated. A better approach would be to require subjects to only ever recall the first of two groups (so there is no reason to rehearse the second). If erroneous items in the first group still tended to come from the same position in the second group, such retroactive interference of positional information would refute the rehearsal hypothesis.

Finally, preliminary evidence against the rehearsal hypothesis was obtained in a recent pilot experiment by Page & Norris (1996a). Using part-list recall of one of two groups, they found evidence for positional errors even under conditions of articulatory suppression (during both presentation and recall, and with both visual and auditory presentation). In fact, positional errors seemed more prevalent than usual, and yet articulatory suppression should preclude, or at least minimise, rehearsal (Baddeley, 1986). Further experiments are required to confirm these findings and test the rehearsal hypothesis more rigorously.

**Positional Codes Generated during Presentation**

Given no conclusive evidence for the rehearsal hypothesis, and preliminary evidence against it, SEM's assumption that position is automatically coded during presentation, rehearsal and recall will be maintained. The question remains however as to how position is coded relative to the end of a sequence, when the end of a sequence is unknown in advance. Several possible solutions are outlined below.

One possibility is that the end marker does not grow in strength until the very last item, when the end of the sequence is finally confirmed. A similar suggestion was made by Houghton (1990), whose end node was only triggered by termination of a sequence. By assuming further that presentation of items left them transiently activated in memory, the triggering of the end node allowed it take a "snap-shot" of a recency gradient of decaying activations of the last few items in the list. By growing in strength more gradually during recall, this allowed the end node to exert an influence on items earlier in the list. The problem with this solution however is that it does not allow positional tokens to be created until the end of presentation. This is contrary to the assumption of SEM, that a position-sensitive token is generated as soon as each new item is presented.

A more suitable approach within SEM is that, when the length of a sequence is unknown, the end marker might grow in a fixed manner. This growth is irrespective of exact sequence length (though in a manner that might make some allowance for the expected range of lengths). For example, on the very first trial, the strength of the end marker might grow by a constant amount during presentation. (Alternatively, no end marker might be employed, given that SEM can still recall a sequence correctly without an end marker; Chapter 5). During subsequent trials however, when a subject has induced the range of possible list lengths, the end marker might grow more quickly, reaching its maximum value as soon as it is possible for the list to end (e.g., after the fifth item in the Variable condition of Experiment 5). It might then stay at that maximum value during any remaining items in the sequence, reflecting the subject's expectation that the sequence will end soon.

During recall however, when the length of the sequence is known (particularly if the correct number of boxes are provided for recall), the end marker can behave as previously assumed, growing continually towards the end of recall, and reflecting more accurate expectation for the end of the report. Because the end marker behaves differently during presentation and recall, there will be a greater positional uncertainty for the last few items in the list. This is consistent with the finding in Experiment 5, that six-item lists are recalled less well in the Variable condition than the Fixed condition. However, there will still be considerable overlap between the cue for the last position and the token created at the last position in the previous report (Chapter 5). Thus, protrusions will still be more likely between terminal positions than absolute positions when measured with respect to output position.

In summary, it is possible to maintain an expectancy interpretation of the end marker, with a hazy notion of expectancy during presentation that is refined during recall. A hazy notion of expectancy can be modelled by an end marker that grows in a constant manner, irrespective of exact list length. Though this entails the end marker behaving differently during presentation and recall, the more accurate recoding of positions during recall still predicts significant numbers of terminal protrusions from the previous report. This was the approach taken below, in fitting SEM to data from the present experiments.

**Fits of the Start-End Model**

SEM was fitted to data from Experiment 4 and Experiment 5, including modelling the effect of guesses.[3] This entailed an additional assumption about the nature of guesses.

**Modelling Guesses**

In the present experiments, subjects were asked to indicate both explicit guesses and uncertain responses. In SEM, explicit guesses are assumed to arise after an item's phonological representation has been retrieved. If the activation of that representation does not exceed a guessing threshold, $T_G$, then an item is guessed instead. (In previous experiments, omissions may have resulted in such cases, which were modelled by activations below the omission threshold, $T_O$; Chapter 5.) A guess is chosen by competition amongst phonological representations, on the basis of their current activation and suppression, together with additive Gaussian noise with standard deviation $G_G$ (Appendix 3). Thus guesses are random choices from the set of recently perceived items, with a bias against those recently recalled, owing to suppression. The bias against items recently recalled reflects the fact that people are often reluctant to repeat themselves (Chapter 7). Because phonological activations tend to be lower for later items, guesses will increase towards the end of recall, in a manner similar to omissions (Chapter 5). Furthermore, with guesses predominantly at the end of recall, suppression of previous responses will tend to preclude most list items, meaning that many guesses will be intrusions (Experiment 5).

However, not all responses indicated as guesses in the present experiments were likely to be explicit guesses. A considerable number may have been items that came to mind, but not readily enough for subjects to be certain of them. If subjects were obeying instructions, these responses would be indistinguishable from "true" guesses in the present experiments (see Experiment 8). Such uncertain responses can be modelled as phonological activations that do not exceed an uncertainty threshold, $T_U$ (Appendix 3). Because most such responses, though weakly active in memory, will nevertheless be correct, a significant proportion of responses indicated as guesses will be correct (Experiments 4 and 5). When simulating the removal of

3. Modelling guesses seems more appropriate than "correcting" the data for guesses (e.g., Sperling & Melchner, 1976; Drewnowski, 1980a), because the latter would seem impossible to achieve in an atheoretical manner.

uncertain responses in SEM, $T_U$ is set above $T_G$, and any item whose phonological activation drops below $T_U$ is removed. Thus, the two new parameters, $T_G$ and $T_U$, replace the old parameter $T_O$. By setting $T_G>0$ and $T_U=0$, SEM simulates the inclusion of guesses; by setting $T_U>T_G>0$, SEM simulates the exclusion of guesses.

**Fit 7: Interpositions in Variable Groups**

The multiple-trial version of SEM was fitted to all three conditions of Experiment 4, without and without guesses. Most parameters, such as $C_P$, $C_D$, $C_R$, and $C_I$, were fixed by the experimental design (Appendix 3). Specifically, $C_P=C_R=1$ reflected contextual change and phonological decay during presentation and recall of each item, and $C_D=C_I=0$ reflected the immediate recall and unfilled intertrial interval. The parameters $G_G=G_P=0.30$ and $T_O=0.00$ were fixed. This left 5 free parameters, eventually set to $D_G=0.10$, $M_G=0.95$, $G_C=0.06$ and $T_G=0.90$, while $T_U=0.00$ or $T_U=1.10$ was varied to fit the inclusion or exclusion of guesses. Remaining parameters were the same as in Fit 5.

Two simulations of SEM were run on the 720 lists given to subjects in the Ungrouped condition, one with $T_U=0.00$, to simulate the inclusion of guesses, and one with $T_U=1.10$, to simulate the exclusion of guesses. Responses were then split by whether or not they were correct and whether or not they were guesses (Table 6-9; cf. Table 6-4). Of the guesses, .28 were correct, and of the errors, .41 were not guesses (the corresponding figures over 100,000 trials were .30 and .41 respectively). These figures are close to those in Experiment 4 and Experiment 5, supporting the assumption that guesses include both explicit guesses (below $T_G$) and uncertain responses (below $T_U$). In other words, SEM's treatment of guesses appears to provide a reasonable approximation of subjects' behaviour.

| | Not Guess | Guess |
|---|---|---|
| Correct | 2670 | 447 |
| Error | 781 | 1142 |

Table 6-9: Number of guesses and errors in Ungrouped condition from SEM in Fit 7.

Six further simulations were run on 100,000 copies of the same lists given to subjects, to fit each condition with and without guesses. With guesses included, SEM recalled .17 of lists in the Ungrouped condition correctly, .38 of lists in the Grouped 3-4 condition, and .37 in the Grouped 4-3 condition. With guesses excluded, the corresponding figures were .06, .26 and .26 respectively. Overall performance was therefore reasonably matched to the data, though slightly worse for ungrouped lists. This is probably attributable to spontaneous grouping in the Ungrouped condition of Experiment 4, as is common with supraspan lists (Chapter 3). SEM also reproduced the error position curves, with an RMSE over 42 data points of 7.22% (Figure 6-5).

The proportion of errors on critical positions that were interpositions showed a reasonable quantitative fit to the data (Table 6-10; cf. Table 6-2), though the pattern was more pronounced in the model than the data, particularly without guesses. This may reflect more noise in the data than was captured by SEM's assumption about uncertain responses (and, in the ungrouped case, the presence of spontaneous grouping in the data). Nonetheless, the RMSE of *10.58%* over the 16 (untransformed) data points was not a reliable difference, given the variability in the data, $T^2=62.43$, $F(16,6)=0.52$, $p=.86$. The most important point was that SEM reproduced the significant aspect of the data, that terminal interpositions were more frequent than absolute interpositions.

| | Grouped | | Ungrouped | |
|---|---|---|---|---|
| | Third | Final | (Third) | (Final) |
| Grouping 3-4 | | | | |
| Guesses Included | .12 | .28 | .08 | .14 |
| Guesses Excluded | .08 | .38 | .02 | .06 |
| Grouping 4-3 | | | | |
| Guesses Included | .18 | .23 | .14 | .13 |
| Guesses Excluded | .13 | .31 | .06 | .06 |

Table 6-10: Proportion of errors that were interpositions from SEM in Fit 7.

Including guesses and collapsing across the grouped conditions, as in Experiment 4, between-group repetitions were slightly more common than in the data. Nonetheless, the

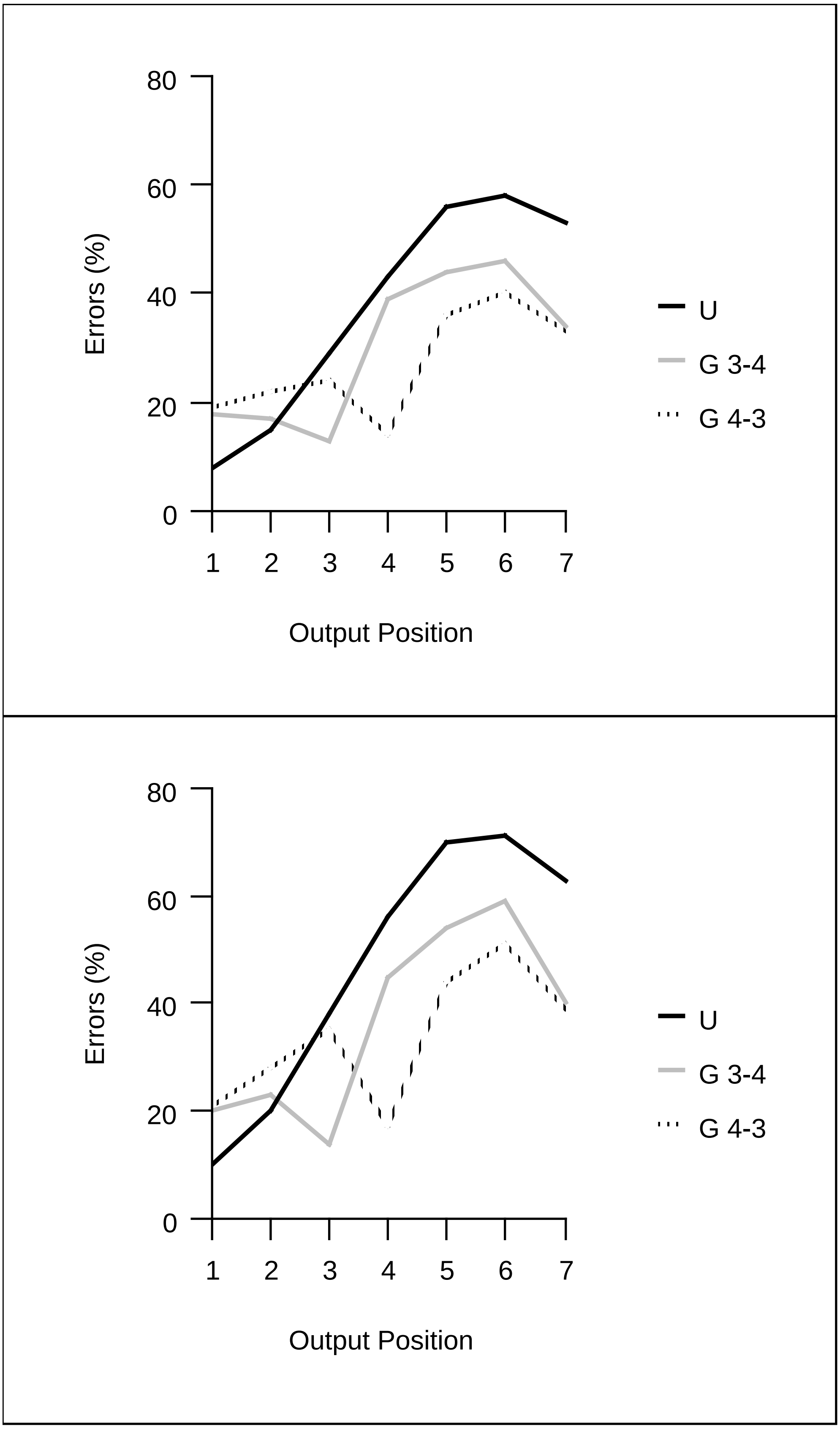



Figure 6-5: Errors by position with guesses (upper panel) and without guesses (lower panel) from SEM in Fit 7.

(U=Ungrouped, G=Grouped.)

proportion that were interpositions was very similar to the data (Table 6-11; cf. Table 6-3). The RMSE over the 8 (untransformed) data points was *8.58%*, which was extremely good, given the noise associated with the small numbers in the data. This pattern of repetitions between groups stems from SEM's assumption that each response is recoded in its output position (Chapter 5). When an item is recalled at the end of the first group, the new token created will be strongly cued again at the end of the second group. Not only does it share the same code for within-group position, but its more recent encoding means its general context will overlap more with the recall context than other, as yet unrecalled tokens. Thus, providing the token is cued strongly enough to overcome the suppression of its type representations a few responses earlier, its repetition at the end of the second group is quite likely.

| | Grouped | | Ungrouped | |
|---|---|---|---|---|
| | Third | Final | (Third) | (Final) |
| Between-group Repetitions | .26 | .25 | .23 | .28 |
| Interpositions | .23 | .41 | .26 | .25 |

Table 6-11: Frequency of between-group repetitions and proportion that were interpositions from SEM in Fit 7.

**Fit 8: Protrusions in Variable Lists**

The multiple-trial version of SEM was fitted to both conditions of Experiment 5, without and without guesses. All parameter values were identical to the Ungrouped condition of Fit 7, except the value $G_C=0.01$. This one degree of freedom was to accommodate differences in the experimental procedure. A new parameter $N_M$ was also introduced. The value $N_M=5$ was fixed by the experimental design and reflected the minimum list length expected by subjects in the Variable condition (given that lists varied from five to seven items). This meant that the end marker coding the positions of items in the list grew exponentially to a value $E_{0,I}=0.60$ during presentation of the fifth item, and then stayed constant at that value during presentation of any further items (Appendix 3). During recall, when the list length was known, the end marker behaved as normal (Figure 6-6).

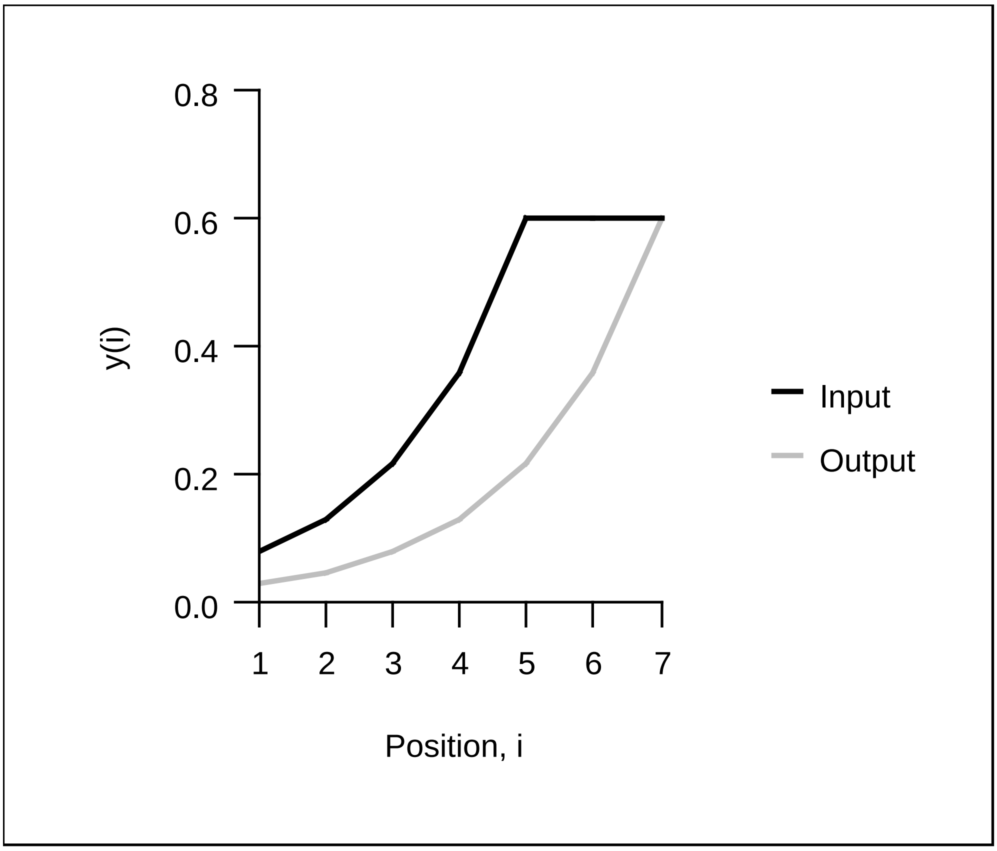



Figure 6-6: Strength of the end marker during presentation (input) and recall (output) of an unpredictable seven-item list in the Variable condition of Fit 8.

Four simulations were run on 100,000 copies of the same lists given to subjects, to fit each condition with and without guesses. In close agreement with both Fit 7 and the data, .34 of guesses were correct and .38 of errors were not guesses. With guesses included, SEM recalled .57 of lists correctly in the Fixed condition, and .80 of five-item lists, .35 of six-item lists and .04 of seven-item lists in the Variable condition. With guesses excluded, the corresponding figures were .33, .64, .18 and .01. Some of these figures were lower than in the data, but the important trends were present, including better performance on six-item lists in the Fixed condition than the Variable condition. SEM also produced similar error position curves, with an RMSE over 48 data points of *10.56%* (Figure 6-7).

In agreement with the data, removing guesses from the Fixed condition decreased the proportion of errors that were intrusions, from .38 to .26, but not the proportion that were protrusions, which increased slightly from .18 to .20. Thus SEM's assumptions about guessing

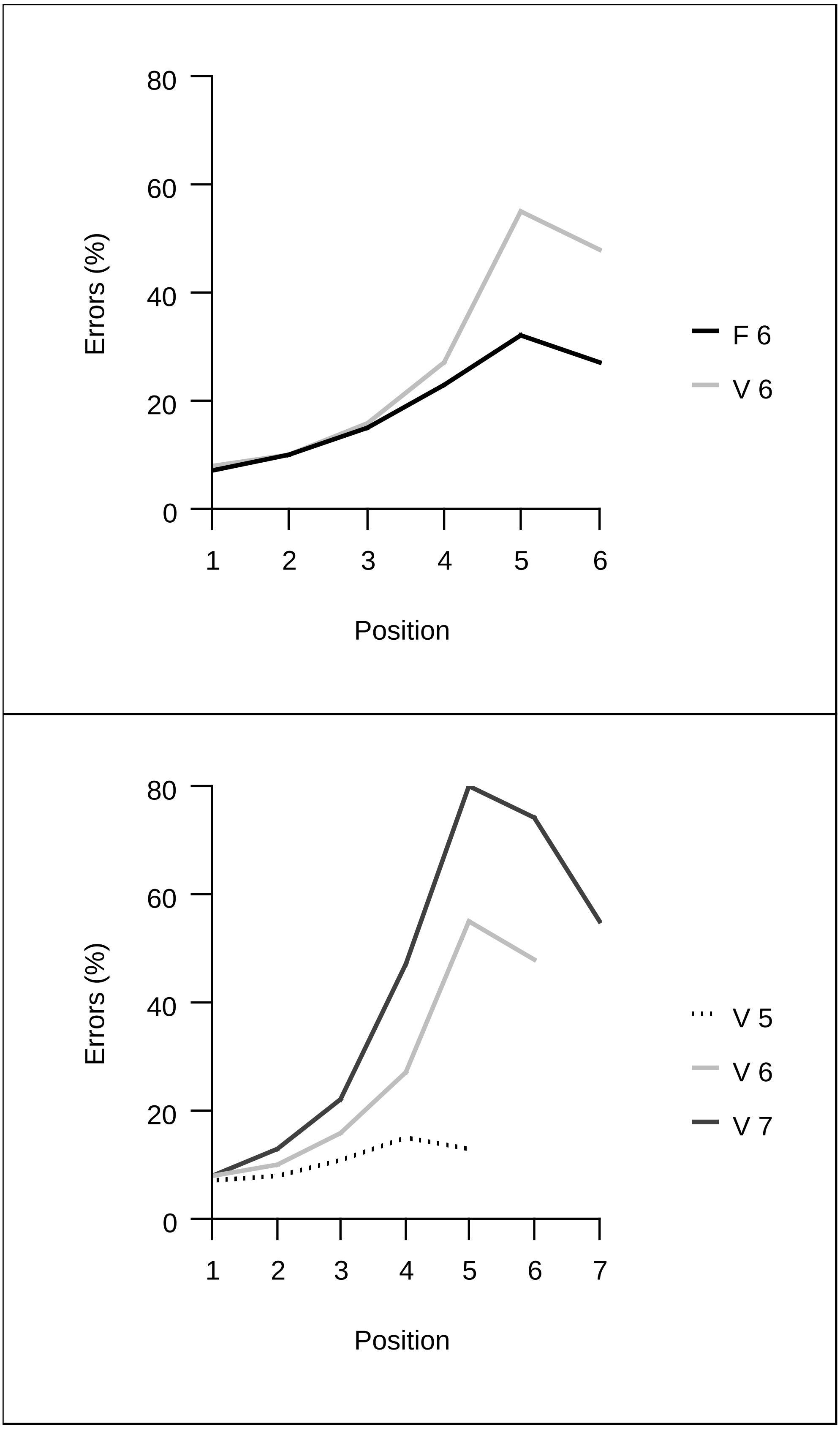



Figure 6-7: Errors by position for six-item lists in Fixed and Variable conditions (upper panel) and five-, six-, and seven-item lists in the Variable condition (lower panel) from SEM in Fit 8. (F=Fixed; V=Variable.)

can explain these somewhat paradoxical results, by producing intrusions that arise from both random guesses (most intrusions) and from response competition (mainly protrusions).

The frequency of immediate intrusions, including guesses, was slightly lower than in the data. Nevertheless, the proportion that were protrusions showed a reasonable quantitative fit (Table 6-12; cf. Table 6-7). The proportion of protrusions on the last position in the Fixed condition was higher than in the data, though again this may reflect the effects of spontaneous grouping in the data. The RMSE over the 8 (untransformed) data points was *6.67%*, a difference that was not reliable given the variability in the data $T^2=0.75$, $F(8,14)=0.16$, $p=.99$. The most important point was that SEM reproduced the significant aspect of the data, that terminal protrusions were more frequent than absolute protrusions, attributable to SEM's assumption that position in a list is coded relative to the end of that list.

| | Variable | | Fixed | |
|---|---|---|---|---|
| | Fifth | Final | Fifth | Final |
| Immediate Intrusions | .13 | .08 | .08 | .13 |
| Protrusions | .22 | .33 | .33 | .56 |

Table 6-12: Frequency of immediate intrusions, including guesses, and proportion that were protrusions from SEM in Fit 8.

In summary, the assumption of an end marker geared to the minimum expected list length provided a reasonable fit to the data, together with a plausible explanation for the general impairment in the Variable condition (though a greater ease of spontaneous grouping in the Fixed condition is equally plausible). The fit was therefore a reasonable first approximation to modelling unpredictable lists in SEM, and demonstrates how an expectancy interpretation of the strength of the end marker might be maintained.

**Future Work**

Though the present finding that positional errors respect terminal positions of sequences suggests that relative rather than absolute position is coded, the concept of relative position includes more than just terminal positions. To demonstrate relative position more generally, it would be necessary to test whether positions in the middle of sequences are also coded relative to the start and the end of those sequences. For example, coding of relative position predicts that the middle item of a sequence of three (Item 2) is likely to substitute with the middle item of a sequence of five (Item 3). Such a finding would confirm that present results reflect more than something special about the first and last item in a sequence. Preliminary support for relative position comes from a study by Banks, White and Mermelstein (1980), who showed that, in judgements of relative order, an item added to the middle of a four-item list immediately behaved like the middle item of a five-item list.

However, SEM does not necessarily predict that relative position is symmetrical with the respect to the start and end of a sequence. The exact overlap between positional codes for middle positions depends on the particular parameter values of the start and end markers. With an end marker weaker than the start marker (Chapter 5), the middle item of a sequence of three might be more likely to substitute with the second item of a sequence of five, because the influence of the end marker on these positions will be less than that of the start marker. Further experiments will therefore not only help clarify the issue of relative position and its symmetry, but may also help determine the relative strengths of the start and end markers.

## Chapter Summary

Two experiments demonstrated that positional errors between sequences of different lengths respect both the start and the end of those sequences, confirming one of the core assumptions of SEM. These errors were not simply position-sensitive guesses. These results were predicted by SEM, and are problematic for all other models of serial recall. Nevertheless, the demonstration of positional errors between the ends of sequences, even when the ends of those sequences are unpredictable, prevents any simple interpretation of the end marker in SEM. Two more subtle interpretations were suggested, one of which was implemented in the model and fitted to the present data. Future work may help clarify the nature of the end marker and, in particular, test 1) whether position is really coded during presentation as well as recall, and 2) whether position is truly relative (i.e., extending to more than just terminal positions). The next chapter examines a different core assumption of SEM, that the order of items in short-term memory is stored over token representations.

# Chapter 7: Item repetition in serial recall

## *Repetition facilitation, inhibition and contamination*

The experiments and simulations in previous chapters used lists of unique items. This chapter describes three experiments that examine serial recall of lists in which an item is repeated. The results of these experiments, together with those described in Henson (1996b), are consistent with SEM's assumptions of positional tokens and suppression of type representations. Nevertheless, the results also highlight additional assumptions required before SEM, or any other model, can fully explain item repetition effects in serial recall.

### Ranschburg Repeated

The presence of a repeated item in a list has important effects on the serial recall of that list. Foremost are effects on the recall of the repeated item itself, the *Ranschburg* effects (Jahnke & Bower, 1986). When two occurrences of an item are close together in a list, recall of both occurrences is generally superior to recall of two different items at corresponding positions in control lists with no repeated items (*repetition facilitation*; e.g., Crowder, 1968a; Lee, 1976b). However, when the two occurrences are separated by a number of intervening items, recall of one or both occurrences is generally inferior to recall of two different items at corresponding positions in control lists (*repetition inhibition*; e.g., Crowder, 1968a; Jahnke, 1969b). Repetition can also affect recall of other, nonrepeated items in a list (*repetition contamination*), particularly those immediately following a repeated item (Wickelgren, 1966). Though these effects have been demonstrated in many previous studies, they have been measured, and interpreted, in several different ways. The main purpose of the present experiments was to attempt a unified measurement and interpretation.

The issue of repeated items in serial recall is important because it raises questions about the representation of items in memory. For example, do two occurrences of a repeated item activate the same type representation in memory, as in Wickelgren's associative theories (Wickelgren, 1969), or does each occurrence form a separate token representation in memory, as in Wickelgren's nonassociative theories? This question is particularly relevant to chaining theory, because associative chaining models predict a detrimental effect of item repetition (given that repeated items are associated with more than one successor; Chapter 1). One way to overcome this problem is to assume nonassociative chaining models, which chain along token rather than type representations (Wickelgren, 1969).

The representational issue is particularly apparent in explicit, computational models, whether or not they employ chaining. For example, the Articulatory Loop Model (Burgess & Hitch, 1992, 1996b) assumes type representations, whereas the Primacy Model (Page & Norris, 1996b) requires token representations. SEM also assumes token representations, though these are linked to a common type representation at output, in order to produce a categorical response (Chapter 5). Surprisingly however, none of these models has directly addressed recall of repeated items; simulations have always assumed lists of unique items.

Models that do address repetition in more detail have generally dealt with long-term memory, such as the typing model of Rumelhart and Norman (1982) and the spelling model of Houghton, Glasspool and Shallice (1994). These models assume special mechanisms for dealing with repetition (particularly immediate repetition). The repetition mechanism of Houghton et al. (1994) is supported by neuropsychological data, and is justified theoretically in order to overcome the temporary suppression of previous actions during sequential behaviour (e.g., Houghton & Tipper, 1996; MacKay, 1987). This may correspond to the suppression of SEM's type representations during serial recall, suggesting one possible explanation for repetition inhibition.

Another reason for the interest in item repetition is in connection with repetition blindness (Kanwisher, 1987). Repetition blindness is the failure to detect or respond to repeated elements in rapidly presented sequences (of around 100 ms per item). Though repetition blindness is unlikely in the present experiments, where presentation rates are 500 ms per item or more, some previous studies have used serial recall to index repetition blindness and have been potentially confounded therefore by repetition inhibition. Indeed, several researchers have suggested that repetition blindness is no more than a memory phenomenon such as repetition inhibition (e.g., Armstrong & Mewhort, 1995; Fagot & Pashler, 1995; Whittlesea, Dorken & Podrouzek, 1995; Whittlesea & Podrouzek, 1995), though others have argued that repetition blindness is a separate perceptual phenomenon (Bavelier, 1994; Kanwisher, Driver & Machado, 1995; Luo & Caramazza, 1995; Park & Kanwisher, 1994). Much of the confusion has resulted from a failure to distinguish type and token representations in the scoring of serial recall. The new scoring scheme introduced below may help resolve some of the confusion.

Repetition facilitation and repetition inhibition are simple to demonstrate in short-term, serial recall, and are robust to experimental manipulations such presentation rate or presentation modality (Mewaldt & Hinrichs, 1973). Consequently, the effects have been subject to considerable research. In order to summarise, the separate occurrences of a repeated item are the *repeated elements* of a *repetition list*. Nonrepeated items at the same positions in *control lists*, containing no repeated items, are the *control elements*. These are the *critical elements*; remaining nonrepeated elements in repetition and control lists are *context elements*.

In a parametric study varying the separation between two critical elements, Crowder (1968a) compared recall on positions of repeated elements with recall on positions of control elements. He demonstrated superior recall of repeated elements one or two positions apart, and inferior recall of repeated elements three or more positions apart. Wickelgren (1965c) showed a similar transition between repetition facilitation and repetition inhibition using item-scoring (i.e., whether a critical element was recalled anywhere in a subject's report) rather than position-scoring. The repetition inhibition stemmed from a failure to recall the repeated item more than once. The importance of recalling both critical elements was reinforced by Lee (1976b). Lee showed that, when estimating the probability that at least one of the critical elements was recalled somewhere, there was no repetition facilitation or inhibition for any separation of repeated elements. Only when estimating the probability that both critical elements were recalled somewhere did the effects arise.

Repetition contamination can also be measured in different ways. Under item-scoring, context elements are usually recalled better in repetition lists than in control lists (Wickelgren, 1965c). Under position-scoring however, an advantage for context elements arises only when repeated elements are close together (i.e., under conditions of repetition facilitation, Crowder, 1968a; Mewaldt & Hinrichs, 1973). These results may be reconciled by a trade-off between two opposing factors, one which increases the recall of context items generally, and another which decreases the probability of recalling them in the correct position.

The general increase in correct recall of context elements in repetition lists may owe to a reduced set of possible responses: If an item is forgotten, there is one less item to guess from in repetition lists than control lists. This can improve recall of context elements under both position- and item-scoring. One factor that might impair positioning of context elements is the presence of *associative intrusions* following repeated elements (Wickelgren, 1966).[1] These are transpositions between the context elements that immediately follow repeated elements (and therefore only occur when repeated elements are separated in the list). When conditionalised on correct recall of the repeated element, Wickelgren found such associative intrusions to occur significantly more often than equivalent transpositions following correct control elements in control lists. Wickelgren used associative intrusions to support an associative chaining theory of serial recall (Chapter 1), stating that "...the prior item is an important cue in short-term memory for serial lists and there is only one representative of an item no matter how many times it is presented." (Wickelgren, 1966, p. 858). One question asked here is whether associative intrusions, together with a guessing bias, are a sufficient account of repetition contamination.

Thus the simple presence of repeated elements can have a number of different effects on serial recall. It is somewhat surprising therefore that theoretical interpretation of these effects remains unclear. Repetition facilitation has been variously attributed to chunking (Wickelgren, 1965c), isolation or distinctiveness (Lee, 1976a), and repetition tagging (Lee, 1976b). Repetition inhibition has been attributed to output interference (Crowder, 1968b; Jahnke, 1969a), proactive interference (Jahnke, 1972b) and guessing strategies (Hinrichs, Mewaldt & Redding, 1973). Repetition contamination has been attributed to associative chaining (Wickelgren, 1966) and grouping strategies (Henson et al., 1996). The present set of experiments aim to resolve some of these issues.

However, all previous studies of these repetition effects have overlooked the type/

1. Though these errors are not really intrusions in the present terminology (they are transpositions and a special type of associate; Chapter 1), Wickelgren's term is maintained for consistency with previous studies. Note that a similar effect of repetition contamination arises in speech errors: the "repeated phoneme effect" (Dell, 1984).

token distinction. As well as being an important theoretical issue, this distinction has implications for the measurement of repetition effects. If the two occurrences of a repeated item are separate tokens in memory, the experimenter, who deals only with categorical responses, has no way of distinguishing them. This presents a problem in the conventional scoring of serial recall.

Fortunately, it is possible to overcome this problem with a new method of scoring control lists. This method is introduced in Experiment 6 and employed in all subsequent experiments. Indeed, the main aim of Experiment 6 was to determine whether repetition facilitation and repetition inhibition remain significant effects under this new method of scoring. Experiment 6 also examined the interaction between these effects and temporal grouping of lists, another theme continued through subsequent experiments. Experiment 7 examined people's ability to detect and remember the repetition of an item, on-line with serial recall (complementing the off-line task in Experiment 2 of Henson, 1996b). There is good evidence that memory for the repetition plays an important role in Ranschburg effects, and the ability to detect and remember repetition may be affected by grouping. Finally, Experiment 8 examined the role of guessing in Ranschburg effects; in particular, to test whether repetition inhibition is more than a bias against guessing repeated items.

## Experiment 6

The main aim of the first experiment was to measure repetition facilitation, inhibition and contamination under a new method of scoring control lists that overcomes any bias in favour of repetition lists. This bias arises because, unlike a transposition between two control elements, a transposition between two repeated elements can not be detected by the experimenter. This bias potentially confounds all previous demonstrations of repetition facilitation and repetition inhibition.

For example, consider recall of a control list represented by the sequence *123456,* and a repetition list represented by the sequence $12R_1R_256$, where $R_1$ and $R_2$ are the two repeated elements. A subject's report of the control list as *124356* would be marked as containing incorrect positioning of the control elements *3* and *4*. However, a subject's report of the repetition list as *12*$R_2R_1$*56* would be marked as containing correct positioning of the repeated elements, because the experimenter has no way of distinguishing $R_1$ and $R_2$. Such a discrepancy in scoring the two types of list might lead to an overestimation of repetition facilitation and an underestimation of repetition inhibition. Indeed, given that the transposition between adjacent items in the above example is a very common type of error in serial recall (the locality constraint; Chapter 4), any repetition facilitation found in such situations could be no more than a scoring bias against control elements.

One way to try to overcome the bias is to use an item-scoring criterion, where an element is scored correct irrespective of its recall position. However, this does not address people's ability to recall a list in the correct order, for which a position-scoring criterion is required. An alternative approach is not to score a transposition between two control elements as incorrect (Wickelgren, 1965c). This is a more suitable approach. However, the most general way to control for the bias is to count either control element appearing in a critical position as correct. For example, all four reports of the repetition list *1*$R_1$*34*$R_2$*6* in Table 7-1 will look identical to the experimenter, even though they may represent different outputs from short-term memory. To control for this, all four reports of the control list *123456*, in Table 7-1 would be judged as correct under the *modified control scoring.* There are several points worth noting about the reports in Table 7-1:

1. Under conventional position-scoring, such as serial position curves (e.g., Crowder, 1968a), the second, third and fourth reports of the control list will lead to errors on Position 2 or Position 5, underestimating performance on control elements relative to repeated elements.

2. Even under conventional item-scoring, the third and fourth reports of the control list will reduce Lee's (1976b) probability of recalling both control elements somewhere, relative to recalling both repeated elements somewhere.

3. Repetitions of a response in the third and fourth reports of the control list might appear rare from the meta-analysis of Chapter 4, but that was when subjects knew that lists never contained repeated items. When subjects are aware that some lists contain repeated items, as in the present experiments, a greater incidence is likely to result. Furthermore, an even higher incidence is likely when lists are grouped (Chapter 3).

| Reports of Repetition List | | | | | |
|---|---|---|---|---|---|
| 1 | $R_1$ | 3 | 4 | $R_2$ | 6 |
| 1 | $R_2$ | 3 | 4 | $R_1$ | 6 |
| 1 | $R_1$ | 3 | 4 | $R_1$ | 6 |
| 1 | $R_2$ | 3 | 4 | $R_2$ | 6 |
| **Reports of Control List** | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 5 | 3 | 4 | 2 | 6 |
| 1 | 2 | 3 | 4 | 2 | 6 |
| 1 | 5 | 3 | 4 | 5 | 6 |

Table 7-1: Example reports of a repetition list *1R34R6* and a control list *123456* that are correct under modified control scoring.

4. If Item 6 were recalled in place of Item 3, Wickelgren (1966) would have scored it as an associative intrusion in all four reports of the repetition list, but only in the first and third reports of the control list (because the previous control element would not be correct in the second and fourth report), potentially biasing measures of repetition contamination.

In addition to the modified scoring of control lists, the present study introduces a single index of repetition facilitation and repetition inhibition. These effects have been scored in several different ways in previous studies, often because they have been studied separately from one another. The new index, *delta*, provides a unified measure of both repetition facilitation and repetition inhibition. Specifically, delta is the difference between the probability of recalling two repeated elements and the probability of recalling two control elements under modified control scoring (where correct recall is further defined under either position- or item-scoring criteria). A positive value of delta implies facilitation; a negative value implies inhibition. This measure requires recall of more than one repeated element (Lee, 1976b), though it does not distinguish which repeated element benefits or suffers in recall. Again, this reflects the fact that the experimenter cannot be certain that the first repeated element recalled is the first repeated element presented (though previous studies suggest it is mainly the second repeated element to be recalled that is affected, e.g., Crowder, 1968a; Jahnke, 1969a; Wickelgren, 1965c).

If repetition facilitation and repetition inhibition arise above possible scoring biases, then previous research suggests that both should be demonstrable in the present experiment simply by varying the separation between critical elements. Plotting delta against *repetition separation* should give some form of facilitation-inhibition continuum, with positive delta at short separations and negative delta at longer ones. In the present experiments, repetition separation was defined over several different *repetition formats*, depending on the exact positions of repeated elements. This was to examine the role of serial position in addition to repetition separation.

The second aim of Experiment 6 was to examine how this continuum changes with the introduction of temporal grouping. Strong interactions might be expected between grouping and repetition effects. For example, will repetition facilitation remain for adjacent repeated elements that straddle a group boundary? Or will repetition inhibition remain for more widely separated repeated elements that occur at the same position within groups? These questions may shed further light on the nature of repetition effects. Another reason for investigating the effect of temporal grouping is that people will often spontaneously group lists subjectively (Chapter 3). It is possible that the choice of grouping strategy will be influenced by the presence of repeated elements in a list. This possibility was also noted in passing by Walsh and Schwartz (1977):

> "Both during presentation and during recall, subjects often imposed their own intonation groupings or rhythmic patterns upon the items within the sequence. This suggests a possible source of interference which would apply to the experimental but not to the control sequences, i.e. a conflict between the subjective grouping imposed by the subject and the objective grouping imposed by intrasequence repetition." (p. 68).

The presence of objective, temporal grouping of lists should override any subjective grouping strategies triggered by particular repetition formats, and hence eliminate the bias suggested by Walsh and Schwartz.

The final aim of Experiment 6 was to measure repetition contamination, particularly errors following correct recall of a repeated element (*contamination errors*). Wickelgren's associative intrusions are contamination errors following one repeated element that are transpositions of the item that immediately followed the other repeated element. For example, recall of a list *1R34R6* as *1R63R4* contains two contamination errors (Items 6 and 4), the first of which (Item 6) is an associative intrusion.

The evidence for associative intrusions in Wickelgren's (1966) experiment was rather weak. Indeed, Wolf and Jahnke (1968) failed to find a significant incidence of associative intrusions. Moreover, associative intrusions could be the result of subjective grouping, rather than item-item chaining. The presence of repeated elements in a list might affect (and even effect) spontaneous grouping of that list, as a number of subjects in subsequent experiments reported. For example, the list *12R4R6* might be grouped as three groups of two, the repeated elements triggering groups of this size during rehearsal. If repetition lists were grouped in this manner, with repeated elements starting groups, then associative intrusions could be no more than a special case of interpositions between groups (Chapter 3). If this explanation were true, then overriding subjective grouping by objective grouping should remove any difference in the incidence of associative intrusions in repetition lists compared with control lists.

In summary, the present experiment measured repetition facilitation, inhibition and contamination within the same design, by manipulating the position of repeated elements (within subjects) and the presence or absence of temporal grouping (between subjects).

## Method

### Subjects

Twenty-four subjects from the APU Subject Panel were assigned to two conditions. There were ten women and two men in the ungrouped condition (mean age of twenty-nine), and eight women and four men in the grouped condition (mean age of twenty-eight).

### Materials

The letters *J, H, R, Q, V, M* were used to generate 144 lists of 6 items. Two-thirds of the lists (the repetition lists) contained one repeated item; the rest (the control lists) contained all six items. The repetition lists were divided into eight repetition formats, according to the positions of the repeated elements (Table 7-2). The repeated elements were between one and four positions apart. Over all lists, each item was repeated equally often and occurred approximately equally often at each position.

| List Type | Repetition Format | Repetition Separation | No. of Lists |
|---|---|---|---|
| Control | 1 2 3 4 5 6 | - | 48 |
| Repetition | 1 R R 4 5 6 | 1 | 12 |
| Repetition | 1 2 R R 5 6 | 1 | 12 |
| Repetition | 1 R 3 R 5 6 | 2 | 12 |
| Repetition | 1 2 R 4 R 6 | 2 | 12 |
| Repetition | 1 R 3 4 R 6 | 3 | 12 |
| Repetition | 1 2 R 4 5 R | 3 | 12 |
| Repetition | R 2 3 4 R 6 | 4 | 12 |
| Repetition | 1 R 3 4 5 R | 4 | 12 |

Table 7-2: Composition of lists in Experiment 6.

Control and repetition lists were distributed equally over 6 blocks of 24 trials. The order of trials was pseudorandomised, with the constraint that two consecutive trials never contained the same repetition format nor repetition of the same item. Subjects did not know in advance whether the next trial would contain a repeat, though the ratio of repetition to control lists would lead them to expect repetition more often than not.

**Procedure**

Each letter was presented in the centre of a VDU, replacing the previous one, at a rate of two every second (400 ms on, 100 ms off). The grouped condition included an further 500 ms pause after every third letter. The sixth letter was followed by a sequence of three distractor digits (drawn randomly without replacement from the set *1-9*), presented at the same rate as the letters. Subjects were instructed to vocalise each letter and digit as it appeared, but to recall only the letters. Vocalisation of the letters was monitored by the experimenter to ensure that subjects perceived the items correctly, particularly the repeated items. (In fact, errors in vocalisation were extremely rare.) Though vocalisation of the items introduced potentially complicating effects of echoic, auditory information, vocalisation of the distractor digits should minimise such effects, each digit acting much like an auditory suffix.

Immediately after the last digit, a visual cue appeared to prompt spoken, forward recall of the letters, which were written down by the experimenter. Subjects were encouraged to guess if they were unsure, or to say *blank* if no letter came to mind. They were alerted to the fact that some lists contained repeated letters. After ten practice trials, each subject attempted all six blocks, with the order of blocks counterbalanced across subjects. The whole experiment took approximately 50 minutes.

**Results**

In brief, significant repetition facilitation and repetition inhibition were found under modified control scoring. In ungrouped lists, repetition facilitation was found for adjacent repeated elements and repetition inhibition was found as soon as one or more context elements intervened. In grouped lists, no repetition facilitation was found for adjacent repeated elements that straddled a group boundary, and no repetition inhibition was found for three-apart repeated elements that occurred at the end of groups. Furthermore, repetition facilitation and repetition inhibition were dissociable by different scoring criteria: repetition facilitation reflected superior positioning of repeated elements, whereas repetition inhibition reflected inferior recall of repeated elements anywhere.

There was a nonsignificant trend for a greater proportion of errors following repeated elements to be associative intrusions than following correct elements, irrespective of grouping. However, this trend may have reflected a guessing bias between the two list-types, and therefore does not constitute evidence for associative chaining models of serial recall.

## Position-scoring of Serial Recall

The probability of recalling critical elements on both critical positions was calculated for repetition and control lists (Table 7-3). A three-way ANOVA on log-odds showed a significant effect of list type, $F(1,330)=45.21, p<.001$, and repetition format, $F(7,330)=31.18, p<.001$. There was no significant effect of grouping, $F(1,22)<1$, but grouping did interact with repetition format, $F(7,330)=5.35, p<.001$. The interaction between list type and repetition format was also significant, *F(7,330)=15.91, p<.001*, as was the three-way interaction between list type, repetition format and grouping, *F(7,330)=3.01, p<.005*..

| Condition | 1RR456 | 12RR56 | 1R3R56 | 12R4R6 | 1R34R6 | 12R45R | R234R6 | 1R345R |
|---|---|---|---|---|---|---|---|---|
| | Repetition Format | | | | | | | |
| Ungrouped | | | | | | | | |
| Repeated | .74 (.16) | .75 (.14) | .55 (.16) | .49 (.15) | .43 (.14) | .64 (.17) | .50 (.16) | .44 (.15) |
| Control | .62 (.15) | .62 (.15) | .62 (.14) | .62 (.15) | .63 (.15) | .69 (.15) | .61 (.17) | .67 (.16) |
| Grouped | | | | | | | | |
| Repeated | .88 (.09) | .66 (.15) | .41 (.17) | .40 (.16) | .52 (.15) | .82 (.13) | .39 (.16) | .48 (.15) |
| Control | .70 (.15) | .75 (.14) | .64 (.14) | .62 (.15) | .66 (.14) | .75 (.15) | .67 (.14) | .63 (.16) |

Table 7-3: Correct recall of critical elements under position-scoring in Experiment 6. (Calculated from weighted log-odds.)

The general pattern of facilitation and inhibition can be seen by collapsing repetition formats with the same repetition separation. To calculate delta, a repeated measure, the differences between recall of repeated elements and recall of control elements were averaged over subjects. The resulting delta values are shown in Figure 7-1, expressed as weighted, empirical log-odds. (A log-odds value of +1.0, for example, represents an increase of .20 in recall of repeated elements over recall of control elements, given recall of control elements of .60.) The upper panel shows delta for ungrouped lists; the lower panel shows delta for grouped lists. Even making a Bonferroni correction for eight pairwise comparisons, delta was significantly different from zero in all cases, *Z(12)>3.21, p<.006*, except for one-apart and three-apart repeated elements in the grouped condition, *Z(12)<1.17, p>.24*.

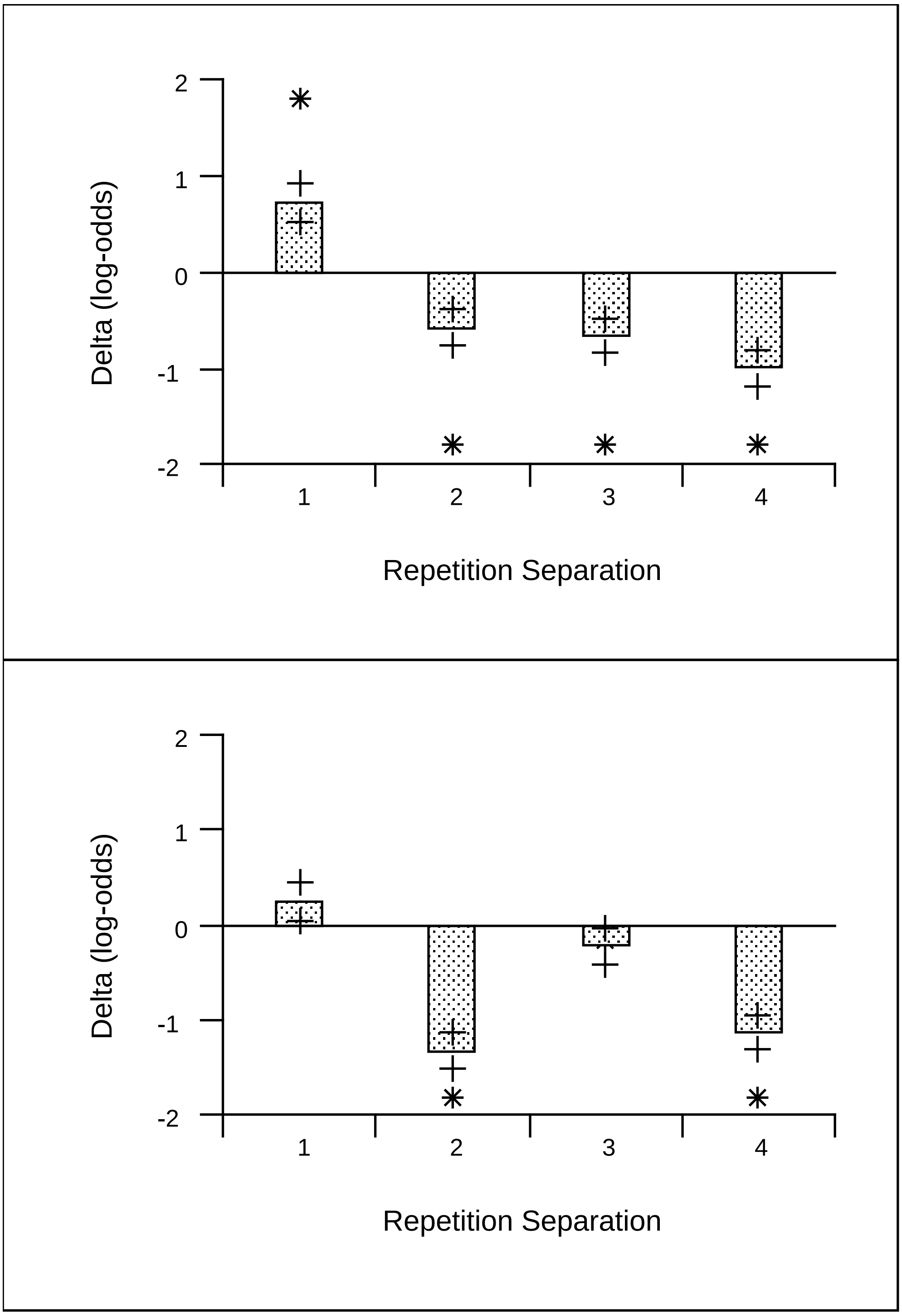



Figure 7-1: Delta under position-scoring as a function of repetition separation in ungrouped (upper panel) and grouped (lower panel) conditions of Experiment 6.
(Cross-hairs show standard error of delta scores above and below mean; asterisks indicate delta values significantly different from zero under Holm's method of multiple comparisons.)

To examine the three-way interaction between list type, repetition format and grouping in more detail, delta values were tested via weighted log-odds for each repetition format in grouped and ungrouped conditions. Delta was significantly below zero for ungrouped repetition formats *1R34R6*, *R234R6* and *1R345R*, and grouped repetition formats *1R3R56*, *12R4R6*, *1R345R* and *R234R6*, $Z(12)>3.12$, $p<.005$. Delta was significantly above zero for grouped repetition format *1RR456*, $Z(12)=3.93$, $p<.0001$. No other delta values reached significance under Holm's correction for sixteen comparisons, $Z(12)<2.70$, *family-wise* $p>.05$.

Differences between weighted delta values were also tested across grouped and ungrouped conditions. Grouping had no significant effect on delta for repetition format *1RR456*, $Z(24)=1.21$, $p=.23$, but significantly decreased delta for repetition format *12RR56*, $Z(24)=3.03$, $p<.005$. Grouping also decreased delta for repetition formats *1R3R56*, *12R4R6* and *R234R6*, but not necessarily significantly, $Z(24)<2.44$, *family-wise* $p>.05$. In contrast, grouping increased delta for repetition format *12R45R*, such that delta changed sign, though again the change did not quite reach significance, $Z(24)=1.88$, $p=.06$. Repetition formats *1R34R6* and *1R345R* showed little change, $Z(24)<0.94$, $p>.35$.

One reason why some contrasts across grouped and ungrouped conditions did not quite reach significance may be because subjects in the ungrouped condition were spontaneously grouping lists in threes. Several subjects reported such grouping in debriefing. Spontaneous grouping may also explain why there was no significant repetition inhibition for repetition format *12R45R* in ungrouped lists, $Z(12)=1.14$, $p=.25$. This lack of repetition inhibition was not simply because the second repeated element in repetition format *12R45R* was the last item in the list (Crowder, 1968a), because highly significant repetition inhibition was found for repetition format *1R345R*, $Z(12)=4.28$, $p<.0001$.

In summary, the failure to find significant repetition facilitation for one-apart repeated elements in the grouped condition came from a reduction in delta when repeated elements straddled a group boundary (i.e., repetition format *12RR56*): Elements repeated immediately within a group (i.e., repetition format *1RR456*) continued to show repetition facilitation. The failure to find significant repetition inhibition for three-apart repeated elements in the grouped condition came from an increase in delta for repeated elements at the end of groups (i.e., repetition format *12R45R*): Elements repeated in the middle of groups (i.e., repetition format *1R34R6*) showed little change.

**Other scoring of Serial Recall**

Repetition facilitation and repetition inhibition were investigated further by using an item-scoring criterion, measuring how often critical elements were recalled in any two positions in a report. Delta under item-scoring is shown for each repetition separation in Figure 7-2. The most striking observation is that only repetition inhibition was evident under item-scoring; there was no significant repetition facilitation. In other words, recall of two repeated elements somewhere was always less likely than recall of two control elements somewhere. Further tests of weighted, log-odds showed delta was significantly less than zero for all repetition formats, $Z(12)>2.58$, $p<.01$, except *1RR456*, both when grouped and when ungrouped, *12RR56* when ungrouped, and *12R45R* when grouped, $Z(12)<2.16$, *family-wise* $p>.05$. Thus repetition inhibition was clear in all cases except for adjacent repeated elements that did not straddle a group boundary, and three-apart repeated elements at the end of groups.

If repetition facilitation did not arise through better item recall of repeated elements, it must have arisen through better positioning of those elements. This was confirmed by analysing the conditional probability of recalling critical elements in the two critical positions, given that critical elements were recalled at two positions somewhere. The results under this scoring criterion are shown in Figure 7-3. There was clear repetition facilitation for adjacent repetition, but no repetition inhibition for any other repetition separation, except for two-apart repeated elements in the grouped condition. Indeed, separate analysis of each repetition format showed delta was significantly above zero for repetition format *1RR456*, both grouped and ungrouped, and repetition format *12RR56* when ungrouped, $Z(12)>3.42$, $p<.001$ in each case, but not significantly different from zero for any other format, $Z(12)<2.70$, *family-wise* $p>.05$.

**Repetition Contamination**

In general, context elements in repetition lists were recalled better when repeated elements were recalled better (i.e., under conditions of repetition facilitation). This was reflected in better recall of the lists as a whole. For example, approximately *61%* of ungrouped lists with repetition format *1RR456* were recalled correctly, compared with *52%* of control

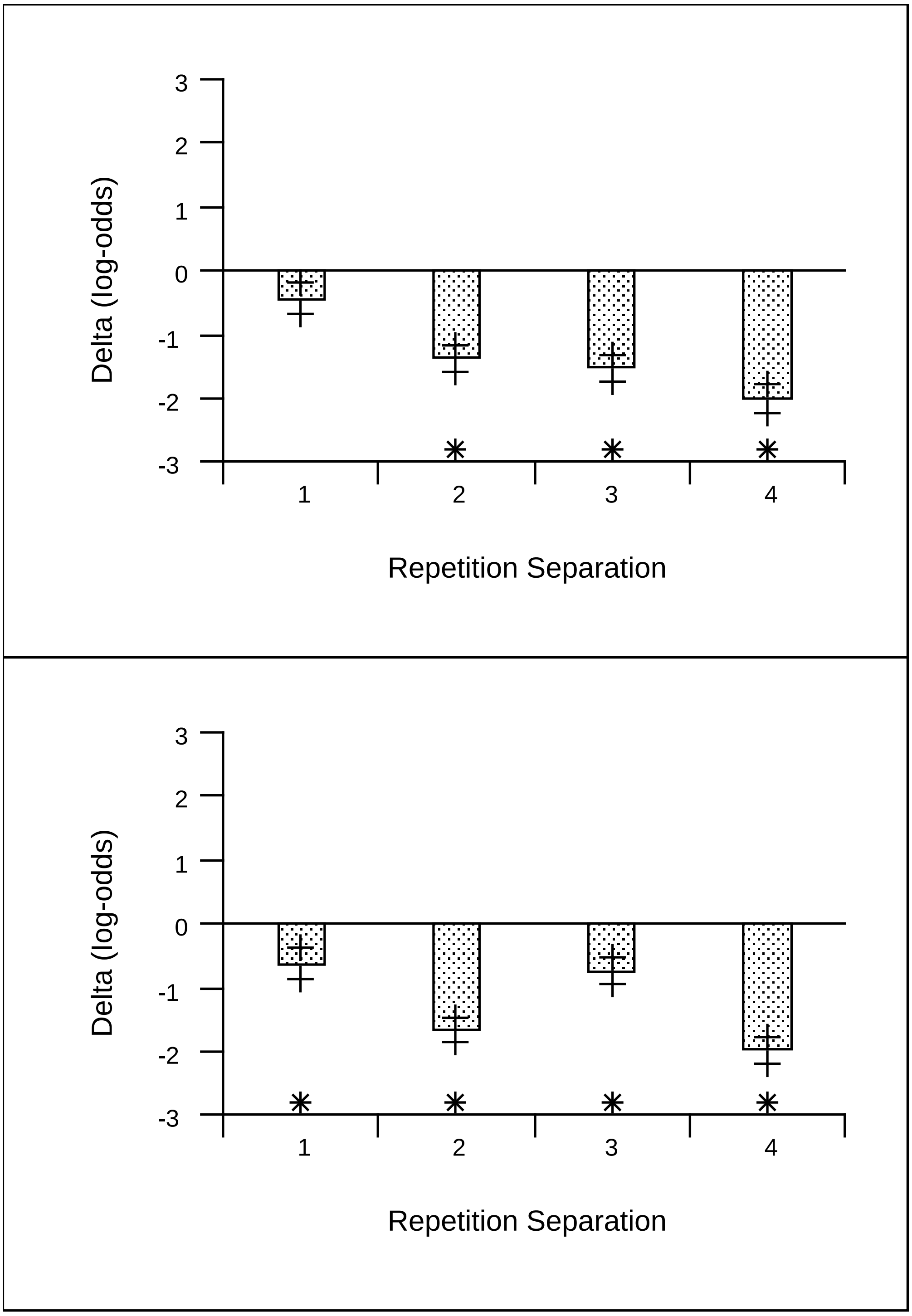



Figure 7-2: Delta under item-scoring as a function of repetition separation in ungrouped (upper panel) and grouped (lower panel) conditions of Experiment 6.
(Cross-hairs show standard error of delta scores above and below mean; asterisks indicate delta values significantly different from zero under Holm's method of multiple comparisons.)

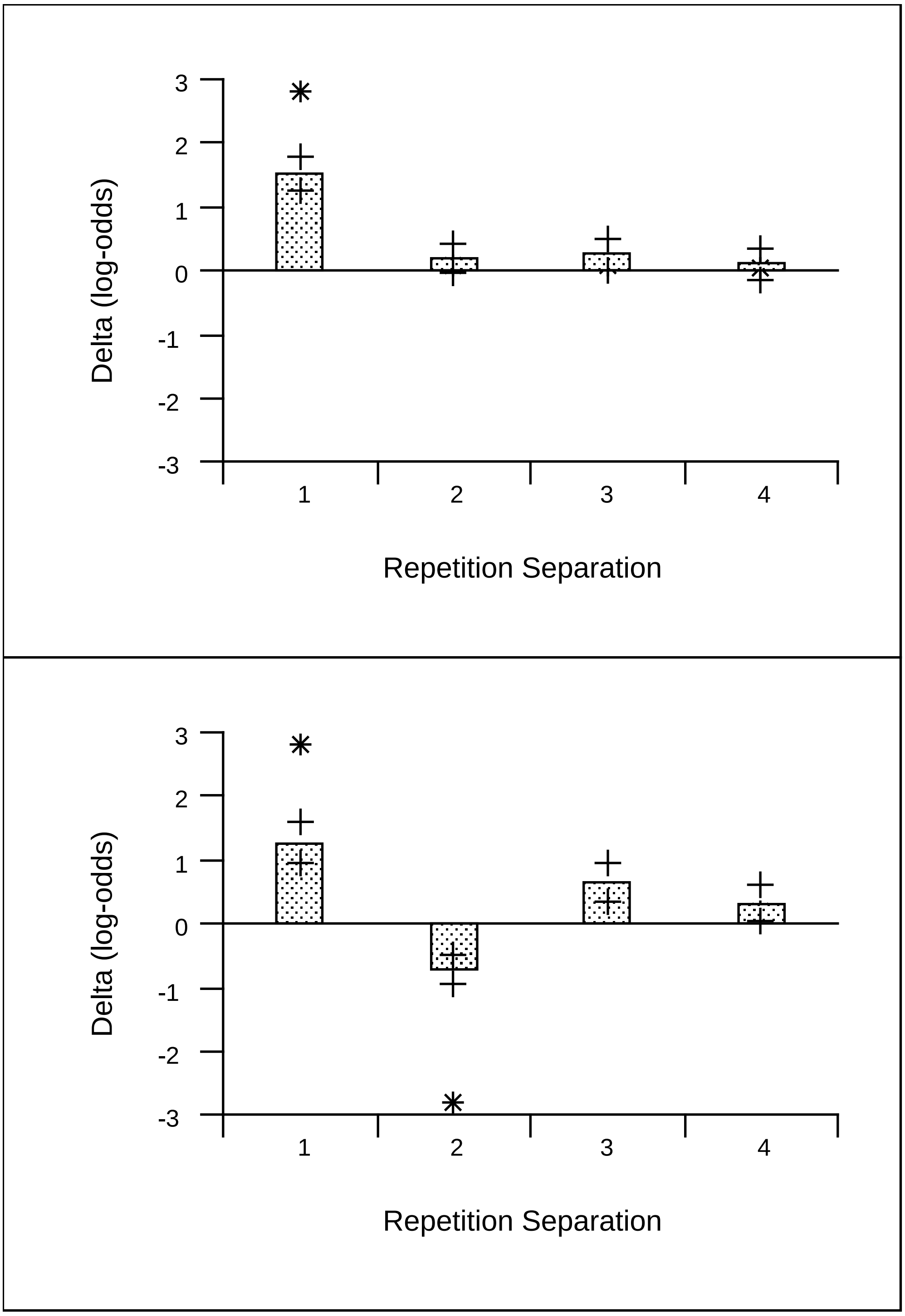



Figure 7-3: Delta under position-scoring, given correct item recall, as a function of repetition separation in ungrouped (upper panel) and grouped (lower panel) conditions of Experiment 6. (Cross-hairs show standard error of delta scores above and below mean; asterisks indicate delta values significantly different from zero under Holm's method of multiple comparisons.)

lists. The converse was true for repetition formats that produced repetition inhibition. For example, only *35%* of ungrouped lists with format *1R345R* were recalled correctly. This mirrors previous findings (e.g., Crowder, 1968a).

A more detailed analysis of repetition contamination examined the errors immediately following correct critical elements, using modified control scoring and collapsing over repetition formats with at least one context element following each critical element (i.e., repetition formats *1R3R56*, *12R4R6*, *1R34R6*, and *R234R6*). The proportion of responses following correct critical elements that were errors and the proportion of such contamination errors that were associative intrusions (following Wickelgren, 1966) are shown in Table 7-4.

| | Contamination Error | Associative Intrusion |
|---|---|---|
| Ungrouped | | |
| Repeated | .18<br>(.05) | .36<br>(.18) |
| Control | .15<br>(.05) | .25<br>(.18) |
| Grouped | | |
| Repeated | .17<br>(.05) | .24<br>(.16) |
| Control | .15<br>(.05) | .19<br>(.14) |

Table 7-4: Proportion of responses following correct critical elements that were errors and the proportion of such contamination errors that were associative intrusions in Experiment 6. (Calculated from weighted log-odds.)

A two-way ANOVA on the log-odds of a contamination error, given correct recall of the preceding critical element, showed a significant effect of list-type $F(1,22)=4.58$, $p<.05$, reflecting a greater incidence of contamination errors following repeated elements, but no significant effect of grouping, or interaction, $F(1,22)<1.11$, $p>.30$. A two-way ANOVA on the log-odds that a contamination error was an associative intrusion showed a significant effect of grouping, $F(1,22)=6.06$, $p<.05$, reflecting a greater incidence of associative intrusions in ungrouped lists, but no significant effect of list-type, or interaction, $F(1,22)<1$.

Thus, though both ungrouped and grouped conditions showed a trend for a greater proportion of contamination errors to be associative intrusions, this trend did not quite reach significance, even under direct tests of weighted log-odds, $Z(12)<1.59$, $p>.11$. More importantly, there is an alternative reason for this trend. If subjects guessed a list item at random when unsure, the proportion of contamination errors that were associative intrusions expected by chance would be .25 in repetition lists (given the four possible erroneous items), and .20 in control lists (given the five possible erroneous items). In other words, there is a guessing bias for more associative intrusions following repeated elements than control elements. Consistent with this account, tests of weighted log-odds showed that the incidence of associative intrusions in repetition and control lists was not significantly greater than these chance levels in either condition, $Z(12)<1.02$, $p>.31$.

Nevertheless, a guessing account does not explain why the proportion of responses following correct repeated elements that were errors was greater than the proportion following correct control elements. However, this difference was accompanied by a lower frequency of a correct repeated elements ($M=.66$, $SD=.09$) than correct control elements ($M=.80$, $SD=.07$), a difference that was significant, $Z(12)=4.63$, $p<.001$. This reflected the fact that recall of repeated elements was worse than control elements for the subset of repetition formats used to examine repetition contamination (i.e. those with nonadjacent repeated elements, which showed repetition inhibition). The greater proportion of contamination errors in repetition lists may therefore reflect greater difficulty, on average, in recalling those lists. This caveat emphasises the care required of conditional analyses of errors (e.g., Bower & Springston, 1970; Healy, 1982).

**Discussion**

Using a single measure of performance and a position-scoring criterion, the present experiment showed both repetition facilitation and repetition inhibition simply by varying the positions of repeated elements. In the ungrouped condition, there was a transition from facilitation to inhibition as soon as one or more context elements intervened between the repeated elements. Furthermore, temporal grouping removed repetition facilitation for adjacent repeated elements that straddled a group boundary, and appeared to remove repetition inhibition for repeated elements at the ends of groups. This striking interaction between grouping and item repetition represents an important new finding.

The new method of scoring control lists confirmed that these effects arise over and above any potential bias in favour of repetition lists (owing to the inability of the experimenter to tell which response represents which repeated element). The results under this modified scoring were in broad agreement with previous studies, except that significant repetition inhibition was found for two-apart repeated elements, where none has been found before (e.g., Crowder, 1968a; Lee, 1976b). In fact, when contrasted with conventional scoring, the modified scoring of control lists produced increases of up to 7% in the probability of recalling control elements. Though such differences are small in absolute terms, they resulted in a 50% increase in delta in some cases. Modified control scoring was therefore maintained in Experiments 7 and 8.

One might seek to explain the variability in delta as a function of repetition format by examining differences in overall performance levels. Ceiling effects, for example, may account for the failure to find significant repetition inhibition when repeated elements occur at the end of groups. However, the probabilities in Table 7-3 are rarely greater than .80, and performance on control elements is quite constant, around .63 when ungrouped and .67 when grouped. Indeed, the largest difference in performance on control elements over all sixteen conditions is only .14. Even when the data are reanalysed using a relative measure of performance (Jahnke, 1969b), where delta is divided by the probability of recalling control elements, the pattern of results remains essentially unchanged. Thus, though the magnitude of delta is undoubtedly sensitive to overall performance levels, this is not sufficient to explain the effects of repetition separation and grouping.

Further investigation of repetition facilitation and repetition inhibition was made possible by considering different scoring criteria. Under item-scoring, recall of repeated elements nearly always suffered compared to control elements, no matter how far apart the repeated elements. This most probably reflected a failure to recall a repeated item more than once (Jahnke, 1969b; Lee, 1976b). By conditionalising the positioning of critical elements on their recall somewhere, this bias was removed and repetition facilitation was shown to arise from better positioning of adjacent repeated elements (Drewnowski, 1980a). The dissociation of repetition facilitation and repetition inhibition by different scoring criteria suggests that (at least) two different factors contribute to the effects: one that reduces the probability of recalling a repeated item more than once, and another that increases the probability of positioning the repeated item, provided it is recalled more than once.

Regarding repetition contamination, the present results showed that recall of context elements generally correlated with recall of repeated elements. This is not surprising, given the interdependencies between responses in a report (Chapter 4): Better positioning of one item will necessarily improve positioning of other items, explaining why positioning of context elements benefits only under conditions of repetition facilitation (Crowder, 1968a). Conversely, context elements will be less well positioned under conditions of repetition inhibition, even though they may be more often recalled somewhere (Wickelgren, 1965c), given the smaller set of items to chose from in repetition lists than control lists.

A more specific measurement of repetition contamination showed a trend for a greater proportion of contamination errors to be associative intrusions in repetition lists than control lists, as predicted by the associative chaining theory of Wickelgren (1965c). This trend did not appear due to subjective grouping elicited by repeated elements, as suggested in Henson et al. (1996), because no significant interaction was found between this trend and the presence or absence of objective grouping. However, the trend did not reach significance in the present experiment. One reason for this may be a lack of statistical power, given the small numbers of associative intrusions and considerable variability across subjects. However, several points are worth noting in this respect.

Firstly, Wickelgren had even fewer data points in his 1966 experiment, using a sign test over subjects. Even then, there was only a significant incidence of associative intrusions in three of his eight conditions (only one of which would be significant under correction for multiple comparisons). Secondly, Wickelgren did not use the modified control scoring used in the present experiment. This marking scheme allows a greater number of control elements to be judged correct, which may increase the proportion of errors following correct control elements that are scored as associative intrusions. Though such a marking scheme may not be fair to theories assuming type representations of items (such as the chaining account of associative intrusions), it does suggest an alternative explanation of Wickelgren's results, by virtue of an underestimation of associative intrusions in control lists.

Thirdly, there is another possible explanation of Wickelgren's results, in terms of a guessing bias. Because there are a smaller number of items from which to chose a response in repetition lists than control lists, the baseline chance of an associative intrusion in repetition lists is higher than in control lists. Thus, if people occasionally recalled a critical element correctly, but forgot the item that followed it, they have a greater chance of guessing the item that followed the other critical element in repetition lists than control lists. Though such situations are probably rare, they may be sufficient to account for the small trend observed in the present experiment, and the significant trend found by Wickelgren (1966).

Finally, even if associative intrusions are more than a guessing bias, they still represent an extremely small proportion of errors (less than 2% of errors in the present experiment). Indeed, attempts to measure associative intrusions in Experiments 7 and 8 were thwarted by the scarcity of such errors. Thus, even if repeated items do represent ambiguous cues in a process of item-item chaining, this fact has an almost negligible effect on the recall of repetition lists, especially in relation to the effects of repetition facilitation and repetition inhibition on recall of context elements. Even if the alternative explanations of Wickelgren's finding prove incorrect, associative chaining models of serial recall would surely predict a much stronger effect of repetition on cuing (cf. the effect of similarity on cuing; Chapter 2).

In summary, the present experiment replicated both the repetition facilitation and the repetition inhibition reported in previous studies, within a single design and under a new, unbiased scoring of control elements. Moreover, these effects were shown to be sensitive to the grouping of lists; a factor overlooked in previous studies (Walsh & Schwartz, 1977). Given that grouping is such a prevalent and powerful effect in serial recall (Chapter 3), this sensitivity has important implications for both measurement and interpretation of repetition effects. In particular, the role of grouping in the detection of repeated items was examined in Experiment 7.

## Repetition Memory

The previous experiment demonstrated a transition from repetition facilitation to repetition inhibition in ungrouped lists as the separation between repeated elements increased. How does varying repetition separation have such dramatic effects on the probability of recalling repeated elements? One factor that covaries with repetition separation is the probability of detecting the repetition event (i.e., the fact that an item was repeated). Lee (1976b), for example, showed that the probability that subjects detected repetition in a list decreased as the separation between repeated elements increased. Jahnke (1972a) showed a similar effect in a recognition task: Recognition for a pair of repeated elements was worse when they were far apart in a list than when they were close together. In a striking demonstration of people's general failure to detect repetition, Malmi and Jahnke (1972) reported that even when 100% of lists contained repeated elements, subjects only guessed around 40% had repeated items on debriefing. Thus one important contribution to the repetition effects in Experiment 6 may be the probability of detecting repetition: Detection may be necessary for repetition facilitation, while failure to detect a repetition may result in repetition inhibition. A reduced probability of detecting repetition may explain why repetition facilitation was not found across a group boundary, while an increased probability of detecting repetition may explain why repetition inhibition was not found at the end of groups.

Why should people fail to detect repetition of an item? It is not simply because subjects do not expect repetitions: Repetition inhibition arises even when they are told in advance to expect repetition (Experiment 6), are reminded that repetition in lists is possible (Jahnke, 1969b), experience a high frequency of repetition (Hinrichs et al., 1973; Wickelgren, 1965c) or have considerable practice in recalling lists with repeated items (Crowder, 1968b). Repetition inhibition is not the result of repetition blindness either (e.g., Kanwisher, 1987), which is unlikely with the slow presentation rates of the above studies. Indeed, repetition blindness would predict greatest repetition inhibition for repeated elements closest together in time, in striking contrast to the repetition facilitation found for adjacent repeated elements in Experiment 6. Moreover, repetition blindness is often viewed as a perceptual problem, and yet repetition inhibition arises even when subjects vocalise repeated items as they are presented

(Experiment 6; note that one can correctly read aloud an item twice, without explicitly registering that there was repetition of that item). Thus repetition blindness is not sufficient to explain repetition inhibition.

The problem may not rest with detection either, because one still has to remember the repetition event at recall. In other words, it is possible to detect a repetition during presentation of a list, but forget that repetition during recall. Jahnke (1969b) tested this possibility by reminding subjects whether or not a repetition had occurred immediately before recall of each list. Though it reduced the amount of repetition inhibition, this reminding did not eliminate it. Thus failure to remember a repetition event during recall may contribute to repetition inhibition, but it is not a sufficient reason. A further requirement may be to remember not only that a repetition occurred, but which particular item was repeated. If correct positioning of the repeated elements is necessary, the requirement may be even more stringent: One may have to remember where the repetition occurred.

Even if memory for the repetition is an important factor in serial recall, questions remain as to how such memory might improve recall of repeated elements (repetition facilitation), or why a lack of such memory might impair recall of repeated elements (repetition inhibition). The latter question is dealt with in Experiment 8. As for the former question, several roles for detection and memory of a repetition have been suggested. One possibility is that the detection of a repetition leads to increased attention to, or rehearsal of, the repeated elements. A similar reason is often given for Restorff isolation effects (e.g., Potts & Shiffrin, 1970). However, this type of explanation has problems explaining why repetition effects are stronger when presentation rates are faster (Wickelgren, 1965c). It is also inconsistent with Lee's (1976b) finding that the probability of recalling at least one repeated element is no greater than recalling at least one control element: If increased attention were given to repeated elements, then the probability of recalling at least one repeated element should exceed that of recalling at least one control element.

Alternatively, repeated elements may be recoded into a single unit or chunk (e.g., *double-five*), reducing overall memory load (Wickelgren, 1965c). However, such chunking, at least as defined by Johnson (1972), implies all-or-none recall of repeated elements. This is again inconsistent with Lee's (1976b) findings. Instead, Lee proposed that detection of a repetition leads to a *repetition tag* associated with the repeated item. Memory for this tag is independent of memory for the item that was repeated (e.g., separate but associated representations of *five* and *doubled*). Tagging produces repetition facilitation through increasing the probability that both occurrences of a repeated item are recalled. Lee used this fact to explain why the probability of recalling both occurrences was greater than that predicted if recall of each occurrence were statistically independent.

Lee (1976b) assumed further that repetition tags can become separated from a repeated item over time and become associated with a different item. This can produce repetition of the wrong item, consistent with Lee's finding of more such repetition errors in repetition lists than control lists. A proposal similar to Lee's repetition tags has been suggested for long-term memory as well. Following the model of Rumelhart and Norman (1982), Houghton et al. (1994) introduced a *geminate node* in their model of spelling, which is associated with a particular position in a word (rather than a particular letter, as with Lee's repetition tags). The job of the geminate node is to double the output of the letter at that position (and thus is used only for immediate repetition). The fact that this node can sometimes become triggered at an earlier or later position during output accounts for the common typographical errors where the wrong letter is repeated in adjacent positions (e.g, *school* typed as *schhol*, or *scholl*).

Repetition facilitation may also be related to the detection of distinctive items, another reason given for isolation effects (Hunt, 1995). Experiments that vary the acoustic similarity of critical and context elements (e.g., Jahnke & Melton, 1968; Lee, 1976a) have found *contrast facilitation* for critical elements that, though not repeated, are in contrast with context elements (e.g., two adjacent, phonologically nonconfusable items are recalled better when surrounded by confusable items than when surrounded by other nonconfusable items). Importantly, similarity between critical items does not, on its own, lead to facilitation (so repetition is not just an extreme case of phonological similarity). Nevertheless, repetition facilitation and contrast facilitation may have different underlying causes, because repetition facilitation remains over retention intervals where contrast facilitation has disappeared (Lee, 1976a) and does not appear to interact with phonological similarity of items (Drewnowski, 1980a). Contrast facilitation could certainly not arise through the mechanisms Houghton et al. (1994) proposed for immediate repetition for example.

In summary, there may be a role for a separate *repetition memory* in repetition facilitation and repetition inhibition. Detection and memory of a repeated item may improve its recall, through some form of repetition tagging for example. A failure to detect or remember the repetition may result in a failure to recall both repeated elements.

Experiment 2 in Henson (1996b) tested people's repetition memory as a function of repetition format and the presence or absence of grouping. By using a subset of the repetition formats in Experiment 6, it was possible to compare the results of this repetition memory task with the results of the serial recall task. In fact, exactly the same presentation conditions were employed in both experiments; the only difference was whether the task was to remember all elements (Experiment 6), or just the repeated elements (Experiment 2, Henson, 1996b). Two specific hypotheses of interest were (1) whether grouping improved repetition memory for repetition format *12R45R*, where repeated elements occurred at the end of groups, and (2) whether grouping impaired repetition memory for repetition format *12RR56*, where repeated elements straddled a group boundary. The former was confirmed, but the latter was not confirmed (in fact, if anything, grouping improved repetition memory for repeated elements straddling a group boundary). Thus, the fact that repeated elements were detected and remembered better when at the end of groups can explain the lack repetition inhibition for this condition in Experiment 6. However, the lack of repetition facilitation for repetition format *12RR56* when grouped in Experiment 6 seems inexplicable in terms of poorer detection or memory for the repetition. An alternative explanation was offered in terms of repetition tagging (Henson, 1996b); an account elaborated here in the General Discussion.

The other main finding of Experiment 2 in Henson (1996b) was that people were extremely good at remembering a repetition event (on over 95% of occasions for repetition formats *12RR56* and *12R4R6*). They were less accurate at remembering which item was repeated, but even then, they were correct over 75% of the time. This high level of repetition memory could be taken to question the role of repetition memory in serial recall. However, there are several reasons why dismissing such a role would be premature.

Firstly, repetition inhibition may still arise even with high levels of repetition memory. For example, if repetition inhibition arose by default whenever repetition memory failed, it would only take a 20% failure rate of repetition memory to cause a decrement of 12% in the probability of recalling repeated elements relative to control elements (if control elements were correct 60% of the time). This is comparable to the magnitude of repetition inhibition in Experiment 6. Secondly, the concurrent memory demand in Experiment 6, to remember all elements and their order, may have produced much lower levels of repetition memory than measured in Experiment 2 of Henson (1996b). These issues were addressed in Experiment 7.

## Experiment 7

Experiment 7 differed from Experiment 2 in Henson (1996b) by measuring repetition memory on-line with serial recall. This tested repetition memory under a larger memory load than in the latter experiment; the additional memory load required for serial recall of all the elements in the list. Furthermore, by measuring repetition memory and serial recall of critical elements on a trial by trial basis, the two could be directly correlated. A strong correlation would support the hypothesis that repetition memory is an important determinant of repetition facilitation and inhibition. This would complement the indirect evidence for this hypothesis in Henson (1996b) and other studies (e.g., Lee, 1976b).

However, by attempting an on-line measurement of repetition memory, there are risks of contamination of one memory task (reporting the repeated item) by the other (recalling all items in order). Requiring subjects to indicate repeated items first may affect their subsequent serial recall. For example, Crowder (1968b) showed that a redundant response prefix produced “repetition inhibition” if the prefix item also occurred in the list to be recalled. Conversely, requiring serial recall before subjects indicate repeated items may allow them to base their decision about which item was repeated by simply inspecting or remembering what they had previously recalled. There seems to be no perfect solution to this dilemma.

Nevertheless, the approach taken in Experiment 7 was to ask subjects for written, serial recall, followed by a requirement to say aloud any item they thought was repeated. To minimise the risk of subjects performing the second task by inspecting their written reports, they were required to cover their responses before reporting any repeated items. The problem remained that what subjects report in the repetition memory task may still depend on what they recalled in the serial recall task. For example, a subject may repeat the wrong item in recall, and then “remember” the repetition of this item rather than the correct item. To minimise this risk, instructions for the serial recall task emphasised that subjects were only to write items they were certain about. Furthermore, a smaller subset of repetition formats was used than in Experiment 6, to increase overall performance under the dual task conditions.

In addition to investigating repetition memory, the present experiment was an important replication of Experiment 6, with a more powerful, within-subjects design. Three repetition formats were tested both ungrouped and grouped for each subject. Of particular interest was whether repetition inhibition would remain when subjects have specific reason to keep in mind a repeated item. Finally, the number of distractor digits was varied (either one or three), in order to test repetition memory and serial recall over different retention intervals.

### Method

#### Subjects

Twelve subjects from the APU Subject Panel were tested; three were men, nine were women and their mean age was thirty-four years.

#### Materials

One hundred and twenty lists of six items were generated from the same set of letters as Experiment 6. This time however, there was an equal number of repetition lists and control lists, and only three repetition formats from Experiment 6 were employed (Table 7-5). Control and repetition lists were distributed equally over 4 blocks of 30 trials. The order of trials was pseudorandomised in the same manner as previous experiments.

#### Procedure

Each subject attempted four conditions generated from factorial combination of two levels of retention interval (a short delay of one distractor digit and a long delay of three distractor digits) and two levels of grouping (ungrouped and grouped). The order of grouping conditions was constrained such that subjects always attempted the two ungrouped conditions

| List Type | Repetition Format | Repetition Separation | No. of Lists |
|---|---|---|---|
| Control | 1 2 3 4 5 6 | - | 60 |
| Repetition | 1 2 R R 5 6 | 1 | 20 |
| Repetition | 1 2 R 4 R 6 | 2 | 20 |
| Repetition | 1 2 R 4 5 R | 3 | 20 |

Table 7-5: Composition of lists in Experiment 7.

before the two grouped conditions, to reduce the incidence of spontaneous grouping (Experiment 2). The order of short or long retention intervals was counter-balanced within this constraint, as was the order blocks. The remaining procedure was similar to that of Experiment 6, except for three important differences. Firstly, subjects wrote rather than spoke their responses. Secondly, subjects were instructed not to guess, being told:

> "Most importantly, please do NOT guess at letters. In other words, only write a letter in a particular box if you are SURE that it occurred in that position. Otherwise, put a line through the box, before going on to try to recall the next one. It is better to indicate a blank than to respond with a letter which you are unsure about."

(Approximately 11% of responses were omissions, considerably greater than the figure of 4% in Experiment 6, suggesting that subjects did obey this instruction.) Thirdly, when subjects had finished the serial recall task, they were asked to cover their responses with a piece of card, before saying aloud any letter that they remember as being repeated in the list. They were told to report a repeated item even if they had not managed to recall that item correctly. If they did not remember any items as being repeated, they were told to say *none*. They were informed that half of the lists did contain a repeated item. The experiment took 50 minutes.

**Results**

In brief, results of the serial recall task were similar to those of Experiment 6, except that significant repetition facilitation was found for repeated elements at the end of groups. As expected, the concurrent memory demands of the serial recall task lowered performance on the repetition memory task relative to Experiment 2 in Henson (1996b). Most importantly, performance on the two tasks was highly correlated, consistent with the hypothesis that repetition memory is an important determinant of item repetition effects in serial recall.

**Position-scoring of Serial Recall**

As in Experiment 6, the probability of recalling either critical element on the two critical positions was calculated for both repetition and control lists (Table 7-6). Unlike Experiment 6 however, the equal number of control and repetition lists meant that each control list could be paired randomly with one of the three repetition formats (removing the risk of correlations between recall of more than two critical elements from the same control list). A four-way ANOVA on log-odds showed a significant effect of repetition format, $F(2,255)=18.43$, $p<.001$, grouping, $F(1,255)=72.57$, $p<.001$ and retention interval, $F(1,255)=23.20$, $p<.001$. Unlike Experiment 6, there was no significant effect of list type, $F(1,255)=2.46$, $p=.12$, but this factor did interact with repetition format, $F(2,255)=18.43$, $p<.001$, and retention interval, $F(1,255)=12.18$, $p<.001$. There were no further significant interactions. The lack of a main effect of list type suggested that repetition facilitation and repetition inhibition for the three different formats cancelled out overall. The interaction of list type with retention interval reflected a greater decrement in recall of control elements than repeated elements as retention interval was increased.

Collapsing across short and long retention intervals, Figure 7-4 shows delta against repetition format. Tests of weighted log-odds showed delta was only significantly different from zero for repetition format *12RR56* when ungrouped and repetition format *12R45R* when grouped, $Z(12)>2.63$, $p<.01$ in both cases; $Z(12)<1.88$, $p>.06$ for all other repetition formats. Thus, there was not only significant repetition facilitation for adjacent repeated elements when ungrouped, as in Experiment 6, but also for repeated elements at the end of groups, a nonsignificant trend in Experiment 6.

**Item-scoring of Serial Recall**

Though repetition inhibition was not reliable under position-scoring, it was clearly present under item-scoring. Figure 7-5 shows delta against repetition format, again collapsing across retention interval. Delta was significantly below zero for repetition format *12R4R6* in both ungrouped and grouped conditions, $Z(12)>2.65$, $p<.01$. Surprisingly, delta for grouped

| Condition | Short 12RR56 | Short 12R4R6 | Short 12R45R | Long 12RR56 | Long 12R4R6 | Long 12R45R |
|---|---|---|---|---|---|---|
| Ungrouped | | | | | | |
| Repeated | .68 (.21) | .39 (.22) | .53 (.25) | .59 (.23) | .35 (.22) | .66 (.24) |
| Control | .59 (.23) | .52 (.24) | .58 (.24) | .42 (.24) | .37 (.24) | .46 (.24) |
| Grouped | | | | | | |
| Repeated | .77 (.18) | .60 (.23) | .86 (.14) | .70 (.21) | .51 (.24) | .83 (.15) |
| Control | .76 (.19) | .78 (.19) | .80 (.18) | .59 (.23) | .51 (.24) | .67 (.22) |

Table 7-6: Correct recall of critical elements under position-scoring in Experiment 7. (Calculated from weighted log-odds.)

repetition format *12R45R* was still positive under item-scoring, thought this did not quite reach significance, *Z(12)=2.29, family-wise p>.05*. No other delta values were significant, *Z(12)<1.1, p>.27*, in all cases. Weighted tests of delta across ungrouped and grouped conditions showed that grouping significantly increased delta for repetition format *12R45R, Z(12)=2.30, p<.05*, but not for the other two repetition formats, *Z(12)<0.53, p>.60*.

**Item-scoring of Repetition Memory**

A three-way ANOVA on log-odds of memory for the repeated item showed significant effects of repetition format, *F(2,121)=12.85, p<.001*, and grouping, *F(1,121)=14.54, p<.001*, and a significant interaction between them, *F(2,121)=3.71, p<.05*. There was no significant effect of retention interval, nor any other significant interaction. Collapsing across retention interval, the accuracy of repetition memory is shown in Table 7-7. Performance was below the near-ceiling levels of Experiment 2 in Henson (1996b), probably explaining why the interaction between repetition format and grouping reached significance. This interaction arose from grouping producing a significant improvement for repetition format *12R45R*,

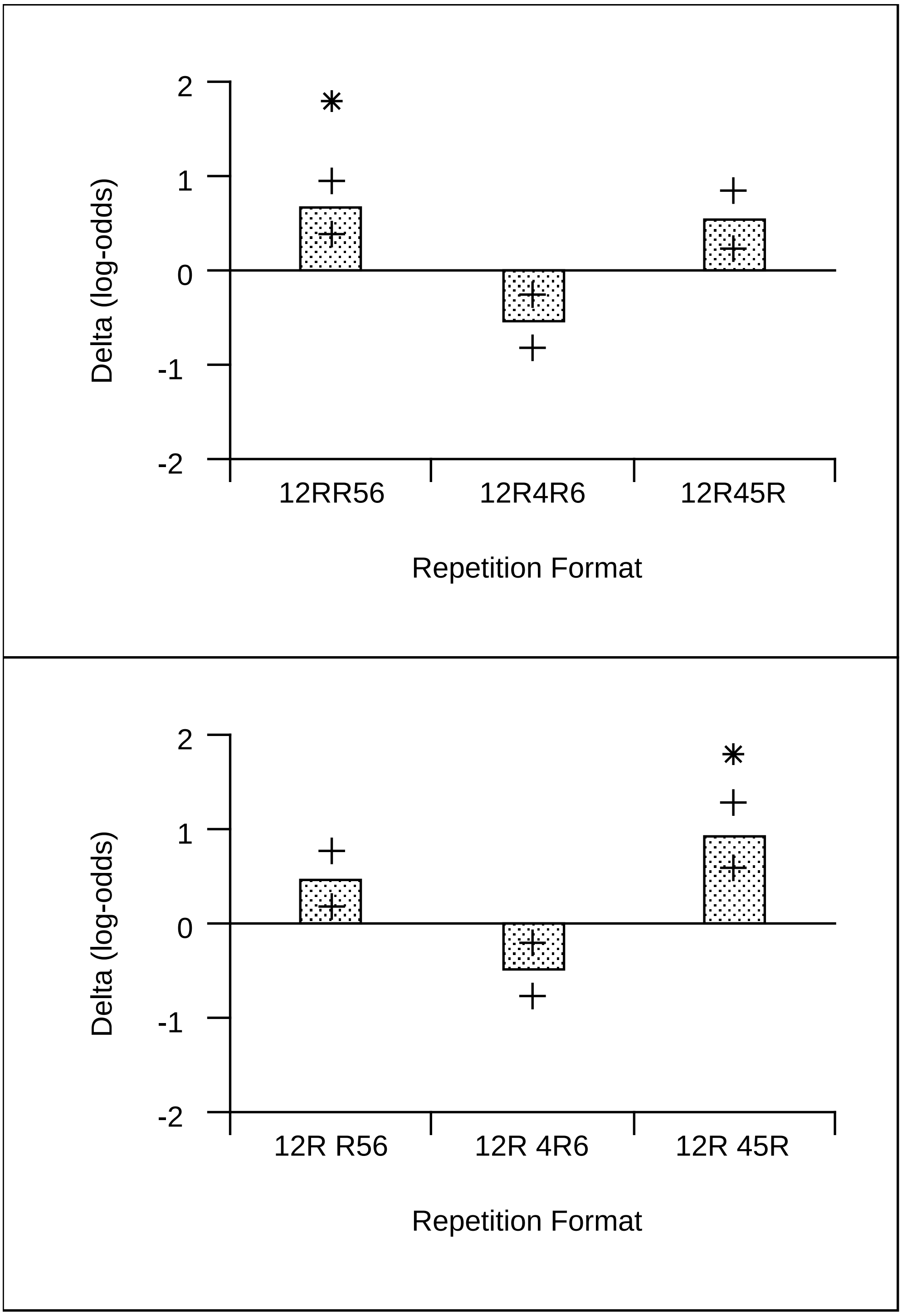



Figure 7-4: Delta under position-scoring as a function of repetition separation in ungrouped (upper panel) and grouped (lower panel) conditions of Experiment 7.
(Cross-hairs show standard error of delta scores above and below mean; asterisks indicate delta values significantly different from zero under Holm's method of multiple comparisons.)

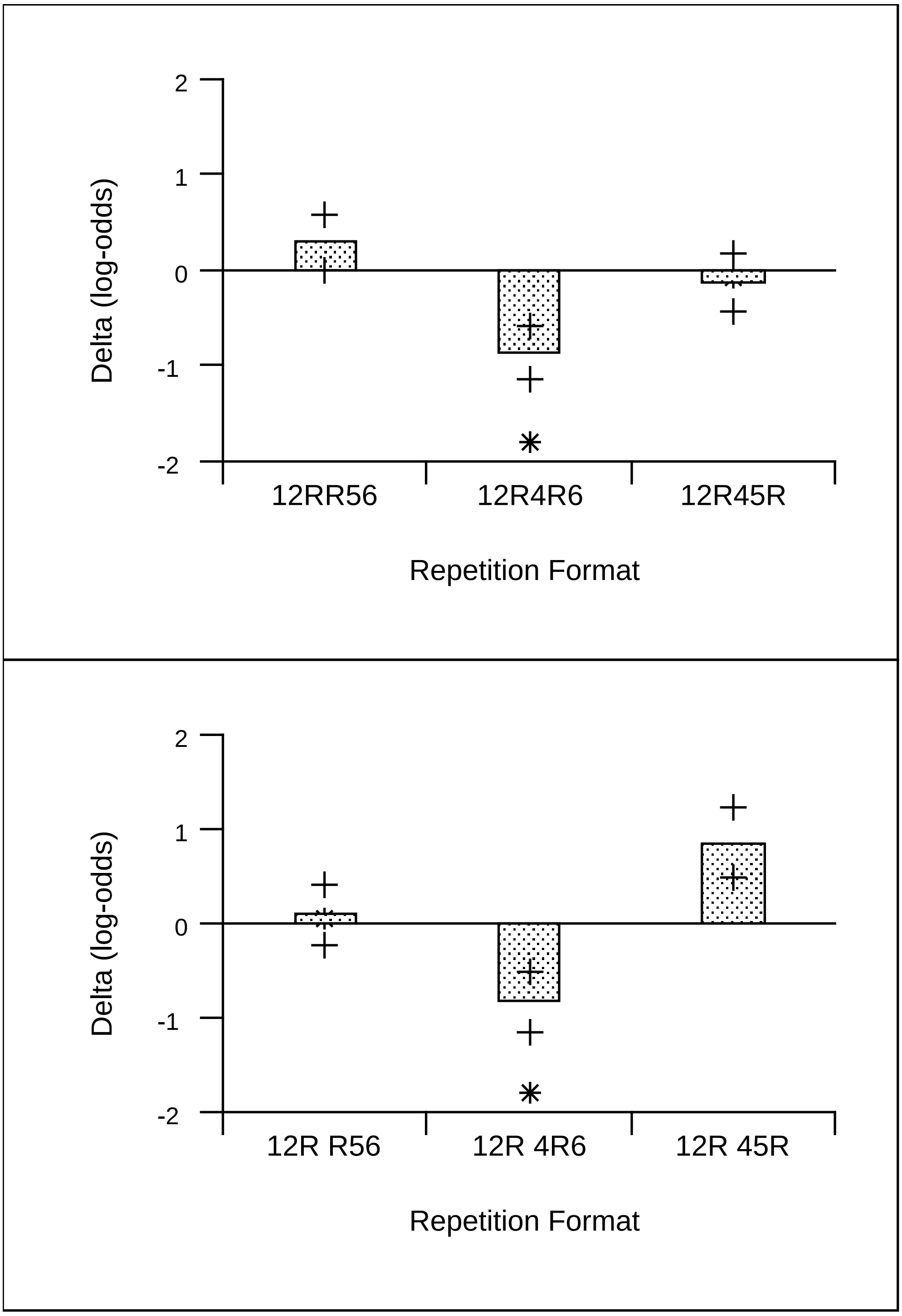



Figure 7-5: Delta under item-scoring as a function of repetition separation in ungrouped (upper panel) and grouped (lower panel) conditions of Experiment 7.
(Cross-hairs show standard error of delta scores above and below mean; asterisks indicate delta values significantly different from zero under Holm’s method of multiple comparisons.)

*Z(12)=2.00, p<.05* (and perhaps repetition format *12R4R6*, *Z(12)=1.73*, *p=.08*), but having little affect on repetition format *12RR56, Z(12)=0.21*, *p=.83,* in agreement with the results of Experiment 2 in Henson (1996b). Also in agreement with that experiment, repetition of a different item was the most common error in the repetition memory task for ungrouped repetition format *12RR56*. The other repetition formats produced approximately equal numbers of incorrect repetitions and failures to detect any repetition.

| | Repetition Format | | |
|---|---|---|---|
| Condition | 12RR56 | 12R4R6 | 12R45R |
| Ungrouped | .75<br>(.15) | .57<br>(.16) | .71<br>(.18) |
| Grouped | .75<br>(.15) | .66<br>(.16) | .80<br>(.14) |

Table 7-7: Correct repetition memory under item-scoring in the Experiment 7. (Calculated from weighted log-odds.)

**Correlation between Repetition Memory and Serial Recall**

Collapsing across repetition formats, contingency tables for performance on the repetition memory task and the serial recall task (under item-scoring) were determined for each subject. A combined test of significance of these tables showed an extremely high correlation between recall of repeated items and memory for their repetition, *Z(12)=20.30, p<.0001*. Table 7-8 shows the contingency table summed across subjects. Unfortunately, the correlation was so high that correct recall in the absence of correct repetition memory and correct repetition memory in the absence of correct recall were so rare that the direction of causality remained unclear. In other words, there was no direct way of telling whether recall depended on repetition memory, repetition memory depended on recall, or both factors influenced each other. If, for example, performance on the repetition memory task depended on serial recall of the whole list (overtly in the present experiment; covertly in Experiment 2 of Henson, 1996b), then there would be no need to postulate a separate repetition memory.

Indirect support for a separate repetition memory came from the observation that an item incorrectly repeated in serial recall was often the same item that was repeated on the

| | Repetition Memory | |
|---|---|---|
| | Correct | Incorrect |
| Serial Recall | | |
| Correct | 521 | 17 |
| Incorrect | 25 | 157 |

Table 7-8: Number correct and incorrect under item-scoring in the repetition memory and serial recall tasks of Experiment 7.

previous trial (and yet the same item was never repeated in the lists of two successive trials). This might be expected if a separate repetition memory were prone to proactive interference, but would not be expected if repetition memory were based solely on serial recall of a list.

Finally, a considerable number of the 22% of occasions where incorrect serial recall was accompanied by incorrect repetition memory were cases where subjects recalled the wrong item twice, and reported remembering that item as being repeated. Prima facie, this would support the hypothesis that repetition memory was based on what was recalled in the serial recall task. However, such repetition errors in serial recall could equally well arise from a failure of repetition memory. Furthermore, though there may well be contamination of repetition memory from prior serial recall in the present experiment, this is not evidence against the existence of a separate repetition memory (which may be updated during recall).

**Discussion**

The present experiment was an important, within-subjects replication of the repetition effects found in Experiment 6. The pattern of delta values was similar, except for a general increase in delta across all conditions. There were probably two reasons for this increase in delta: greater attention towards repetition, as required by the concurrent repetition memory task, and the smaller set of repetition formats used. Nevertheless, the robustness of repetition inhibition was confirmed by the fact that it was still found for two-apart repeated elements in a situation where subjects were required to explicitly detect and remember repetition of items.

The present experiment was also an important extension of the results of Experiment 2 in Henson (1996b), confirming that grouping increases repetition memory for repeated elements at the end of groups, but not for repeated elements that straddle a group boundary. By pulling performance off ceiling levels, the present experiment also showed that increasing repetition separation can impair repetition memory, at least across repetition formats *12RR56* and *12R4R6*. The fact that repetition format *12R45R* showed high levels of repetition memory in the ungrouped as well as grouped condition probably reflected some spontaneous grouping in the ungrouped condition (Chapter 3).

In addition to corroborating the previous experiments, the present experiment went further to correlate serial recall and repetition memory on a trial-by-trial basis. The extremely high correlation leaves no doubt that the two are interdependent. The direction of this dependency is somewhat unclear, though indirect evidence suggested that a separate repetition memory does influence serial recall. One reason was that the same item was often repeated in two consecutive trials, which would be expected if a separate repetition memory were prone to proactive interference, but not otherwise. Another reason was that the serial recall task showed a significant effect of retention interval, whereas the repetition memory task did not. This might be expected if a separate repetition memory (an item memory) were longer lasting than the memory underlying serial recall (a serial memory). Taken together, these facts support the hypothesis that repetition memory plays an important role in repetition facilitation and repetition inhibition.

The most surprising difference between the repetition effects in the present experiment and those in Experiment 6 concerned repetition format *12R45R* when grouped. Under item-scoring, this condition showed a delta value close to zero in Experiment 6, and yet showed a delta value considerably greater than zero in the present experiment. One reason could be the increased attention to repetition in the present experiment, as discussed above. However, an alternative account was offered in Henson (1996b) in terms of *repetition schema* (i.e., memory for the structure of repetition in a list); an account elaborated here in the General Discussion.

In summary, the present experiment reinforced the effects of repetition separation and grouping on repetition memory and serial recall. It also provided reasonably good support, given the problems of measuring repetition memory on-line, for the role of a separate repetition memory in serial recall. That role may be to ensure that an item is recalled twice, preventing repetition inhibition. In addition, special types of repetition memory, such as repetition tagging and repetition schemata (Henson, 1996b) may further cause repetition facilitation. Repetition facilitation at the end of groups in particular may be attributable to the ease of abstracting a repetition schema for repeated elements in these salient positions.

## Repetition Inhibition

The previous experiment suggested that failure to remember which item was repeated may result in repetition inhibition. Anything that affects the probability of detecting and remembering repetition, such as repetition separation and grouping, will therefore affect the magnitude of repetition inhibition. However, the question remains as to why failure to remember the repetition of an item should lead repetition inhibition in the first place. There was no doubt that subjects in the previous experiments could vocalise and therefore presumably encode both occurrences of a repeated item. Yet why did they often fail to recall more than one occurrence?

One possibility is that people fail to repeat a previous response because of output interference: The act of recalling an item in the past makes it less available for recall in the future. As Jahnke remarked, repetition inhibition "...is, at least in part, a result of interference arising from the act of sequential recall..." (Jahnke, 1969a, p. 620). Jahnke supported this claim with data suggesting that repetition inhibition was stronger for the second repeated element to be recalled, rather than the second repeated element presented, whether that was in forward or backward recall. Output interference also explains why repetition inhibition is absent when recall of both repeated elements is unnecessary, such as in probe recognition (Wolf & Jahnke, 1968) or probed recall (Jahnke, 1970). If output interference were an automatic process, repetition inhibition would be expected even when the first repeated element is a redundant, response prefix (Crowder, 1968b). Moreover, repetition inhibition would still be expected when subjects are well aware that repetition of responses is necessary (Crowder, 1968a; Hinrichs et al., 1973; Jahnke, 1969b).

There is an alternative to the output interference hypothesis however. Hinrichs et al. (1973) suggested that repetition inhibition may not reflect the operation of memory per se, but rather the guessing strategies used by subjects when their memory has failed. If people have a default reluctance to repeat themselves, they would be biassed against guessing a repeated item. In other words, they would be more likely to recall a control element correctly from a lucky guess than a repeated element. Thus repetition inhibition may arise not so much from impaired performance on repetition lists because of output interference, but from improved performance on control lists due to more successful guessing. This guessing hypothesis can not only account for most of the findings above, but is supported by further findings that are troublesome for the output interference hypothesis.

Firstly, Greene (1991, Experiment 1) showed that repetition inhibition disappeared when subjects were instructed not to guess, by virtue of poorer recall of control elements, but not repeated elements. This is exactly the pattern predicted by the guessing hypothesis. Secondly, when the vocabulary size is increased, repetition inhibition is reduced (e.g., Hinrichs et al., 1973). Again, this reduction comes from poorer recall of control elements; performance on repeated elements remains unaffected (Jahnke, 1974). According to the guessing hypothesis, a larger vocabulary reduces the probability of guessing a control element correctly. Finally, the fact that overt output is not required for repetition inhibition was demonstrated by Mewaldt and Hinrichs (1977), who found repetition inhibition in a situation where subjects had to report the missing item in a modified Cloze task. This was confirmed by Greene (1991, Experiments 2 and 3), who found that repetition inhibition could occur in a partial report task, where recall of only one repeated element was required. Importantly, this repetition inhibition was contingent on the remaining items being displayed during recall, to bias guesses against these items.

Jahnke (1972b) suggested that another important factor in repetition inhibition is proactive interference. He showed that when the experimental vocabulary was large enough, such that an item never occurred in more than one trial, there was no repetition inhibition. In other words, it appeared that repetition inhibition depended on intertrial repetition as well as intratrial repetition. Jahnke also showed that, when there was intertrial repetition, repetition inhibition was normally absent on the first trial, and increased over subsequent trials, again suggesting a role for proactive interference. However, the role of proactive interference is far from clear. Walsh and Schwartz (1977) showed that repetition inhibition was unaffected by a category shift across trials, unlike proactive interference. More importantly, the pattern of results of Jahnke (1972b) can also be explained by the guessing hypothesis. The larger vocabulary needed to prevent intertrial repetition will necessarily reduce the probability of guessing control elements, and hence reduce repetition inhibition. With a small vocabulary, the build up of repetition inhibition over trials can attributed to subjects gradually learning the vocabulary and hence constraining their range of sensible guesses.

Several puzzles remain even for the guessing hypothesis however. One puzzle is why subjects are still reluctant to guess an item they have already recalled when they are well aware that lists can contain repeats (e.g., Jahnke, 1969b). Though Mewaldt and Hinrichs (1977; Hinrichs & Mewaldt, 1977) showed that repetition inhibition was reduced when subjects experienced greater frequencies of repetition, it was clearly not eliminated. A second puzzle is that, in direct contradiction to the guessing hypothesis, subjects in the Walsh and Schwartz (1977) study reported no conscious avoidance of guessing repeated items during debriefing. Indeed, they often reported taking into account the presence of a repeated item when guessing. A third puzzle is why Jahnke (1972b) found that most errors on critical positions were omissions, rather than the substitutions predicted by guessing accounts. Jahnke (1974) also failed to find a correlation between the number of errors made by each subject and the magnitude of their repetition inhibition effect. If most errors were guesses (as would be expected from Experiments 4 and 5), a guessing bias would predict a strong correlation, with more errors resulting in a greater magnitude of repetition inhibition.

One final puzzle concerning the guessing hypothesis is that Walsh and Schwartz (1977), unlike Greene (1991), failed to find a significant effect of guessing instructions on repetition inhibition. Greene argued that Walsh and Schwartz used large vocabularies, which tend to reduce the magnitude of repetition inhibition (Hinrichs et al., 1973), and hence would have reduced the probability of Walsh and Schwartz observing a significant effect of instructions. However, the fact remains that Walsh and Schwartz still found considerable repetition inhibition even with strict instructions not to guess. Furthermore, Experiment 7 used a very small vocabulary and not only found significant repetition inhibition with instructions not to guess, but also with instructions specifically to remember which item was repeated.

One possible solution to these puzzles is that people are not always sure of when exactly they are guessing and when they are not. This is why instructions not to guess may not always be effective. In addition, when people do guess, it may not be that they are consciously avoiding guesses that would repeat previous responses, but that previous responses simply do not come to mind as possible guesses. In other words, both the output interference hypothesis and the guessing hypothesis can assume that the unavailability of repeated elements comes from an automatic, unconscious bias. In this case, the difference boils down to whether this bias causes forgetting of repeated elements (the output interference hypothesis), or prevents guessing of repeated elements already forgotten (the guessing hypothesis).

## Experiment 8

The aim of Experiment 8 was to test the output interference and guessing hypotheses. Rather than instructing subjects not to guess, they were asked to indicate which of their responses were guesses, to see whether these guesses did present a bias against repeated elements. This provides a test of the guessing hypothesis. However, given that subjects may not always be certain of what constitutes a guess (Chapter 6), they were further asked to indicate responses that they were simply not sure about. Both these confidence ratings (guesses and uncertain responses) were measured on-line during recall, through subjects moving up and down an array of response boxes in the same manner as in Experiments 4 and 5. A bias towards control elements in uncertain responses would further support the guessing hypothesis. However, if significant repetition inhibition remained even when both guesses and uncertain responses were removed from analysis, then there would also be support for the output interference hypothesis.

### Method

#### Subjects

Twelve subjects from the APU Subject Panel were tested; three were men, nine were women and their mean age was thirty years.

**Materials**

The same materials were used as in Experiment 7.

**Procedure**

The procedure was similar to that of Experiment 7, except for the following differences. Firstly, subjects were only tested on serial recall of lists; there was no additional requirement to remember which item was repeated. Secondly, subjects were not instructed to avoid guessing, but rather could indicate three levels of confidence for each response: *confident*, *unsure* and *guess*. The confidence of a particular response was indicated by where subjects wrote it in an array of three rows: The top row was used for confident responses, the middle row for unsure responses and the bottom row for guesses. Subjects were told they could move up and down the rows as much as they liked, providing they always gave exactly one response per column. In other words, they were always required to give six responses (omissions were not allowed), even if that meant guessing randomly from the vocabulary. The whole experiment took approximately 50 minutes.

**Results**

In brief, position-scoring of all responses replicated the results of Experiment 6, except that significant repetition facilitation was again found for repeated elements at the end of groups, as in Experiment 7. Item-scoring revealed exactly the same pattern of repetition inhibition as in Experiment 6, whether or not guesses were included in the analysis. Further exclusion of uncertain responses removed repetition inhibition for some conditions, but significant repetition inhibition remained for two repetition formats when grouped. These results support a role for both guessing and output interference in repetition inhibition.

**Position-scoring of Serial Recall**

With all responses included, the probability of recalling either critical element on the two critical positions was calculated for both repetition and control lists. A four-way ANOVA on log-odds showed a significant effect of repetition format, $F(2,255)=6.36$, $p<.005$, grouping, $F(1,255)=23.38$, $p<.001$, and retention interval, $F(1,255)=7.82$, $p<.01$. Like Experiment 7, there was no significant effect of list type, $F(1,255)<1$, but this factor did interact with repetition format, *F(1,255)=13.20, p<.001*. There were no further significant interactions. Collapsing across retention interval, these probabilities are shown in Table 7-9.

| | Repetition Format | | |
|---|---|---|---|
| Condition | 12RR56 | 12R4R6 | 12R45R |
| Ungrouped | | | |
| Repeated | .65 (.16) | .36 (.17) | .49 (.19) |
| Control | .41 (.19) | .50 (.18) | .39 (.18) |
| Grouped | | | |
| Repeated | .73 (.16) | .55 (.18) | .94 (.06) |
| Control | .74 (.15) | .79 (.14) | .77 (.15) |

Table 7-9: Correct recall of critical elements under position-scoring in Experiment 8. (Calculated from weighted log-odds.)

Figure 7-6 shows delta against repetition format. Tests of weighted log-odds showed delta was significantly different from zero in all cases, *Z(12)>2.70, p<.01*, except for one-apart repeated elements in grouped lists, and three-apart repeated elements in ungrouped lists, *Z(12)<1.73, family-wise p>.05*. Tests of weighted delta values showed that grouping significantly reduced delta for one-apart repeated elements, *Z(12)=2.80, p<.01*. Grouping did not significantly reduce delta for two-apart repeated elements, *Z(12)=1.19, p=.23*, or significantly increase delta for three-apart repeated elements, *Z(12)=1.72, p=.09*, though both these trends were in exactly the same direction as the nonsignificant trends in Experiment 6 and Experiment 7. The fact that three experiments show these trends seems to warrant the general conclusion that grouping impairs recall of repeated elements at different positions within groups and improves recall of repeated elements at the end of groups.

These results replicated those of Experiments 6 and 7, even when subjects were forced to guess. Thus repetition facilitation and repetition inhibition were robust to increased levels of guessing. However, the main purpose of the present experiment was to see if these effects,

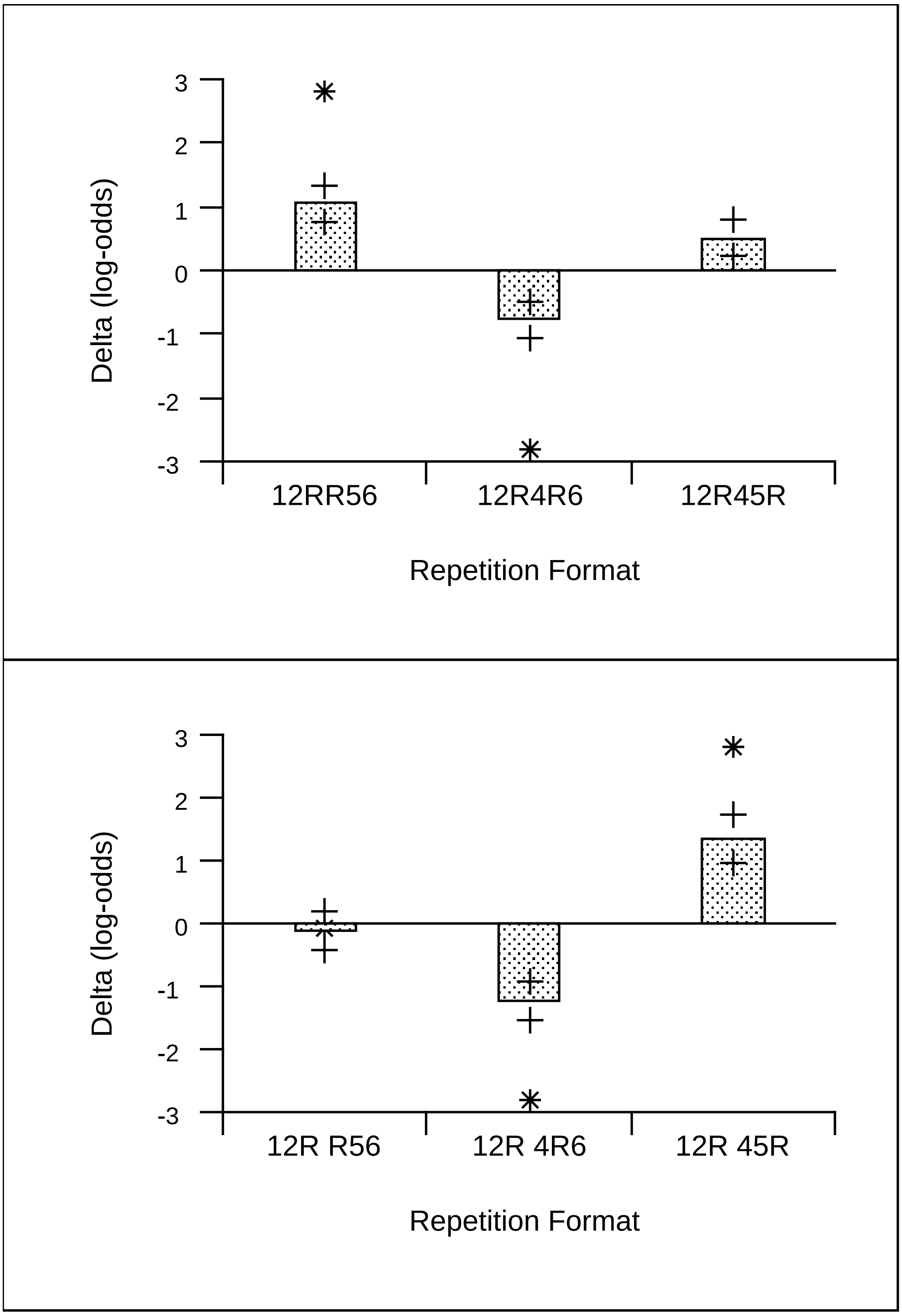



Figure 7-6: Delta under position-scoring as a function of repetition separation in ungrouped (upper panel) and grouped (lower panel) conditions of Experiment 8.
(Cross-hairs show standard error of delta scores above and below mean; asterisks indicate delta values significantly different from zero under Holm’s method of multiple comparisons.)

particularly repetition inhibition, were robust to decreased levels of guessing. This possibility was tested under item-scoring, by successively removing less confident responses.

**Uncertain Responses**

Over all subjects, approximately 9% of responses were indicated as guesses, and an additional 24% were indicated as unsure. Interestingly, these percentages were identical for both repetition and control lists. In all subsequent discussion, the term *uncertain responses* will refer to the 33% of responses that subjects either guessed or were unsure about.

A three-way ANOVA on the log-odds of an uncertain response showed significant effects of grouping, $F(1,253)=121.42$, $p<.001$, retention interval, $F(1,253)=134.34$, $p<.001$, and output position, $F(5,253)=27.84$, $p<.001$. Grouping interacted significantly with output position, $F(5,253)=3.67$, $p<.005$, but none of the other interactions was significant. The longer retention interval produced a higher frequency of uncertain responses, as would be expected. Generally, uncertain responses increased towards the end of recall, though this pattern was modified by grouping, which reduced uncertain responses more for the second group than the first, and tended to equate the certainty of responses within groups (Chapter 6).

**Item-scoring of Serial Recall**

To investigate the impact of uncertain responses on repetition inhibition, a four-way ANOVA was conducted on the log-odds of recalling two critical elements anywhere. Collapsing across retention interval, the four factors were list type, repetition format, grouping and response certainty (Table 7-10). The three levels of response certainty were either to include all responses, to include all responses except guesses, or to include all responses except uncertain responses. There were significant effects of list type, $F(1,389)=84.38$, $p<.001$, repetition format, $F(2,389)=19.32$, $p<.001$, grouping, $F(1,389)=203.74$, $p<.001$, and response certainty, $F(2,389)=152.14$, $p<.001$. As expected, repetition format interacted with both list type $F(2,389)=21.67$, $p<.001$, and grouping, $F(2,389)=16.05$, $p<.001$. Interestingly, response certainty also interacted with both list type, $F(2,389)=9.95$, $p<.001$, and grouping, $F(2,389)=4.71$, $p<.01$. Two three-way interactions were significant, that between list type, grouping and repetition format, $F(2,389)=15.68$, $p<.001$ and that between list type, grouping and response certainty, $F(2,389)=6.07$, $p<.005$. No other interactions approached significance.

| Condition | All Responses 12RR56 | All Responses 12R4R6 | All Responses 12R45R | Without Guesses 12RR56 | Without Guesses 12R4R6 | Without Guesses 12R45R | Without Uncertain Responses 12RR56 | Without Uncertain Responses 12R4R6 | Without Uncertain Responses 12R45R |
|---|---|---|---|---|---|---|---|---|---|
| Ungrouped | | | | | | | | | |
| Repeated | .72 (.15) | .51 (.16) | .59 (.19) | .71 (.15) | .46 (.16) | .53 (.19) | .58 (.16) | .25 (.13) | .34 (.18) |
| Control | .82 (.13) | .88 (.10) | .83 (.12) | .68 (.18) | .72 (.16) | .69 (.16) | .33 (.16) | .31 (.17) | .28 (.16) |
| Grouped | | | | | | | | | |
| Repeated | .78 (.14) | .62 (.18) | .94 (.06) | .75 (.15) | .60 (.17) | .89 (.09) | .59 (.18) | .44 (.18) | .74 (.17) |
| Control | .93 (.07) | .91 (.08) | .91 (.08) | .90 (.09) | .83 (.12) | .85 (.12) | .65 (.20) | .67 (.17) | .62 (.18) |

Table 7-10: Correct recall of critical elements under item-scoring in Experiment 8. (Calculated from weighted log-odds.)

The interaction between response certainty and list type reflected a greater reduction in the probability of recalling control elements than repeated elements when guesses and uncertain responses were removed. This is consistent with the guessing hypothesis of Hinrichs et al. (1973). The interaction between response certainty and grouping reflected a greater reduction in the probability of recalling critical elements when guesses and uncertain responses were removed from ungrouped lists than from grouped lists. This is expected if grouping not only improves performance, but also increases confidence levels. The interaction between response certainty, grouping and list type reflected a greater interaction between response certainty and list type in ungrouped lists than grouped lists.

Delta for all responses is shown in Figure 7-7, delta without guesses is shown in Figure 7-8, and delta without uncertain responses is shown in Figure 7-9. With all the responses analysed, the pattern of delta values was identical to that in Experiment 6. Significant repetition inhibition was found for all repetition formats, $Z(12)>3.63$, $p<.001$,

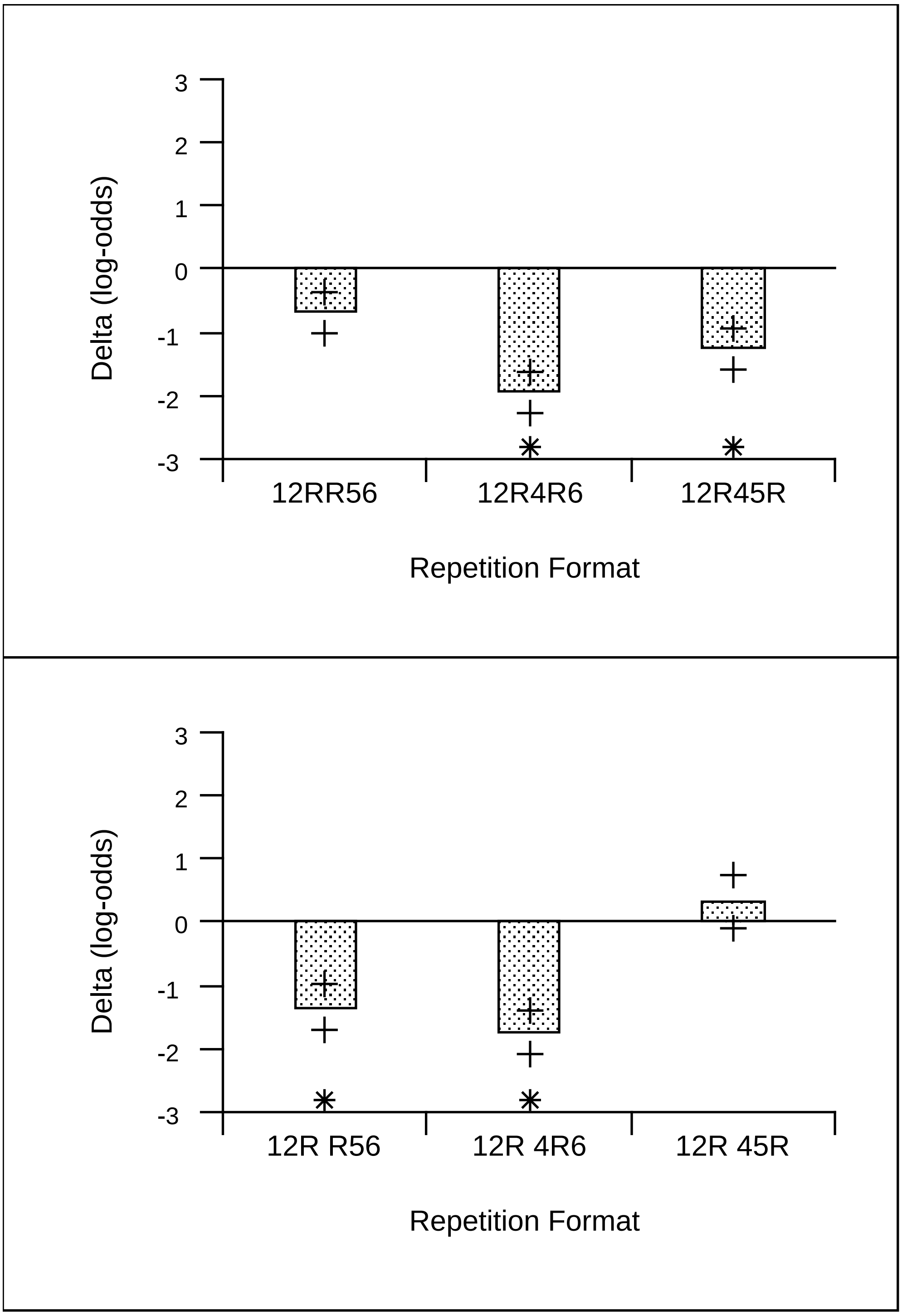



Figure 7-7: Delta under item-scoring, including all responses, as a function of repetition separation in ungrouped (upper panel) and grouped (lower panel) conditions of Experiment 8. (Cross-hairs show standard error of delta scores above and below mean; asterisks indicate delta values significantly different from zero under Holm's method of multiple comparisons.)

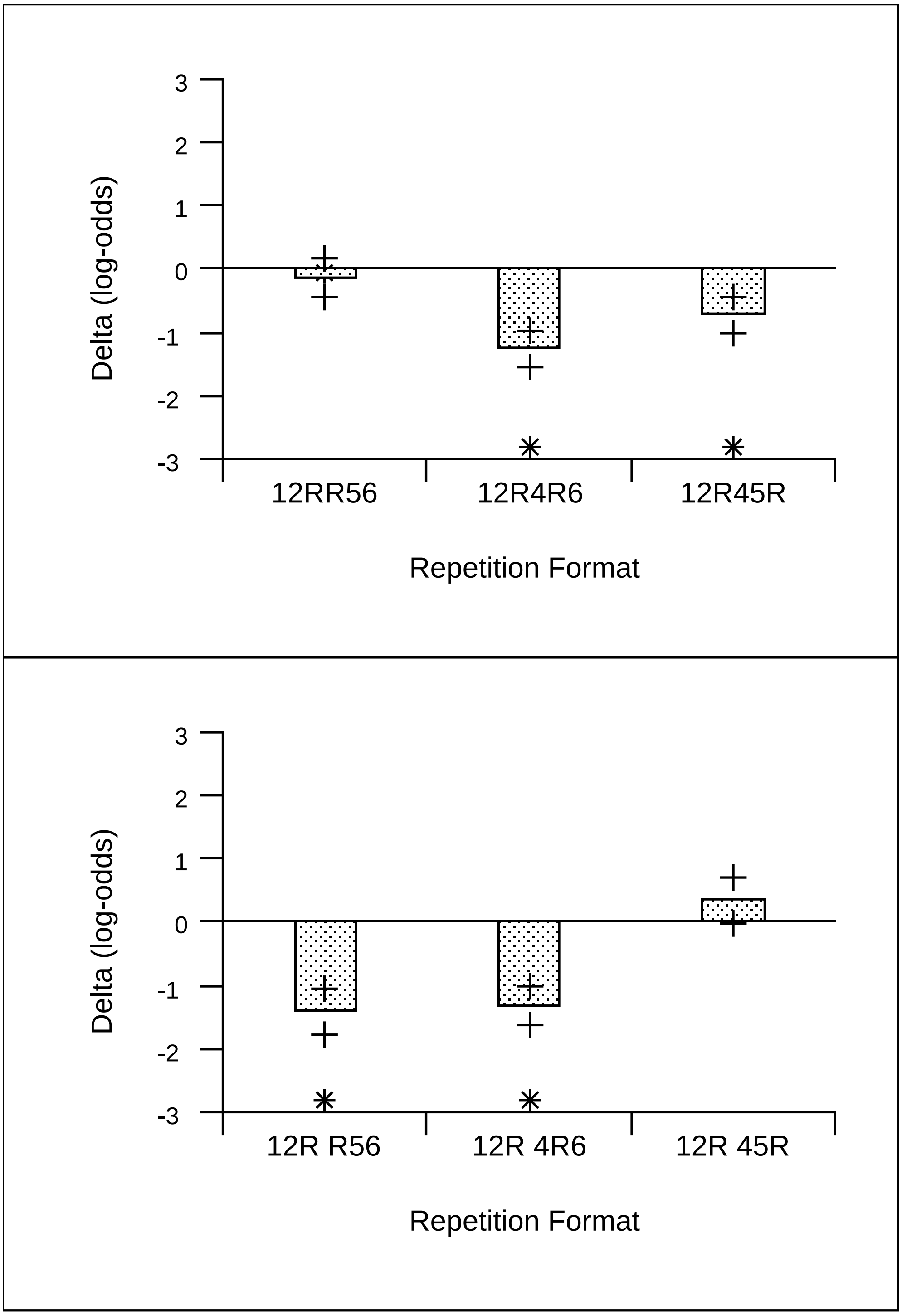



Figure 7-8: Delta under item-scoring, excluding guesses, as a function of repetition separation in ungrouped (upper panel) and grouped (lower panel) conditions of Experiment 8. (Cross-hairs show standard error of delta scores above and below mean; asterisks indicate delta values significantly different from zero under Holm's method of multiple comparisons.)

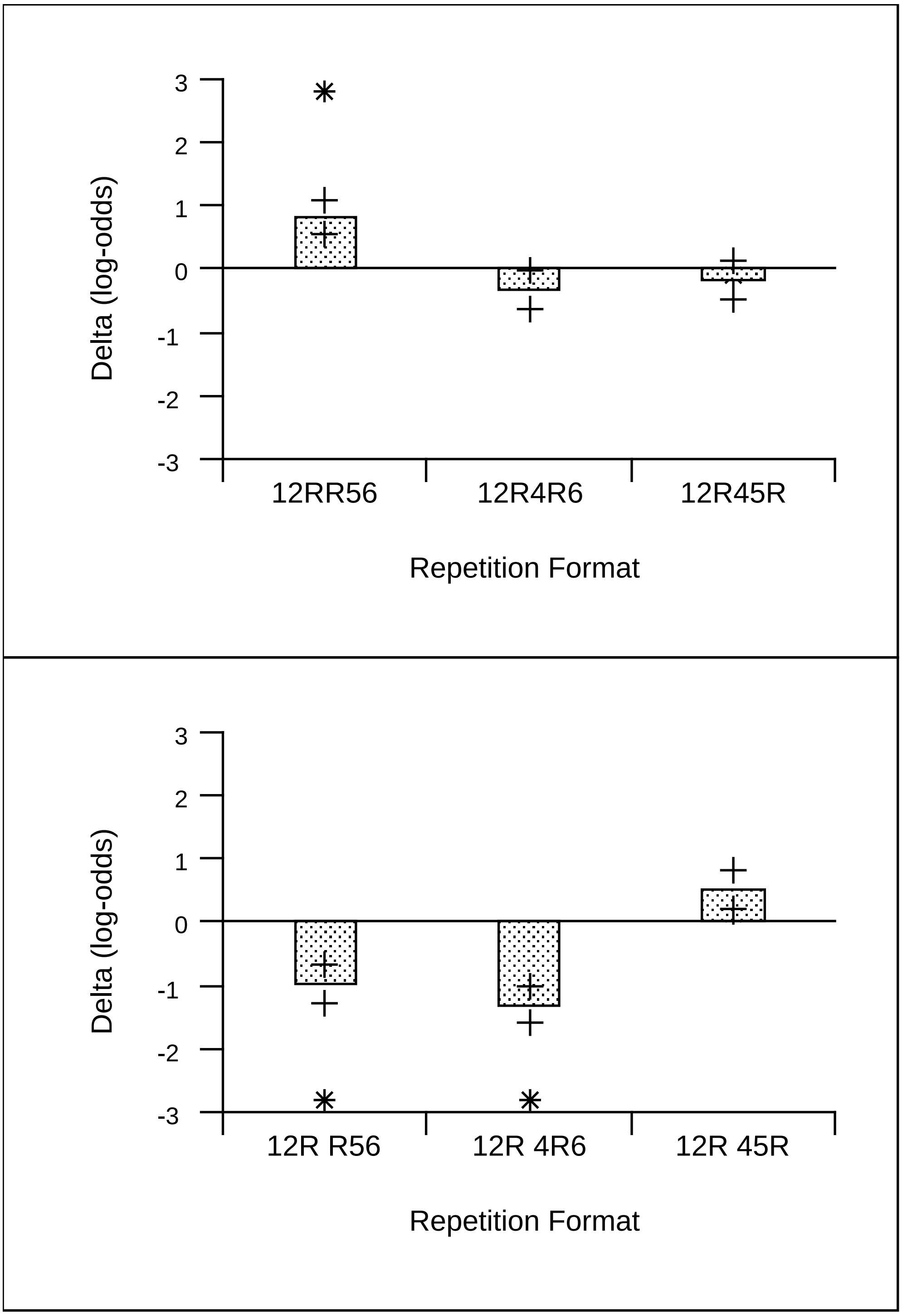



Figure 7-9: Delta under item-scoring, excluding uncertain responses, as a function of repetition separation in ungrouped (upper panel) and grouped (lower panel) conditions of Experiment 8. (Cross-hairs show standard error of delta scores above and below mean; asterisks indicate delta values significantly different from zero under Holm's method of multiple comparisons.)

except one-apart repetition in ungrouped lists and three-apart repetition in grouped lists, *$Z(12)<2.16$, family-wise $p>.05$*.

With guesses removed, delta increased for all conditions. However, there was still significant repetition inhibition for all repetition formats, *$Z(12)>2.54$, $p<.05$*, except one-apart repetition in ungrouped lists and three-apart repetition in grouped lists, *$Z(12)<0.95$, family-wise $p>.05$*. In other words, removing guesses did not change the pattern of significant results.

With all uncertain responses removed, delta increased further still. There was no longer significant repetition inhibition for any condition except for one-apart and two-apart repeated elements in grouped lists. Nevertheless, repetition inhibition for both these conditions was still highly significant, even under a Bonferroni correction, *$Z(12)>3.18$, $p<.008$*. In addition, there was significant repetition facilitation for one-apart repetition in ungrouped lists, *$Z(12)=2.94$, $p<.008$*. No other delta values differed significantly from zero, *$Z(12)<1.65$, family-wise $p>.05$*.

**Discussion**

The results of the present experiment confirm the guessing hypothesis of Hinrichs et al. (1973), that guessing strategies bias recall against repeated elements and in favour of control elements. This bias was apparent by successive removal of responses that subjects guessed or were simply unsure about. Though one might not want to call all such uncertain responses “guesses”, the effect of removing them was to reduce dramatically the probability of recalling control elements; the probability of recalling repeated elements was not affected to the same extent. Thus, a considerable part of repetition inhibition reflects an increased likelihood of guessing control elements relative to repeated elements.

However, the present data also suggest that guessing strategies are not the only cause of repetition inhibition, because highly significant repetition inhibition remained under some conditions even when all uncertain responses were removed. This is consistent with the significant repetition inhibition found when subjects were instructed not to guess in Experiment 7. It is surprising then that this repetition inhibition was not found by Greene (1991) with similar instructions. One reason may be the more sensitive measure of repetition inhibition used in the present study. Greene compared only error rates on critical positions in serial position curves, but did not take into account the probability of recalling two critical items, nor the modified control scoring introduced in the present study. It is noteworthy that the figure of 33% uncertain responses in the present experiment was much greater than the percentage of guesses suggested by Greene's data (no more than 10%). This suggests that the present experiment, as a test of guessing theories, erred on the conservative side, if at all.

Note that the residual repetition inhibition found after removing uncertain responses is unlikely to be because subjects were simply less certain of repeated responses than control responses. For example, a natural reluctance for people to repeat themselves may make them more likely to indicate a second repeated element as uncertain than a second control element. However, by this account, repetition inhibition should have increased rather than decreased as uncertain responses were removed. Moreover, the overall numbers of uncertain responses were almost identical in control lists and repetition lists. It seems more likely therefore that people simply did not recall the second repeated element.

How can this persistence of repetition inhibition be explained? The hypothesis outlined below is that both guessing strategies and output interference play a role. Moreover, though originally presented as competing hypotheses, both can be viewed as consequences of a more general process of suppression during serial recall (as in SEM; Chapter 5).

After an item is recalled once, its type representation is assumed to be suppressed, reducing its probability of output again in the immediate future, much like Crowder's (1968b) output interference hypothesis. This will cause forgetting of the second repeated element to be recalled (and perhaps its replacement by a unsuppressed context element). If this suppression is an automatic, unconscious process, repetition inhibition can remain even when both guesses and uncertain responses are removed. Moreover, suppression can also cause a guessing bias. Suppression of an item already recalled may prevent that item coming to mind when one does decide to guess (as in SEM's guessing in Chapter 6). This would explain why control elements are greater affected by the removal of uncertain responses than are repeated elements. Thus suppression can not only cause forgetting of a repeated item, but also prevent its guessing.

The unconscious nature of suppression can explain why repetition inhibition remains even when people are fully aware that repetition is necessary in order to recall correctly, and why people do not always report any conscious bias against guessing a repeated item (Walsh & Schwartz, 1977). Because suppression can prevent retrieval of a repeated item, as well as prevent its guessing, suppression also explains why repetition inhibition can remain when people are instructed not to guess (Experiment 7; Walsh & Schwartz, 1977) and why repetition inhibition can result from omission as well as substitution errors (Jahnke, 1972b). The automatic nature of suppression is further supported by an unpublished study by Baddeley and Andrade (1996). They found that the magnitude of repetition inhibition was not affected by a concurrent secondary task, which would presumably attenuate any conscious guessing bias.

In SEM, suppression is usually assumed to be partial rather than complete. In other words, suppression will not always prevent correct recall of both repeated elements: It simply reduces the probability of recalling both. More accurately, SEM assumes that suppression is temporary, wearing off over time, to explain repetitions even when there are no repeated items in a list (Chapter 5). In the present experiment for example, approximately 24% of control lists contained such repetition errors. Though a pure guessing hypothesis can appeal to a similar notion of forgetting of previous responses over time, this is problematic for its account of repetition inhibition. If such a significant baseline probability of forgetting previous responses were operating in repetition lists too, a much higher incidence of guessing would be required to explain the magnitude of repetition inhibition. Interestingly, if the elderly were less effective at suppressing or inhibiting previous responses (e.g., Hasher & Zacks, 1988; Koriat, Ben-Zur & Sheffer, 1988), they should not only make more repetitions in control lists, but, somewhat ironically, be less prone to repetition inhibition in repetition lists.[2] The fact that repetition inhibition is a within-subject measure makes it particularly attractive to the study of developmental changes in inhibitory processes.

Considerable numbers of erroneous repetitions were found in repetition lists too. In the present experiment, approximately 32% of repetition lists contained repetition of the wrong item. The greater percentage of such errors in repetition lists than control lists is usually taken as evidence that people sometimes detect a repetition event, but forget which item was

2. Preliminary evidence suggests that age does indeed reduce the repetition deficit in production tasks analogous to serial recall, but not in perception tasks (i.e., repetition blindness), in which age accentuates rather than reduces the repetition deficit (MacKay, Abrams & Pedroza, 1996).

repeated (Henson, 1996b). The suppression account suggests an alternative, or perhaps additional, reason: When people fail to recall the second repeated element due to suppression, they are likely to substitute another, less suppressed item. The smaller set of such items in repetition lists than control lists means a repetition error is more likely to result in the former.

The refractory nature of suppression may also explain why Crowder (1968a, 1968b) found greater repetition inhibition when repeated elements were three positions apart than when they were more than three positions apart: the further apart the repeated elements, the longer the time for suppression to wear off. One reason why this trend was not found in the present experiments may be because of the smaller range of repetition separations tested. Also, in addition to allowing greater recovery from suppression, increasing repetition separation will reduce the chance of detecting repetition (Lee, 1976b; Henson, 1996b). The trade-off between these two factors may depend on the exact repetition formats used. This reinforces the potentially complex nature of item repetition effects in serial recall.

Finally, an alternative explanation of present results is worth discussing. The repetition inhibition remaining for some grouped repetition formats, in spite of the removal of uncertain responses, might result when people thought a different item was repeated. Henson (1996b) showed such errors of repetition memory are quite common for these formats. Can incorrect repetition memory cause repetition inhibition, by overriding repetition of the correct item with repetition of a different item? If so, there may be no need to appeal to the notion of suppression. However, this alternative account seems unlikely for several reasons.

Firstly, repetition errors were rare when uncertain responses were removed, so correct repetitions were not always replaced by incorrect repetitions. Secondly, the repetition memory errors in Henson (1996b) may have arisen when subjects detected a repetition, but forgot which item was repeated, and so resorted to guessing, or even covert serial recall of the whole list. In either case, the repetition memory errors would not cause repetition inhibition per se, but rather reflect situations where repetition inhibition would have resulted anyway. Finally, memory for an incorrect repetition can not be a sufficient reason for repetition inhibition because repetition inhibition still occurred in repetition formats where such errors were rare: where it was more likely for a complete failure to detect any repetition (e.g., repetition format

*1R345R* in Henson, 1996b). This is confirmed by Jahnke (1969b), who showed repetition inhibition remained even when reports containing repetitions were removed from analysis. Suppression seems the only way to explain the residual repetition inhibition in Experiment 8.

In summary, repetition inhibition may be attributed to two causes. One is output interference, which can cause forgetting of the second repeated element to be recalled. The other is a bias against guessing repeated items when an item is forgotten; a bias which may operate unconsciously as well as consciously. A similar argument for two causes underlying repetition inhibition was made by Arbuthnott (in press) for repetition effects in sequential arithmetic problems. Both conscious (Baddeley, Emslie, Kolodny & Duncan, 1995) and unconscious (Brugger, Monsch & Johnson, 1996) causes have also been suggested for people's failure to give appropriate numbers of immediate repetitions in random generation tasks. Nevertheless, it is possible that both output interference and guessing biases are consequences of a more general process of suppression. The automatic suppression of previous responses, assumed necessary for serial recall, can not only cause failure to retrieve an item more than once, but can also prevent it coming to mind should one decide to guess.

## General Discussion

The present series of experiments confirmed that the presence of repeated items has important effects on short-term, serial recall, even under a new, conservative scoring scheme. These effects were mainly restricted to the repeated elements (repetition facilitation and repetition inhibition); there was little evidence for a direct effect of repetition on surrounding context elements (repetition contamination). Furthermore, the effects of repetition facilitation and repetition inhibition were shown to interact in a reliable, yet complex manner with repetition separation and grouping. The complexity of this interaction suggests that several factors play a role. This is probably why there have been numerous demonstrations of repetition facilitation and repetition inhibition in the literature, and yet no comprehensive theoretical interpretation has emerged. A summary of the empirical findings related to the present chapter is given below, followed by one such attempt at a more comprehensive theory of item repetition effects. Finally, the effects are discussed in relation to models of serial recall from short-term memory, and SEM in particular.

**Summary of Empirical Findings**

All repetition effects in the present study were measured under a modified scoring scheme that treated repeated elements identically to control elements. This scheme overcomes a potential bias against the scoring of control elements, which may have caused an overestimation of repetition facilitation and underestimation of repetition inhibition in previous studies. This may be why the present experiments found repetition inhibition as soon as one context element intervened between two repeated elements, where previous studies reported repetition inhibition only after two or more intervening context elements (e.g., Crowder, 1968a; Lee, 1976b). The new scoring scheme may also explain why significant repetition inhibition was found in the absence of guesses, where Greene (1991) failed to find such an effect. Nevertheless, other results were in broad agreement with previous studies, and, given no theoretical reason to chose one scoring scheme over another (i.e., no accepted theory of whether repeated items are represented as types or tokens in short-term memory), the present scheme seems preferable as a conservative and unbiased method.

In the serial recall tasks, repetition facilitation reflected mainly superior positioning of two repeated elements relative to two control elements. It was found only for immediate repetition that did not straddle a group boundary (Experiments 6, 7, 8), and repetition at the end of groups (Experiments 7, 8). Repetition inhibition reflected mainly inferior recall of two repeated elements anywhere in a report. It was typically found for all repetition formats that did not show repetition facilitation (Experiments 6, 7, 8). Repetition inhibition was reduced by discouraging (Experiment 7) or removing (Experiment 8) guesses. Nevertheless, significant repetition inhibition remained even when all uncertain responses were removed from analysis (Experiment 8), and when subjects concentrated on remembering repetition (Experiment 7).

In the repetition memory tasks, the probability of detecting a repetition event was generally high, even with a concurrent serial recall task (Experiment 7), but decreased slightly as repetition separation increased (Henson, 1996b). The probability of reporting the correct repeated item was lower, and reporting repetition of a different item was common for small repetition separations (Experiment 7; Henson, 1996b). Correct positioning of repeated elements decreased more markedly as repetition separation increased (Henson, 1996b). The only exception to these effects of repetition separation occurred when lists were grouped: Repetition at the end of groups was then much better remembered (Experiment 7; Henson, 1996b). Surprisingly, repetition memory was not impaired for repetition across a group boundary (Experiment 7; Henson, 1996b). Memory for a repeated item correlated very highly with serial recall of that item (Experiment 7).

Apart from a general association between recall of critical elements and recall of their surrounding context elements (as would be expected given the interdependency between responses in serial recall), more specific measures showed little evidence of repetition contamination. The probability of errors following correct repeated elements was significantly greater than following correct control elements, but this was accompanied a greater probability of errors on the repeated elements themselves. There was a trend for more contamination errors to be associative intrusions in repetition lists than control lists, irrespective of grouping (Experiment 6). Unfortunately, such intrusions were so infrequent that this trend could not be tested further in Experiments 7 and 8.

## A General Theory of Item Repetition Effects

There are probably several possible interpretations of the repetition effects found in the present study. Rather than trying to enumerate all of them, one possible interpretation is outlined below. This general theory attempts to incorporate results from the present study with those from previous studies in the literature. Several important components of the theory, such as repetition tagging, repetition schemata, guessing strategies and response suppression, come from Henson (1996b). What follows is an attempt to bring these ideas together.

### Repetition Facilitation and Inhibition

The basic tenet of the theory is that repeated items face a negative bias against repetition during recall, which can be overcome in situations where their repetition is explicitly remembered. The negative bias, underlying repetition inhibition, has several forms. Firstly, people have a natural reluctance to repeat themselves. This will prevent them from guessing a repeated item if they have forgotten the answer to a problem (Hinrichs et al., 1973). Note that this might apply to a range of tasks; such a bias is not necessarily restricted to serial recall (Greene, 1991; Mewaldt & Hinrichs, 1977). Secondly, in serial recall, an additional unconscious bias operates. This is the automatic suppression of previous responses. Suppression of an item after its recall reduces its chances of being retrieved again, and may also prevent that item coming to mind as a guess. Such suppression is often assumed necessary for serial recall (Henson et al., 1996) and may be a general process in the sequencing of actions (Houghton & Tipper, 1996; MacKay, 1987). Repetition inhibition is therefore not necessarily found in tasks with no requirement to output both repeated elements (e.g., Jahnke, 1970; Wolf & Jahnke, 1968).

Both a guessing bias and a suppression process are sufficient to explain why repetition inhibition reflects inferior item recall of repeated elements (Experiments 6, 7, 8) and, in particular, the second repeated element to be recalled (Crowder, 1968a; Jahnke, 1969b; Wickelgren, 1965c). However, both are necessary to explain why discouraging or removing guesses decreases repetition inhibition (Experiments 7, 8; Greene, 1991), but does not eliminate it (Experiments 7, 8; Walsh & Schwartz, 1977). Furthermore, only a guessing bias can explain the effects of vocabulary size and number of trials (Hinrichs et al., 1973; Jahnke, 1969b, 1974), occasional repetition inhibition for the first repeated element (Jahnke, 1969a) and repetition inhibition in tasks other than serial recall (Greene, 1991; Mewaldt & Hinrichs, 1977). On the other hand, only an automatic suppression process can explain why people do not always report a bias against guessing repeated elements (Walsh & Schwartz, 1977), why many errors in recall of critical elements are omissions (Jahnke, 1972b), why repetition inhibition does not necessarily correlate highly with ability (Jahnke, 1974), and why repetition inhibition is unaffected by secondary distraction tasks (Baddeley & Andrade, 1996).

Recall of repeated elements can be aided when their repetition is explicitly remembered. In order to be remembered, the repetition event must first be detected. Though people may correctly encode both occurrences of a repeated item (inferred from the present study because subjects almost invariably vocalised both repeated elements), they do not automatically notice that an item has been repeated. Generally, the probability of detecting repetition is lower the greater the repetition separation (Henson, 1996b; Lee, 1976b). The most important aspect of repetition separation is the number of intervening items, rather than the absolute time, because repetition memory is just as accurate for repeated elements that are separated by a pause between groups as for repeated elements that are not (Experiment 7; Henson, 1996b). Nevertheless, grouping can affect the probability of detecting repetition in other cases. In particular, the distinctive nature of the end of groups (Chapter 5) improves detection of repeated elements at these positions (Henson, 1996b).

However, accurate repetition memory depends not only on detecting the repetition during presentation, but also remembering which particular item was repeated. Sometimes one can remember the repetition event, but forget which item was repeated. Indeed, subjects in the present experiments occasionally reported “knowing” that a repetition occurred, but not being sure which item was repeated. This explains why repetition inhibition remains when repetition is expected (Mewaldt & Hinrichs, 1977), monitored (Experiment 7) or even reminded (Jahnke, 1969b). Memory for the repeated item is assumed to be an item memory separate from the memory for the list itself (Experiment 7). Like Lee’s (1976b) repetition tags, this repetition memory affects retrieval rather than storage of the list. It is also prone to proactive interference, explaining the tendency for people to perseverate the repetition of a particular item across trials (Experiment 7).

When the repetition event is remembered, but the repeated item forgotten, people may guess at an item, or try to reconstruct the item via (covert) serial recall of the list. “Memory” for repetition of a different item can result in both cases. Such errors of repetition memory do not cause repetition inhibition therefore, but arise in situations where repetition inhibition would result anyway. Because repetition detection is better the closer the repeated elements, repetition of the wrong item will necessarily be more frequent in such cases (Henson, 1996b).

Accurate repetition memory can counteract the negative bias against repetition during recall. Correct memory for the repeated item will remove any conscious bias against guessing it, and perhaps overcome its suppression during recall (at least until it has been output twice). This will reduce any difference in recall of repeated and control elements. The magnitude of repetition inhibition will therefore depend mainly on the number of trials in which a repeated item is correctly detected and remembered, relative to the number of trials in which it is not detected or forgotten. This explains why the correlation between repetition memory and recall performance is so high (Experiment 7). In order to remove repetition inhibition completely, repetition memory must be accurate over all trials. Thus, even though repetition across a group boundary may be detected and remembered on most trials (Experiment 7; Henson, 1996b), it may not be accurate enough over all trials to overcome repetition inhibition (Experiment 6, 8).

While memory for the repeated item is necessary to prevent repetition inhibition, it is not sufficient to cause repetition facilitation. Repetition facilitation requires additional forms of repetition memory. One such memory is the tagging of immediate repetition. Such repetition tags are associated with a position in a list, and cause immediate repetition of the item recalled at that position (as in Houghton et al., 1994).[3] This increases the probability of recalling repeated elements above that of control elements, by ensuring that both repeated elements are recalled in adjacent positions. Indeed, such specialised coding of the immediate repetition of an action may have evolved specifically to overcome suppression during the execution of action sequences (e.g., MacKay, 1987).

If adjacent repetition is detected and tagged often enough over trials, repetition facilitation can emerge, sometimes under item-scoring (Experiment 8; Lee, 1976b), but most obviously under position-scoring (Experiment 6), given that there is no opportunity for the second repeated element to transpose with the context elements that follow it as there is for the second of two adjacent control elements. Repetition tagging is not applied to nonadjacent repeated elements however, because immediate repetition of the corresponding item during recall would result in the wrong order of items (e.g., *12R4R6* being recalled as *12RR46*). Neither is tagging is applied to adjacent repeated elements that straddle a group boundary, because immediate repetition would then interfere with the grouped organisation of recall. People tend to pause between recalling groups, and the immediate repetition of an item at the end of one group may impair retrieval of the next, by disrupting the grouped organisation of recall (e.g., *12R..R56* recalled as *12RR..56*). This explains why repetition facilitation does not occur in these situations (Experiments 6, 7, 8), even though detection and memory for the repetition may actually be improved (Experiment 7; Henson, 1996b). For further details about repetition tagging, see Henson (1996b).

3. These repetition tags differ from those of Lee (1976b) and Drewnowski (1980a), which were associated with items rather than positions, and were not restricted to immediate repetition.

Another special type of repetition memory is the repetition schema. A repetition schema represents knowledge of a recurring repetition structure in lists (Jahnke, 1969b; e.g., the particular repetition formats in the present study). In the case of repetition format *12R45R* for example, the schema might be of the form *something, something, repeat; something, something, repeat*. Though people are not normally aware of recurring patterns of repetition (Malmi & Jahnke, 1972), when the subset is very small, and repetition memory is very accurate over a number of trials (Experiments 7, 8), they may extract one or more of the underlying repetition structures. The salience of repetition at the end of groups makes it particularly likely that a schema will be extracted in such cases. The use of repetition schemata can aid both item and position recall of repeated elements (Experiments 7, 8), though they do not have to be employed by every subject on every trial to cause repetition facilitation. They must simply be used by enough subjects on enough trials to overcome the repetition inhibition arising when repetition is not detected or remembered.

**Repetition Tagging and Schemata**

Some parts of this theory are assumptions that may need further justification. For example, why assume that accurate memory for which item was repeated is only necessary to prevent repetition inhibition, and is not sufficient, without additional repetition tagging or schemata, to cause repetition facilitation? Could not the notions of repetition tagging and repetition schemata be subsumed within a single notion of repetition memory (such as Lee's tags, 1976b), with the presence of repetition facilitation or repetition inhibition depending simply on the accuracy of this memory? The main reason for thinking otherwise is that the present study failed to find repetition facilitation in some situations where repetition memory was very accurate, and yet did find repetition facilitation in other situations where there was comparable accuracy of repetition memory. With repetition format *12RR56* for example, there was repetition facilitation when ungrouped, but not when grouped (Experiments 6, 7, 8), even though repetition memory for both conditions was comparable and very accurate (over 75% of trials in Experiment 7 and Henson, 1996b).

There was also indirect evidence that repetition tagging and repetition schemata are qualitatively different from a simple memory for which item was repeated. It was only with adjacent repeated elements that did not straddle a group boundary where people sometimes erroneously repeated a response in two adjacent positions (e.g., *12RR56* recalled as *1RR256*, or *12RR56* recalled as *1255R6*; Henson, 1996b). In all other formats, such repetition errors were much further apart. The former type of error can be attributed to a repetition tag being triggered too early or too late, much like the Houghton et al. (1994) account of typing errors such as *schhol*. Such side-effects of repetition tagging might explain why adjacent repetition does not always improve item recall (Experiment 6). The latter type of error can be attributed to people forgetting which item was repeated, and making a repetition error when all other responses have been suppressed (see Discussion in Experiment 8).

A repetition schema is a qualitatively different type of repetition memory because it is only likely to be extracted when the range of repetition formats is small. This is one reason why repetition facilitation at the end of groups was stronger in Experiments 7 and 8 than in Experiment 6. It is possible that repetition memory for repetition format *12R45R* in Experiment 6 was very accurate on a trial by trial basis, but a corresponding repetition schema was never extracted because the repetition structure was so variable across trials. This would explain why few subjects in that experiment could accurately describe any of the repetition formats during debriefing, whereas most subjects in Experiments 7 and 8 were able to describe repetition format *12R45R*, normally after they had attempted the grouped condition (Henson, 1996b). (Some subjects could also describe adjacent repetition in other formats, but could not always correctly position that repetition). This was the only repetition format with nonadjacent repetition that ever led to repetition facilitation in the present experiments. Finally, use of a repetition schema for repetition format *12R45R* can also explain why serial recall of this format was least sensitive to retention interval in Experiment 7. This was in contrast with repetition format *12RR56*, which showed greater sensitivity to retention interval, presumably because of the greater opportunity for erroneous triggering of repetition tags (above).

The present theory also suggests some ways to dissociate repetition tags and repetition schemata. Repetition tagging is assumed to be a general property of the cognitive system for ordering output of sequences from both short- and long-term memory. Thus repetition tags are assumed to be employed automatically by everyone. Use of repetition schemata on the other hand depends on how well an individual can detect and remember a recurring repetition structure. Some people may extract one or more repetition schemata; others may not. Therefore, a large group of subjects could be split after testing into those who were able to describe some repetition formats and those who were not. This post hoc division should have little effect on repetition facilitation for immediate repetition; for other types of repetition, only subjects who were able to describe a repetition format accurately should show repetition facilitation for that format.

Another means of dissociating repetition tags and repetition schemata might be to employ a secondary distraction task during presentation and recall. The added attentional demands of this task should impair the abstraction of repetition schemata and reduce repetition facilitation in such cases, whereas the automatic nature of repetition tagging should mean that repetition facilitation for immediate repetition is unaffected (Baddeley & Andrade, 1996).

There are many other aspects of the above theory that warrant further investigation. The most obvious questions concern the exact interaction between repetition memory and serial recall. How exactly does memory for a repeated item prevent repetition inhibition? How do repetition tags operate, occasionally incorrectly? How do repetition schemata act during recall to improve recall of repeated items? How much do guessing strategies affect repetition effects like repetition contamination? These questions require a more precise, computational model of serial recall.

### Models of Serial Recall and Item Representation

The ability to detect the repetition of an item clearly demands type representations of items at some level of memory. Indeed, within the general theory outlined above, the process of response suppression is assumed to operate over type representations. However, these assumptions do not imply that serial order is stored over type representations. In fact, though present data indicate that repeated items in lists can impair serial recall of those lists, associative models that store order over type representations would seem to face much greater problems in recalling such lists.

Associative chaining models face problems because a repeated item will be associated with more than one successor, making it an ambiguous cue in chaining. This ambiguity can be reduced by assuming that the cue includes a number of previous responses, as in compound chaining models (Chapter 1). This additional "context" allows disambiguation of repeated elements with different predecessors. However, to the extent that cues following repeated elements retain some similarity, associative chaining models still predict that there will be a greater probability of an error following a repeated element than a control element (Chapter 2). Yet in Experiment 6, the evidence for errors, particularly associative intrusions, was weak, and may well have alternative explanations (e.g., guessing biases). In any case, given the predicted impairment following repeated items, it is unclear whether associative chaining models could match the high level of recall of repetition lists in the present study (at least without an explicit, quantitative model). More generally, associative chaining models would face increasing difficulties as the number of repeated elements in a list increases.

Associative positional models would also appear to face problems in recalling lists with repeated elements. In the Articulatory Loop Model (Burgess & Hitch, 1992, 1996b) for example, any overlap between the positional cues for repeated elements will reinforce the associations between those cues and the type representation of the repeated elements. This would appear to cause a tendency for repeated elements to be recalled too early (for reasons related to the original model's difficulty with phonologically similar items; Chapter 1).

More generally, any model that assumes that repetitions of an item produce multiple associations with the same type representation (or even increased activation of a type representation, as in *strength* models, Hintzman, 1976), would appear to have difficulty explaining Lee's (1976b) finding that the probability of recalling at least one repeated element does not differ from the probability of recalling at least one nonrepeated (control) element. Again, this problem may be more apparent than real, an issue that can be resolved by applying computational models to data from studies like the present one. The above problems do not apply to nonassociative models however, where repeated elements are stored as separate tokens (e.g., Page & Norris, 1996b). SEM is a nonassociative model.

**Item Repetition in SEM**

SEM has not been fitted to the data from Experiments 6, 7 and 8. This is mainly because many new assumptions would be needed, regarding repetition memory, repetition tagging and repetition schema, which themselves remain hypothesis to be tested further (Henson, 1996b). Nevertheless, it is worth considering how the general theory of item repetition effects outlined above might be implemented within SEM.

In SEM, auditory or visual perception of an item produces a position-sensitive token for that item in short-term memory.[4] During this process, there is no automatic registration that some items have occurred before (i.e., that some of the tokens may represent the same type). Detecting that two tokens correspond to a repeated item depends on a secondary process of comparing new tokens with older tokens in memory. The probability of detecting a repetition will therefore be a function of the similarity between tokens. This similarity is determined both by the identity of the items and their positional context (i.e., repetition detection might demand not only identical item codes, but also similar positional codes). The smaller positional overlap for items further apart in a list (Chapter 5) explains why the probability of detecting a repetition generally decreases with increasing repetition separation. In other words, a *Q* at the start of the list and the *Q* at the end of the list may appear quite different in short-term memory. When lists are grouped, the positional context for items in the same position within groups is increased, particularly for repeated elements at the end of groups, where the positional coding is very sharp (Chapter 5). At the same time, the positional codes for repeated elements at different positions within groups will reduce the chance of detecting their repetition. This is consistent with most of the results in Experiment 7.

However, the notion of positional overlap does not explain why repeated elements straddling a group boundary are well detected, because their tokens have very different positional contexts within groups. One reason why they are well detected may be the simple fact that there are no intervening items. Their repetition could then be detected by alternative means, such as residual activation of type representations. (This possibility could be tested by inserting a redundant item between groups.) Though obviously an ad hoc solution at the moment, this additional assumption allows SEM to capture fully the effects of repetition separation and grouping on repetition memory.

---

4. When repeated elements are too close in time however, as in very rapidly presented sequences, it may not be possible to form two separate tokens. Such a limit to the process of token individuation (at least for visual presentation) is one reason sometimes given for repetition blindness (e.g., Kanwisher, 1987).

Once a token is selected for output during recall (Chapter 5), it makes contact with its long-term type representation again, in order to articulate a categorical response. The suppression of these type representations corresponds to the automatic suppression underlying repetition inhibition in the general theory above. This suppression can prevent a repeated item being recalled more than once (even though there are two separate tokens for that item in memory). This may lead to a transposition of an item whose type representation is not suppressed, or even an omission, if most type representations are suppressed. Alternatively, a guess might result instead. A guess chosen from the most active (least suppressed) type representation in memory (Chapter 6) will be unlikely to produce the correct repeated item, and repetition inhibition will still result. Thus SEM already provides a suppression process that can explain repetition inhibition by both failure to retrieve a repeated item and failure to guess a repeated item.

It is less clear how memory for a repeated item interacts with serial recall in SEM. One possibility is that detection of a repeated item causes its type representation to be flagged. The purpose of the flag is to prevent suppression of the type representation during output. In principle, this means that both tokens can be selected and output as effectively as if they represented two different control elements[5]. This prevents repetition inhibition. By assuming further that the flag itself can sometimes be forgotten or separated from its type representation, there is also the possibility of repetition of a different item.

As in the general theory then, the magnitude of repetition inhibition depends on how often the repetition is detected and flagged correctly. Because detection and flagging is normally quite good, repeated elements can be recalled correctly more often than not, and repetition inhibition is usually only of the order of 10%. When detection and flagging are very accurate, then repetition inhibition can be prevented completely.

Repetition tagging can be modelled as a special type of token in memory. When this “doubling” token is selected, it causes immediate repetition of the next token selected, before the corresponding type representation is suppressed. This gives adjacent repeated elements an

5. (though the exact consequences of withholding suppression depend on the equation governing the strength with which categorical representations compete for output; Equation 10-8 in Appendix 3)

advantage over adjacent control elements, producing repetition facilitation. However, this is not always the case, because if the doubling token is selected too early or too late, there can be adjacent repetition on the wrong position or of the wrong item. Modelling tags as tokens, rather than over phonological representations, may explain why effects of immediate repetition do not interact with phonological similarity (Drewnowski, 1980a).

Repetition schemata might be modelled as a structure associated with particular positional codes. When the identity of tokens with these positional codes match, this structure is triggered, and associated with the repeated item. This association suppresses the type representation of the repeated item until the corresponding positions are reached in recall. At this point, the representations are unsuppressed and the repeated item is output. This will increase recall of repeated elements at these positions above corresponding control elements, producing repetition facilitation.

**Repetition Contamination**

Finally, repetition contamination is assumed to have two aspects. The first is that the probability of recalling context elements in a repetition list depends on the probability of recalling the repeated elements. This means context elements are recalled better under conditions of repetition facilitation than repetition inhibition, and is a trivial consequence of competition and suppression in SEM (Chapter 5). The second aspect of repetition contamination is a slight increase in the probability of guessing context elements in repetition lists than control lists. This owes to the smaller set of items to guess from in repetition lists than control lists. This guessing bias explains why a slightly greater proportion of errors following correct recall of one critical element will be the item following the other critical element in repetition lists than control lists. Thus Wickelgren's (1966) associative intrusions are not seen as evidence for an associative chaining theory, but as an artifact of guessing. Being a nonassociative model, SEM also predicts that, provided a repeated element is recalled correctly, the fact that it represents a repeated item has no relevance to recall of the subsequent context element (i.e. no effect of repetition on cuing). The question of whether repeated items do have any effect on recall of subsequent items must await further testing, in which performance on repeated and control elements is equated.

Many of the above ideas are speculative, and must await implementation and simulation in SEM before they can be confirmed theoretically (and subsequent experiments before they can be supported empirically). Nevertheless, they represent a first approach to modelling a tight set of constraints emerging from the present experiments; constraints that also represent a challenge for other models of serial recall, particularly associative ones.

## Chapter Summary

The results from present experiments suggest that repetition inhibition arises because people often fail to retrieve, or guess, the second occurrence of a repeated item, unless they explicitly detect and remember repetition of that item. Detection and memory of a repetition is more likely the more immediate or the more salient the repetition (e.g., at the end of groups). In such cases, additional tagging of immediate repetition or abstraction of repetition schemata can increase the probability of recalling both occurrences and produce repetition facilitation.

These complex yet robust findings represent important challenges for models of serial recall from short-term memory. Those that assume order is represented over token representations would seem best suited to explaining people's general ability to recall sequences with repeated elements. This is consistent with one of the core assumptions of SEM (Chapter 5). The specific effects that item repetition has on serial recall, at least in short-term memory, may well arise from special mechanisms geared towards the detection of repetition (such as repetition tagging or repetition schema), or the output of a response (such as response suppression), though the implementation of such mechanisms in SEM remains a task for the future. Indeed, the issue of item representation in both short- and long-term memory remains an open one, whose resolution may well depend on demonstrations that particular models with particular representations can account for the empirical data, such as those in this chapter.

# Chapter 8: Conclusions

## *A Solution to the Problem of Serial Order?*

How then do we store and retrieve a sequence of items in the correct order? The argument in this thesis has been that, for short-term memory at least, order is stored by associating each item with the start and end of the sequence. The relative strengths of these associations provide an approximate code for the item's position in the sequence. This code is stored together with the item to form a position-sensitive token. The order is retrieved by reinstating the positional codes and using them to cue tokens in memory. The evidence for this argument is summarised below.

### Summary of Thesis

Chapter 1 introduced three possible solutions to the problem of serial order: chaining theory, positional theory, and ordinal theory. Experiment 1 in Chapter 2 failed to find any evidence for chaining theory in immediate serial recall, whereas Experiments 2 and 3 in Chapter 3 found evidence supporting positional theory. This evidence took the form of positional errors, either as transpositions between groups that maintain their position within groups (Experiment 2) or intrusions between trials that maintain their position within a trial (Experiment 3). These errors cannot be explained by ordinal theory.

Chapter 4 examined three more specific models of serial recall and used meta-analyses to argue that none was sufficient to capture the complete pattern of errors in short-term, serial recall. A new positional model was developed in Chapter 5 (the Start-End Model, SEM) that could reproduce the complete pattern of errors. This model was based on positional theory, and demonstrated how items can be ordered by cuing with approximate positional codes.

Chapter 6 examined the nature of those positional codes in more detail: in particular, whether the codes represent absolute position, or position relative to the start and end of a sequence. Experiments 4 and 5 resembled Experiments 2 and 3, except that they used groups and lists of different lengths. The positional errors in these cases supported the notion of relative position, in agreement with the predictions of SEM.

Finally, Chapter 7 examined the effects of repeated items in sequences, which pose problems for most models of serial recall. The robust yet complex nature of these effects suggested several additional factors contribute to recall of repeated items. Nonetheless, the effects appeared best explained by SEM: in particular, with its assumption that items are stored as position-sensitive tokens.

### The Start-End Model

It is important to distinguish theory and model in this thesis. The theory that memory for serial order utilises positional information, where that information is defined relative to the start and end of a sequence, is based on experimental results from tasks such as serial recall. The model, SEM, is a more specific implementation of this theory, which makes further assumptions about short-term memory and the serial recall process, in order to fit data quantitatively. The success of SEM in fitting present data supports the more general positional theory. Nonetheless, the validity of the theory does not depend on the success of the model; SEM may be refuted by future data without necessarily refuting the theory.

SEM was reviewed in Chapter 5, where it was discussed in relation to further aspects of short-term memory, and compared with other models. However, it is useful to step away from the details of experiments and models, and consider more general implications of a positional theory of memory for serial order.

## Serial Order in Short-term Memory

The problem of serial order in the present thesis has been confined to short-term memory for a novel sequence of items. The example given in Chapter 1 was of holding an unfamiliar telephone number in memory long enough to dial it. The present solution, in terms of storing each digit with a positional code defined by start and end markers, and reinstating these codes during dialling, may not seem particularly intuitive. Several questions might be asked of such a solution. For example, is not generating and reinstating positional codes for each digit in the telephone number somewhat laborious? What are the start and end markers that define these codes? How is an end marker used if the length of the telephone number is unknown? How is the order of positional codes themselves reinstated?

The answers to these questions are interrelated. Firstly, generating and reinstating positional codes is probably an automatic, unconscious process. One does not necessarily have to think “a 5 was near the start of the telephone number and a 9 was near the end”. In other words, the start and end markers may represent intrinsic properties of short-term memory that are not open to introspection. This also makes their interpretation difficult (which is why their psychological definition has remained vague in the present thesis). In fact, it is unclear whether much elaboration of start and end markers is possible in psychological terms. For example, there may be one group of neurons in the brain whose activity is triggered by the first digit in the telephone number and decreases with each subsequent digit, and another group of neurons whose activity increases with each subsequent digit. What is the psychological interpretation of these neurons? The answer can only be simply that the activity of the first group represents proximity to the start of the telephone number and the activity of the second group represents proximity to the end of the telephone number. This is equivalent to saying the groups of neurons are start and end markers.[1]

The question of how an end marker can grow towards the end a sequence, when the end is not known in advance, was addressed in Chapter 6, with one answer requiring only an approximate level of expectation for the end of the sequence. With respect to the present example however, two points are worth noting. Firstly, an end marker may not be employed. The order of digits in the telephone number may be stored with reference to a start marker only (similar perhaps to the start-of-list context proposed by Page & Norris, 1996b). SEM is still able to store order with a single marker. Secondly, one may split the telephone number into groups of a predetermined size (i.e., the number may be grouped subjectively). In this case, an end marker can be employed to help define positions within a group, because the end of a group is known in advance. Alternatively, coding of position relative to the end of a sequence might be achieved during cumulative rehearsal of the digits (Chapter 5). Thus prior knowledge of the length of a telephone number, though helpful, is not necessary, particularly if the number is grouped subjectively or rehearsed cumulatively.

1. Note that few other positional theories have adequate psychological interpretation of the positional information they assume (e.g., Lee & Estes, 1981; Nairne, 1991), except perhaps those that attribute this information to temporal oscillators (e.g., Burgess & Hitch, 1996b), which have problems explaining the results in Chapter 6.

Finally, there is the concern that a positional theory simply displaces the problem of ordering the digits of the telephone number to the problem of ordering the positional codes associated with each digit. However, the problem of ordering the positional codes can have a simpler solution than the problem of ordering the digits. For example, in the Houghton (1990) model, all that need be done is activate the start marker. This activation then decays in a fixed manner, and the activation of the end marker grows in a yoked manner. Positional codes are therefore ordered automatically, as a consequence of the simple dynamics of the start marker. Though the data in Chapter 6 argue against a yoked end marker, positional codes in SEM are also an automatic consequence of two simple functions, one decreasing from the start of a sequence and one increasing towards the end. Moreover, these functions can be a fixed property of short-term memory: Though the order of digits must be learned for each novel telephone number, the order of positional codes is invariant (analogous perhaps to a pegword mnemonic). Ordering positional codes is less of a problem than it might first appear.

In summary, the present account of our ability to remember a novel telephone number for a short period of time does not necessarily suffer from conflicting intuitions. Indeed, it is defensible with respect to data from short-term memory for similar sequences in the present experiments. The next question is whether this solution to the problem of serial order in short-term memory generalises to the problem of serial order in other aspects of memory.

## Serial Order in Long-term Memory

The problem of serial order is fundamental to most aspects of memory. Chapter 1 gave two examples of serial order in procedural memory (ordering phonemes in speech) and in episodic memory (ordering life-events). Is positional theory appropriate for these examples?

### Procedural Memory

Though positional theory is supported by data from short-term memory, it faces problems when applied to long-term, procedural memory. These problems were mentioned in Chapters 1 and 5, with the main problem being the interference problem: People can store and retrieve numerous sequences of the same basic elements, with little to no interference between such sequences. If those sequences were coded by a single start and end marker, much greater interference would be expected. In other words, the interference between groups in STM is rarely found between chunks or motor programs in procedural memory (the possible exception of speech errors is discussed below).

One solution to the interference problem is that each sequence has its own, unique start and end marker (Houghton, 1990). However, for procedural memory at least, a simpler solution may be an ordinal one. Indeed, several ordinal accounts have been suggested along neurally-inspired dimensions (e.g., Grossberg, 1978; Nigrin, 1993; Page, 1994). For example, order might be stored in a primacy gradient of synaptic strengths between a neuron representing the sequence and others representing the sequence elements, a hypothesis which has some neurophysiological support (Granger, Whitson, Larson & Lynch, 1994). Because the synaptic strengths associated with one “sequence neuron” are independent of those associated with another, as many sequences can be stored, in principle, as there are free neurons. Order can be retrieved by activating a sequence neuron, which primes the “element neurons” in proportion to their associated synaptic strength, and activating each element neuron in a cyclic process of selection and suppression of the most primed neuron. Models based on these ideas have demonstrated long-term learning of temporal sequences that does not suffer from the interference problem (Nigrin, 1993; Page, 1994).

In addition to the interference problem, a further problem concerns memory for long sequences. Even ordinal models face problems with long sequences, given that most biological dimensions are finite (and noisy). Long sequences are best stored as a hierarchy of subsequences, with the hierarchical decomposition continuing until the smallest subsequences are within short-term memory span. This is consistent with experimental evidence. For example, Klahr, Chase and Lovelace (1983) found that latency profiles in probed recall of the alphabet showed marked discontinuities across certain letter pairs. These discontinuities suggested a chunk boundary.[2] For Americans, these boundaries were coincident with the phrase boundaries of the “alphabet song”, which is often used to teach the alphabet. However, though it is well known that the presence of such structure aids serial learning (e.g., Martin, 1974), there is a surprising dearth of adequate psychological models for such data.

2. That the same data can be fitted by a linear model with variable interitem strengths (Scharroo, 1994) is not surprising; what is required is a demonstration of how such a chaining model can solve the interference problem.

Models of speech production generally assume separate mechanisms for dealing with the order and content of an utterance (e.g., MacKay, 1982). The order is sometimes phrased in terms of syntax or schemata (Lashley, 1951). For example, some models assume a syllable schema (e.g., Dell, 1988), in order to explain speech errors such as spoonerisms, the transpositions of phonemes between words (e.g., “dear old queen” spoken as “*q*ueer old *d*ean”). Though resembling positional errors in STM, these errors respect more specific syllabic constraints (Hartley & Houghton, 1996). Other models assume that spoonerisms are failures of an editing process that operates over a speech output buffer, rather than indicating syllabic coding in long-term memory per se (Levelt, 1989). In either case, speech errors clearly entail more than simply positional information and, as such, do not really constitute evidence for positional coding in long-term memory.

The assumption of schemata to order the content of an utterance does not directly address the problem of serial order however. The question remains as to how that order is represented. MacKay (1982) assumed a hierarchy of timing nodes, consistent with evidence for a binary hierarchy in speech production (Gordon & Meyer, 1987). However, the nature of such timing nodes is rarely specified. They may represent internal oscillators, as assumed for the rhythm found in the stress patterns of speech (Robinson, 1977). Such oscillators have also been assumed to store order in short-term memory (e.g., Brown et al., 1996; Burgess & Hitch, 1996a, 1996b), though Chapter 6 argued against these models. In general, more evidence is required to determine the relationship between serial order in procedural memory (and speech production in particular), and serial order in short-term memory.

**Episodic Memory**

Another example in Chapter 1 concerned the ordering of events in the past. This is an example of episodic memory, in which temporal order is fundamental (Tulving, 1983). Short-term memory can also be viewed as an a form of episodic memory. As Nairne noted: “When items are forgotten from memory lists, it is not the items themselves that are forgotten, but rather their occurrences in prior spatiotemporal windows.” (Nairne, 1991, p. 332).

In the case of autobiographical memory, considerable research has focused how people date past events. Friedman (1993) distinguished three main theories. His distance theory coded time along a single dimension (e.g., strength of memory), resembling ordinal theory; his location theory coded time by relating it to temporal schemata (e.g., days of the week), resembling positional theory; and his relative theory coded the order of two events (e.g., that “X occurred before Y”), resembling chaining theory. In reviewing the evidence, he argued for a location theory where events are dated via temporal schemata, occasionally supplemented by memory for relative order (and in rare cases, knowledge of exact dates). A single continuum, along which events recede hazily into the past, may be more illusory than real.

The main evidence for temporal schemata comes from scale effects, where memory for fine temporal detail is superior to that for coarse temporal detail. For example, an event attributed to the wrong week can still be attributed to the correct day, and such memory is not explicable simply by guesses based on general knowledge (Friedman & Wilkins, 1985). These multiple scales (e.g., day-in-week, week-in-month) resemble those seen in short-term memory (e.g., positions of item-in-group, group-in-list; Chapter 5). Unlike short-term memory however, the dating of past events is more likely to involve indirect, reconstructive inferences (e.g., “that was on wash-day, Monday”; Larsen & Thompson, 1995), rather than direct positional codes. Nevertheless, in the case of generalisations such as “towards the start of the week” or “towards the end of the week” there may be an additional role for the type of start and end markers assumed here for short-term memory. These markers would allow judgements of the order of events in the same week whose exact days could not be inferred.

In laboratory tests of episodic memory, positional theory is clearly supported. People are not only able to judge the frequency (Hintzman, 1976), recency (Yntema & Trask, 1963) and duration (Block, 1982) of events, but also their position within temporal sequences. Toglia and Kimble (1976), for example, showed that approximate positional judgements were above chance even under incidental learning of long lists of 96 words. Such judgements not only show primacy and recency effects, but also positional errors between lists (Hintzman, Block & Summers, 1973). Even positions of repeated items can be judged accurately, and often independently (Hintzman & Block, 1971), supporting multiple representations of repeated items (as in SEM). Given that covert serial recall of such long lists is highly improbable, these data constitute strong support for positional information in episodic memory.

Hintzman, Block and Summers (1973) attributed positional judgements to contextual associations, distinguishing reinstateable and nonreinstateable contexts, as in SEM (Chapter 5). Block (1982) showed that judgements of relative duration of lists were affected by both intrinsic and extrinsic contextual change between lists, but judgements of position were not. These results can be explained if duration judgements are based on changes in nonreinstateable context (e.g., SEM’s general context), whereas position judgements are based on reinstateable context (e.g., SEM’s positional codes).

Some evidence suggests that knowledge of relative order may also aid episodic judgements. Tzeng, Lee and Wetzel (1979) used a “study-phase retrieval model” to explain why displaced rehearsals do not disrupt judgements of temporal order. Displaced rehearsals are rehearsals of items that are retrieved during the study of later items, and tend to improve judgements of their relative order. This is problematic for positional models, in which displaced rehearsals will, if anything, recode the retrieved and studied items in adjacent positions, and hence impair discrimination of their relative order. In the study-phase retrieval model, displaced rehearsals are used to encode the relative order of the two items. This explains why judgements of relative order in categorised lists are better for intracategory than intercategory judgements, even though the latter are better separated in time and position (Tzeng & Cotton, 1980). However, that the relative order of two items can be coded explicitly is unequivocal; how such codes could underlie all positional judgements is equivocal. Numerous codes for relative order would be necessary to judge positions accurately, and they could not explain positional errors (Hintzman, Block & Summers, 1973). Coding relative order during study-phase retrievals would appear optional rather than necessary or sufficient.

In summary, serial order in episodic memory does utilise positional codes. With the lists of random words employed in most laboratory tasks, the start and end of a sequence provide the only salient means with which to define position. Together with the notion of nonpositional general context, a theory based on start and end markers, like SEM, can explain the basic findings in episodic judgement tasks, including serial position effects and positional errors.[3] With more meaningful events in autobiographical memory, temporal order is likely to

3. Indeed, an alternative reading of SEM is “Search of Episodic Memory” (cf. Raaijmakers & Shiffrin, 1981).

be supplemented by inferences based on temporal schemata, with occasional help from general knowledge and the explicit coding of relative order.

**Other Memory**

There are other examples of serial order that are not easily classified as procedural or episodic memory. For example, memory for the order of British Monarchs is perhaps better classed as semantic memory (assuming the order has not been learned by rhyme, as in the alphabet example above). This order is more likely to be reconstructed by a series of propositional facts (e.g., “William and Mary must have followed James II because they replaced his Catholic rule”); inferences that do not fall naturally into any of the chaining, positional or ordinal theories in Chapter 1. Indeed, it is arguable whether such reconstruction counts as true memory for serial order.

Another distinction often made is between implicit and explicit memory. Though not necessarily distinct memory systems (Schacter & Tulving, 1994), these memories differ in their access. Explicit memory refers to conscious recollection of past episodes, whereas implicit memory refers to nonconscious use of previously acquired information. These notions are distinct from procedural and episodic memory, though obviously related. There is some evidence to suggest that positional information is used only in explicit memory. For example, serial position effects arise in explicit but not implicit tests (Brooks, 1994) and in “remember” but not “know” responses (Jones & Roediger, 1995).

Positional information may also require explicit encoding (which is not necessarily precluded by the “incidental” learning conditions of the studies mentioned in Chapter 6). Nevertheless, it may still be possible to encode order information implicitly, through nonpositional means. This would be necessary to account for implicit learning of temporal sequences (e.g., Stadler, 1993), though this is a contentious issue (Shanks & St. John, 1994), and unlikely to apply to the Hebb effect (Chapter 5; Sechler & Watkins, 1991). In either case, the hypothesis that explicit encoding and retrieval is necessary to utilise positional information, but not order information, appears an interesting and testable hypothesis.

A complete review of other types of memory is beyond present concerns. The purpose of the above discussion is to suggest more than one solution to the problem of serial order in memory. This suggestion is unusual, but not unprecedented (Ebenholtz, 1972; Frensch, 1994). One solution is clearly positional, and this applies not only to the short-term memory considered here, but episodic memory more generally. It may be contingent on explicit encoding and retrieval processes. Another solution is probably ordinal, providing a simpler means of storing and retrieving temporal order in procedural memory, with the potential to be acquired and expressed implicitly. Though these ideas remain speculative, the fundamental nature of serial order clearly warrants their further investigation.

## Chapter Summary

The present thesis has demonstrated that positional information is utilised in short-term memory for serial order. More generally, such positional information appears common to episodic (explicit) memory. It may also underlie serial order in procedural or implicit memory, though other representations of order, such as simpler ordinal representations, should not be discounted. In either case, this chapter has confirmed the claims of Chapter 1, that addressing the problem of serial order in short-term memory provides a good starting point with which to address the problem of serial order more generally.

# References


Aaronson, D. (1968). Temporal course of perception in an immediate recall task. *Journal of Experimental Psychology, 76,* 129-140.

Anders, T. R., & Lillyquist, T. D. (1971). Retrieval time in forward and backward recall. *Psychonomic Science*, *22*, 205-206.

Arbuthnott, K. D. (in press). To repeat or not to repeat: repetition facilitation and inhibition in sequential retrieval. *Journal of Experimental Psychology: General.*

Armstrong, I. T., & Mewhort, D. J. K. (1995). Repetition deficit in rapid-serial-visual-presentation displays: encoding failure or retrieval failure? *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1044-1052.

Avons, S. E., Wright, K. L., & Palmer, K. (1994). The word-length effect in probed and serial recall. *The Quarterly Journal of Experimental Psychology, 47A,* 207-231.

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *The Quarterly Journal of Experimental Psychology, 20,* 249-263.

Baddeley, A. D. (1986). *Working Memory.* Oxford: Clarendon Press.

Baddeley, A. D., & Andrade, J. (1996). [An experiment examining central-executive load on item repetition effects and recall order.] Unpublished data.

Baddeley, A. D., Conrad, R., & Hull, J. (1965). Predictability and immediate memory for consonant sequences. *The Quarterly Journal of Experimental Psychology, 17,* 175-177.

Baddeley, A. D., & Ecob, J. R. (1970). Simultaneous acoustic and semantic coding in short-term memory. *Nature, 227*, 288-289.

Baddeley, A. D., Emslie, H., Kolodny, J., & Duncan, J. (1995). *Random generation and the executive control of working memory.* Manuscript in preparation.

Baddeley, A. D., & Hitch, G. (1993). The recency effect: implicit learning with explicit retrieval. *Memory & Cognition, 21*, 146-155.

Baddeley, A. D., & Hull, A. (1979). Prefix and suffix effects: do they have a common basis? *Journal of Verbal Learning and Verbal Behavior, 18*, 129-140.

Baddeley, A. D., & Lewis, V. (1984). When does rapid presentation enhance digit span? *Bulletin of the Psychonomic Society, 22,* 5.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology, 36A*, 233-252.

Baddeley, A. D., Papagno, C., & Norris, D. G. (1991). A sandwich for TODAM. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating Theory and Data: in Honor of Bennet B. Murdock* (pp. 175-194). Hillside, NJ: LEA.

Baddeley, A. D., & Scott, D. (1971). Short term forgetting in the absence of proactive interference. *The Quarterly Journal of Experimental Psychology, 23*, 275-283.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575-589.

Banks, W. P., White, H., & Mermelstein, R. (1980). Position effects in comparative judgements of serial order: list structure vs. differential strength. *Memory & Cognition*, *8*, 623-630.

Battig, W. F., Brown, S. P., & Schild, M. E. (1964). Serial position and sequential associations in serial learning. *Journal of Experimental Psychology, 67*, 449-457.

Bavelier, D. (1994). Repetition blindness between visually different items: the case of pictures and words. *Cognition, 51*, 199-236.

Bishop, D. V. M., & Robson, J. (1989). Unimpaired short-term memory and rhyme judgement in congenitally speechless individuals: implications for the notion of “articulatory coding”. *The Quarterly Journal of Experimental Psychology, 41A*, 123-140.

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior, 13,* 80-97.

Block, R. A. (1982). Temporal judgements and contextual change. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8*, 530-544.

Bower, G. H., & Springston, F. (1970). Pauses as recoding points in letter strings. *Journal of Experimental Psychology, 83*, 421-430.

Bower, G. H., & Winzenz, D. (1969). Group structure, coding, and memory for digit sequences. *Journal of Experimental Psychology Monograph Supplement, 80*, 1-17.

Broadbent, D. E. (1981). Association lecture: from the percept to the cognitive structure. In J. Long & A. Baddeley (Eds.), *Attention and Performance IX* (pp. 1-25). Hillside, NJ: LEA.

Broadbent, D. E., Cooper, P. J., & Broadbent, M. H. P. (1978). A comparison of hierarchical and matrix retrieval schemes in recall. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 486-497.

Broadbent, D. E., Cooper, P. J., Frankish, C. R., & Broadbent, M. H. P. (1980). Modality differences in relation to grouping in immediate recall. *British Journal of Psychology, 71*, 475-485.

Brooks, B. M. (1994). A comparison of serial position effects in implicit and explicit word-stem completion. *Psychonomic Bulletin & Review, 1*, 264-268.

Brooks, B. M., & Watkins, M. J. (1990). Further evidence for the intricacy of memory span. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 1134-1141.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *The Quarterly Journal of Experimental Psychology, 10*, 12-21.

Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: no rehearsal needed? *Journal of Memory and Language, 34*, 594-621.

Brown, G. D. A., Preece, T., & Hulme, C. (1996). Oscillator-based memory for serial order. Manuscript submitted for publication.

Bruce, D. (1994). Lashley and the problem of serial order. *American Psychologist, 49*, 93-103.

Brugger, P., Monsch, A. U., & Johnson, S. A. (1996). Repetitive behavior and repetition avoidance: the role of the right hemisphere. *Journal of Psychiatry & Neuroscience, 21*, 53-56.

Bunt, A. A. (1976). Context effects in short-term memory: confirmatory evidence from recall of visually presented lists. *Acta Psychologica, 40*, 423-430.

Burgess, N., & Hitch, G. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31*, 429-460.

Burgess, N., & Hitch, G. (1996a). A connectionist model of STM for serial order. In S. Gathercole (Ed.) *Models of short-term memory* (pp. 51-71). UK: Psychology Press.

Burgess, N., & Hitch, G. (1996b). Memory for serial order: a network model of the phonological loop and its timing. Manuscript submitted for publication.

Clifton, C., & Tash, J. (1973). Effect of syllabic word length on memory-search rate. *Journal* *of Experimental Psychology, 99*, 231-235.

Conrad, R. (1959). Errors of immediate memory. *British Journal of Psychology, 50,* 349-359.

Conrad, R. (1960). Serial order intrusions in immediate memory. *British Journal of Psychology, 51*, 45-48.

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior, 4*, 161-169.

Conrad, R. (1967). Interference or decay over short-retention intervals? *Journal of Verbal Learning and Verbal Behavior, 6,* 49-54.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 429-432.

Conrad, R., & Hull, A. J. (1966). The role of the interpolated task in short-term retention. *The Quarterly Journal of Experimental Psychology, 18,* 266-269.

Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science, 10*, 135-136.

Corballis, M. (1967). Serial order in recognition and recall. *Journal of Experimental Psychology, 74*, 99-105.

Cowan, N. (1993). Activation, attention and short-term memory. *Memory & Cognition, 21*, 162-167.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language, 31*, 1-17.

Cowan, N., Wood, N. L., & Borne, D. N. (1994). Reconfirmation of the short-term storage concept. *Psychological Science, 5*, 103-106.

Cox, D. R. & Snell, E. J. (1989). *Analysis of Binary Data* (2nd ed.). London: Chapman & Hall.

Crannell, C. W., & Parrish, J. M. (1957). A comparison of immediate memory span for digits, letters and words. *The Journal of Psychology, 44*, 319-327.

Crowder, R. G. (1968a). Intraserial repetition effects in immediate memory. *Journal of Verbal Learning and Verbal Behavior, 7,* 446-451.

Crowder, R. G. (1968b). Repetition effects in immediate memory when there are no repeated elements in the stimuli. *Journal of Experimental Psychology, 78*, 605-609.

Crowder, R. G. (1969). Behavioral strategies in immediate memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 524-528.

Crowder, R. G. (1978). Memory for phonologically uniform lists. *Journal of Verbal Learning and Verbal Behavior, 17*, 73-89.

Crowder, R. G. (1979). Similarity and order in memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation, 13* (pp. 319-353). New York: Academic Press.

Crowder, R. G. (1993). Short-term memory: where do we stand? *Memory & Cognition, 21*, 142-145.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage. *Perception & Psychophysics, 5*, 365-373.

Cunningham, T. F., Healy, A. F., Till, R. E., Fendrich, D. W., & Dimitry, C. Z. (1993). Is there really very rapid forgetting from primary memory? The role of expectancy and item importance in short-term recall. *Memory & Cognition, 21, 671-688.*

Cunningham, T. F., Healy, A. F., & Williams, D. M. (1984). Effects of repetition on short-term retention of order information. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 575-597.

Dell, G. S. (1984). Representation of serial order in speech: evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 222-233.

Dell, G. S. (1988). The retrieval of phonological forms in production: tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*, 124-142.

Dillon, R. F., & Thomas, H. (1975). The role of response confusion in proactive interference. *Journal of Verbal Learning and Verbal Behavior, 14*, 603-615.

Drewnowski, A. (1980a). Attributes and priorities in short-term recall: a new model of memory span. *Journal of Experimental Psychology: General*, *109*, 208-250.

Drewnowski, A. (1980b). Memory functions for vowels and consonants: a reinterpretation of acoustic similarity effects. *Journal of Verbal Learning and Verbal Behavior, 19*, 176-193.

Drewnowski, A., & Murdock, B. B. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 319-332.

Ebbinghaus, H. (1964). *Memory: a contribution to experimental psychology.* New York: Dover.

Ebenholtz, S. M. (1963). Serial learning: position learning and sequential associations. *Journal of Experimental Psychology, 66,* 353-362.

Ebenholtz, S. M. (1972). Serial learning and dimensional organisation. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation, 5* (pp. 267-314). New York: Academic Press.

Ellis, A. (1980). Errors in speech and short-term memory: the effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior, 19,* 624-634.

Ellis, N. C., & Hennelly, R. A. (1980). A biligual word-length effect: implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology, 71*, 43-51.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Estes, W. K. (1972). An associative basis for coding and organisation in memory. In A. W. Melton & E. Martin (Eds.), *Coding Processes in Human Memory* (pp. 161-190). Washington DC: V. H. Winston & Sons.

Estes, W. K. (1973). Phonemic coding and rehearsal in short-term memory for letter strings. *Journal of Verbal Learning and Verbal Behavior, 12*, 360-372.

Estes, W. K. (1991). On types of item coding and sources of recall in short-term memory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating Theory and Data: in Honor of Bennet B. Murdock* (pp. 175-194). Hillside, NJ: LEA.

Fagot, C., & Pashler, H. (1995). Repetition blindness: perception or memory failure? *Journal of Experimental Psychology: Human Perception and Performance, 21*, 275-292.

Farrand, P., & Jones, D. (1996). Direction of report in spatial and verbal serial short-term memory. *The Quarterly Journal of Experimental Psychology, 49A*, 140-158.

Ferguson, R. P., & Bray, N. W. (1976). Component processes of an overt rehearsal strategy in young children. *Journal of Experimental Child Psychology, 21*, 490-506.

Frankish, C. R. (1974). *Organisational factors in short-term memory.* Unpublished doctoral dissertation, University of Cambridge.

Frankish, C. R. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language, 24,* 200-209.

Frankish, C. R. (1989). Perceptual organisation and precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 469-479.

Frankish, C. R. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology, 9,* 5-22.

Frankish, C. R. (1996). Auditory short-term memory and the perception of speech. In S. E. Gathercole (Ed.), *Models of Short-Term Memory* (pp. 101-128). UK: Psychology Press.

Frankish, C. R., & Turner, J. (1984). Delayed suffix effect at very short delays. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10,* 767-777.

Frensch, P. A. (1994). Composition during serial learning: a serial position effect. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 423-442.

Frick, R. W. (1988a). Issues of representation and limited capacity in the auditory short-term store. *British Journal of Psychology, 79,* 213-240.

Frick, R. W. (1988b). The role of memory in attenuations of the suffix effect. *Memory & Cognition, 16*, 15-22.

Frick, R. W. (1989). Explanations of grouping in immediate ordered recall. *Memory & Cognition, 17,* 551-562.

Frick, R. W. (1995). Accepting the null hypothesis. *Memory & Cognition, 23*, 132-138.

Frick, R. W., & De Rose, A. (1986). The suffix effect and preattentive unit-formation in visual short-term memory. *Canadian Journal of Psychology, 40,* 97-108.

Friedman, W. (1993). Memory for the time of past events. *Psychological Review, 113,* 44-66.

Friedman, W. J., & Wilkins, A. J. (1985). Scale effects in memory for the time of past events. *Memory & Cognition, 13*, 168-175.

Fuchs, A. H. (1969). Recall for order and content of serial word lists in short-term memory. *Journal of Experimental Psychology, 82,* 14-21.

Gathercole, S. E., Gardiner, J. M., & Gregg, V. H. (1982). Modality and phonological effects in serial recall: does ones own voice play a role? *Memory & Cognition, 10,* 176-180.

Gathercole, S. E., & Martin, A. J. (1996). Interactive processes in phonological memory. In S. E. Gathercole (Ed.), *Models of Short-Term Memory* (pp. 73-100). UK: Psychology Press.

Glenberg, A. M. (1990). Common processes underlie enhanced recency effects for auditory and changing state stimuli. *Memory & Cognition, 18*, 638-650.

Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*, 3-15.

Gordon, P. C., & Meyer, D. E. (1987). Control of serial order in rapidly spoken syllable sequences. *Journal of Memory and Language, 26*, 300-321.

Granger, R., Whitson, J., Larson, J., & Lynch, G. (1994). Non-Hebbian properties of long-term potentiation enable high capacity encoding of temporal sequences. *Proceedings of the National Academy of Sciences, USA, 91*, 10104-10108.

Greene, R. L. (1986). Sources of recency effects in free recall. *Psychological Bulletin, 99*, 221-228.

Greene, R. L. (1989). Immediate serial recall of mixed-modality lists. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 266-274.

Greene, R. L. (1991). The Ranschburg effect: the role of guessing strategies. *Memory & Cognition, 19*, 313-317.

Greene, R. L., & Crowder, R. G. (1988). Memory for serial position: effects of spacing, vocalisation and stimulus suffixes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 740-748.

Grossberg, S. (1978). Behavioral contrast in short-term memory: serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology, 17*, 199-219.

Hanley, J. R., & Broadbent, C. (1987). The effect of unattended speech on serial recall following auditory presentation. *British Journal of Psychology, 78*, 287-297.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35*, 1-31.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: a review and a new view. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation, 22* (pp. 193-225). New York: Academic Press.

Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 644-655.

Healy, A. F. (1977). Pattern coding of spatial order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 16,* 419-437.

Healy, A. F. (1982). Short-term memory for order information. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation, 16* (pp. 191-238). New York: Academic Press.

Healy, A. F., Cunningham, T. F., Gesi, A. T., Till, R. E., & Bourne, L. E. (1991). Comparing short-term recall of item, temporal and spatial information in children and adults. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating Theory and Data: in Honor of Bennet B. Murdock* (pp. 175-194). Hillside, NJ: LEA.

Healy, A. F., Fendrich, D. W., Cunningham, T. F., & Till, R. E. (1987). Effects of cuing on short-term retention of order information. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13,* 413-425.

Healy, A. F., & McNamara, D. S. (1996). Verbal learning and memory: does the modal model still work? *Annual Review of Psychology, 47,* 143-172.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning* (pp. 37-46). London: Oxford University Press.

Hendrikx, A. J. P. (1984). Effects of grouping on recall latency. *Acta Psychologica, 55*, 1-18.

Henson, R. N. A. (1994). *Compound chaining models of memory for order.* Unpublished manuscript.

Henson, R. N. A. (1995). [Data on backward recall and grouping]. Unpublished raw data.

Henson, R. N. A. (1996a). *Irrelevant tones and temporal grouping*. Unpublished manuscript.

Henson, R. N. A. (1996b). *Item repetition in short-term, serial recall: Ranschburg repeated.* Manuscript submitted for publication.

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology, 49A*, 80-115.

Hinrichs, J. V., & Mewaldt, S. P. (1977). The Ranschburg effect: modification of guessing strategies by context. *Bulletin of the Psychonomic Society, 9,* 85-88.

Hinrichs, J. V., Mewaldt, S. P., & Redding, J. (1973). The Ranschburg effect: repetition and guessing factors in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 12*, 64-75.

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation, 10* (pp. 47-91). New York: Academic Press.

Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating Theory and Data: in Honor of Bennet B. Murdock* (pp. 39-56). Hillside, NJ: LEA.

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology, 88*, 297-306.

Hintzman, D. L., Block, R. A., & Summers, J. L. (1973). Contextual associations and memory for serial position. *Journal of Experimental Psychology, 97*, 220-229.

Hitch, G. J. (1972). Organisation and retrieval in immediate memory. Unpublished Doctoral Dissertation, University of Cambridge.

Hitch, G. J. (1974). Short-term memory for spatial and temporal information. *Quarterly Journal of Experimental Psychology, 26*, 503-513.

Hitch, G. J., Burgess, N., Shapiro, J., Culpin, V., & Malloch, M. (1995). *Evidence for a timing signal in verbal short-term memory.* Paper presented at the Experimental Psychology Society, Birmingham Meeting, July, 1995.

Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: a working memory analysis. *The Quarterly Journal of Experimental Psychology, 49A*, 116-139.

Hitch, G. J., & Morton, J. (1975). The unimportance of explicit spatial information in serial recall of visually presented lists. *The Quarterly Journal of Experimental Psychology, 27,* 161-164.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish & M. Zock, (Eds.), *Current Research in*

*Natural Language Generation* (pp. 287-319). London: Academic Press.

Houghton, G. (1994). Some formal variations on the theme of competitive queueing. (Technical Report UCL-PSY-CQ1). London: University College.

Houghton, G., Glasspool, D. W., & Shallice, T. (1994). Spelling and serial recall: Insights from a competitive queueing model. In M. Oaksford & G. D. A. Brown (Eds.), *Handbook of spelling: Theory, process and intervention.* Wiley: Chichester.

Houghton, G., & Hartley, T. (1996). Parallel models of serial behaviour: Lashley revisited. *PSYCHE, 2*, 25.

Houghton, G., Hartley, T., & Glasspool, D. W. (1996). The representation of words and nonwords in short-term memory: serial order and syllable structure. In S. E. Gathercole (Ed.), *Models of Short-Term Memory* (pp. 101-128). UK: Psychology Press.

Houghton, G., & Tipper, S. P. (1996). Inhibitory mechanisms of neural and cognitive control: applications to selective attention and sequential action. *Brain & Cognition*, 30, 20-43.

Houston, J. P. (1976). Item versus order information, proactive inhibition, and serial recall. *American Journal of Psychology, 89*, 507-514.

Howell, D. C. (1992) *Statistical methods for psychology* (3rd ed.). Blemont: Duxbury Press.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 685-701.

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology, 86*, 527-536.

Hunt, P. R. (1995). The subtlety of distinctiveness: what Von Restorff really did. *Psychonomic Bulletin & Review, 2*, 105-112.

Jahnke, J. C. (1969a). Output interference and the Ranschburg effect. *Journal of Verbal Learning and Verbal Behavior, 8*, 614-621.

Jahnke, J. C. (1969b). The Ranschburg effect. *Psychological Review, 76*, 592-605.

Jahnke, J. C. (1970). Probed recall of strings that contain repeated elements. *Journal of Verbal Learning and Verbal Behavior, 9*, 450-455.

Jahnke, J. C. (1972a). Distribution of repeated and nonrepeated target elements and short-term recognition memory. *Journal of Experimental Psychology, 96*, 345-353.

Jahnke, J. C. (1972b). The effects of intraserial and interserial repetition on recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 706-716.

Jahnke, J. C. (1974). Restrictions on the Ranschburg effect. *Journal of Experimental Psychology, 103*, 183-185.

Jahnke, J. C., & Bower, R. E. (1986). Are there two Ranschburg effects? *American Journal of Psychology, 99*, 275-288.

Jahnke, J. C., Davis, S. T., & Bower, R. E. (1989). Position and order information in recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 859-867.

Jahnke, J. C., & Melton, A. W. (1968). Acoustic similarity and the Ranschburg phenomenon. *Proceedings of the 76th annual convention of the American Psychological Association,* 65-66.

Jensen, A. R., & Rohwer, W. D. (1965). What is learned in serial learning? *Journal of Verbal Learning and Verbal Behavior, 4*, 62-72.

Johnson, G. J. (1991). A distinctiveness model of serial learning. *Psychological Review, 98,* 204-217.

Johnson, N. (1972). Organisation and the concept of a memory code. In Melton, A. W. & Martin, E. (Eds.), *Coding process in human memory* (pp. 125-159). NJ: Winston & Sons.

Jones, D. M. (1992). Objects, streams, and threads of auditory attention. In A. D. Baddeley and L. Weiskrantz (Eds.), *Attention: selection, awareness and control: A tribute to Donald Broadbent* (pp. 87-104). Oxford: Clarendon Press.

Jones, G. V. (1976). A fragmentation hypothesis of memory: cued recall of pictures and of serial position. *Journal of Experimental Psychology: General, 105*, 277-293.

Jones, D. M., Beaman, P., & Macken, W. J. (1996). The object-oriented episodic record model. In S. E. Gathercole (Ed.), *Models of Short-Term Memory* (pp. 209-238). UK: Psychology Press.

Jones, D. M., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental*

*Psychology: Learning, Memory and Cognition, 21*, 1008-1018.

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19*, 369-381.

Jones, D. M., & Macken, W. J. (1995a). Organisational factors in the effect of irrelevant speech: the role of spatial location and timing. *Memory & Cognition, 23,* 192-200.

Jones, D. M., & Macken, W. J. (1995b). Phonological similarity in the irrelevant speech effect: within- or between-stream similarity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 103-115.

Jones, T. C., & Roediger, H. L. (1995). The experiential basis of serial position effects. *European Journal of Cognitive Psychology, 7*, 65-80.

Jordan, M. I. (1986). *Serial order: a parallel distributed approach* (ICS Report 8604). San Diego: University of California, Institute for Cognitive Science.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kanwisher, N. G. (1987). Repetition blindness: type recognition without token individuation. *Cognition, 27*, 117-143.

Kanwisher, N., Driver, J., & Machado, L. (1995). Spatial repetition blindness is modulated by selective attention to color or shape. *Cognitive Psychology, 29,* 303-337.

Keppel, G., & Saufley, W. H. Jr. (1964). Serial position as a stimulus in serial learning. *Journal of Verbal Learning and Verbal Behavior, 3,* 335-343.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1*, 153-161.

Kidd, G. R., & Greenwald, A. G. (1988). Attention, rehearsal, and memory for serial order. *American Journal of Psychology, 101*, 259-279.

Klahr, D., Chase, W., & Lovelace, E. A. (1983). Structure and process in alphabetic retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 9, 462-477.

Koriat, A., Ben-Zur, H., & Sheffer, D. (1988). Telling the same story twice: output monitoring and age. *Journal of Memory and Language, 27,* 23-39.

Larsen, S. F., & Thompson, C. P. (1995). Reconstructive memory in the dating of personal and public news events. *Memory & Cognition, 23,* 780-790.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior. The Hixon Symposium* (pp. 112-136). New York: John Wiley & Sons, Inc.

LeCompte, D. C., & Watkins, M. J. (1995). Grouping in primary memory: the case of the compound suffix. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 96-102.

Lee, C. L. (1976a). Repetition and acoustic contrast in short-term memory for letter sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 2,* 695-704.

Lee, C. L. (1976b). Short-term recall of repeated items and detection of repetitions in letter sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 2,* 120-127.

Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Memory, 16,* 395-418.

Lee, C. L., & Estes, W. K. (1981). Item and order information in short-term memory: evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 149-169.

Levelt, W. J. (1989). *Speaking: From Intention to Articulation*. London: MIT Press.

Lewandowsky, S., & Li, S. (1994). Memory for serial order revisited. *Psychological Review, 101*, 539-543.

Lewandowsky, S., & Murdock, B. B., Jr. (1989). Memory for serial order. *Psychological Review, 96*, 25-57.

Li, S., & Lewandowsky, S. (1993). Intralist distractors and recall direction: constraints on models of memory for serial order. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*, 895-908.

Li, S., & Lewandowsky, S. (1995). Forward and backward recall: different retrieval processes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 837-847.

Luo, C. R., & Caramazza, A. (1995). Repetition blindness under minimum memory load:

effects of spatial and temporal proximity and the encoding effectiveness of the first item. *Perception & Psychophysics, 57*, 1053-1064.

MacKay, D. G. (1982). The problems of flexibility, fluency and speed-accuracy trade-off in skilled behavior. *Psychological Review, 89*, 483-506.

MacKay, D. G. (1987). *The organization of perception and action.* NY: Springer Verlag.

MacKay, D. G., Abrams, L., & Pedroza, M. J. (1996). *Error detection, functional dysgraphia, and repetition deficits: new age-linked asymmetries between production versus perception*. Manuscript submitted for publication.

Macken, W. J., & Jones, D. M. (1995). Functional characteristics of the inner voice and the inner ear: single or double agency? *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 436-448.

Madigan, S. A. (1971). Modality and recall order interactions in short-term memory for serial order. *Journal of Experimental Psychology, 87,* 294-296.

Madigan, S. A. (1980). The serial position curve in immediate serial recall. *Bulletin of the Psychonomic Society, 15,* 335-338.

Maisto, A., & Ward, L. (1973). Test of the ordinal position hypothesis using serial anticipation and serial recall procedures. *Journal of Experimental Psychology, 101*, 232-236.

Malmi, R. A., & Jahnke, J. C. (1972). Invariant repetition structure and the Ranschburg effect. *Psychonomic Science, 28*, 247-250.

Mandler, G., & Anderson, S. E. (1971). Temporal and spatial cues in seriation. *Journal of Experimental Psychology, 90*, 128-135.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis.* London: Academic Press.

Margrain, S. A. (1967). Short-term memory as a function of input modality. *The Quarterly Journal of Experimental Psychology, 19*, 109-114.

Martin, E. (1974). Serial learning: a multilevel access analysis. *Memory & Cognition, 2*, 322-328

McElree, B., & Dosher, B. A. (1993). Serial retrieval in the recovery of order information. *Journal of Experimental Psychology: General, 122*, 291-315.

McNicol, D. (1975). Fixed and random address models for storing order in short-term memory. *The Quarterly Journal of Experimental Psychology, 27*, 273-288.

McNicol, D. (1978). Initial codes for order in memory. *The Quarterly Journal of Experimental Psychology, 30*, 681-691.

McNicol, D., & Heathcote, A. (1986). Representation of order information: an analysis of grouping effects in short-term memory. *Journal of Experimental Psychology: General, 115*, 76-95.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*, 1-21.

Metcalfe, J., & Sharpe, D. (1985). Ordering and reordering in the auditory and visual modalities. *Memory & Cognition, 13*, 435-441.

Mewaldt, S. P., & Hinrichs, J. V. (1973). The Ranschburg effect: stimulus variables and scoring criterion. *Memory & Cognition, 1*, 177-182.

Mewaldt, S. P., & Hinrichs, J. V. (1977). Repetition and inference in short-term memory. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 572-581.

Mewhort, D. J. K., Popham, D. & James, G. (1994). On serial recall: A critique of chaining in the Theory of Distributed Associative Memory. *Psychological Review, 101*, 534-538.

Miller, G. A. (1956/1994). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 101*, 343-352.

Monsell, S. (1973). Information processing in short term memory tasks. Unpublished Doctoral Dissertation, University of Oxford.

Monsell, S. (1978). Recency, immediate recognition memory and reaction time. *Cognitive Psychology, 10*, 465-501.

Multhaup, K. S., Balota, D. A., & Cowan, N. (1996). Implications of aging, lexicality and item length for the mechanisms underlying memory span. *Psychonomic Bulletin & Review, 3*, 112-120.

Murdock, B. B., Jr. (1960). The distinctiveness of stimuli. *Psychological Review, 67*, 16-31.

Murdock, B. B. Jr. (1968). Serial order effects in short-term memory. *Journal of Experimental Psychology Monograph Supplement, 76*, 1-15.

Murdock, B. B. Jr. (1969). Where or when: modality effects as a function of temporal or spatial distribution of information. *Journal of Verbal Learning and Verbal Behavior, 8*, 378-383.

Murdock, B. B., Jr. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General, 105*, 191-216.

Murdock, B. B., Jr. (1983). A distributed model for serial-order information. *Psychological Review, 90*, 316-338.

Murdock, B. B., Jr. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review, 100*, 183-203.

Murdock, B. B., Jr. (1995). Developing TODAM: Three models for serial order information. *Memory & Cognition, 23*, 631-645.

Murdock, B. B., Jr., & vom Saal, W. (1967). Transpositions in short-term memory. *Journal of Experimental Psychology, 74*, 137-143.

Murray, D. J., (1965). Vocalisation-at-presentation and immediate recall, with varying presentation rates. *The Quarterly Journal of Experimental Psychology, 17*, 47-56.

Murray, D. J., (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology, 21*, 263-276.

Muter, P. (1980). Very rapid forgetting. *Memory & Cognition, 8,* 174-179.

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition, 16*, 343-352.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251-269.

Nairne, J. S. (1991). Positional uncertainty in long-term memory. *Memory & Cognition, 19,* 332-340.

Nairne, J. S., & Neath, I. (1994). Critique of the retrieval/deblurring assumptions of the theory of distributed associative memory. *Psychological Review, 101*, 528-533.

Nairne, J. S., Whiteman, H. L., & Woessner, L. (1995). Symmetrical cuing effects for item and position information. *American Journal of Psychology, 108*, 345-358.

Navey-Benjamin, M. (1990). Coding of temporal order information: an automatic process?

*Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 117-126.

Neath, I. (1993a). Contextual and distinctive processes in the serial position function. *Journal of Memory and Language, 32*, 820-840.

Neath, I. (1993b). Distinctiveness and serial position effects in recognition. *Memory & Cognition, 21*, 689-698.

Neath, I., & Crowder, R. G. (1990). Schedules of presentation and temporal distinctiveness in human memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 316-327.

Neath, I., & Crowder, R. G. (1996). Distinctiveness and very short-term serial position effects. *Memory, 4*, 225-242.

Neath, I., & Nairne, J. (1995). Word-length effects in immediate memory: over-writing trace decay theory. *Psychological Bulletin & Review, 2*, 429-441.

Nelson, T. O., & Chaiklin, S. (1980). Immediate memory for spatial location. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 529-545.

Nigrin, A. L. (1993). *Neural networks for pattern recognition.* Cambridge, MA: MIT Press.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorisation relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Palmer, S. E., & Ornstein, P. A. (1971). The role of rehearsal strategy in serial probed recall. *Journal of Experimental Psychology, 88*, 60-66.

Page, M. P. A. (1994). Modelling the perception of musical sequences with self-organising neural networks. *Connection Science, 6*, 223-246.

Page, M. P. A., & Norris, D. G. (1996a). [A series of experiments on memory for serial order]. Unpublished raw data.

Page, M. P. A., & Norris, D. G. (1996b). The primacy model: a new model of immediate serial recall. Manuscript submitted for publication.

Park, J., & Kanwisher, N. (1994). Determinants of repetition blindness. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 500-519.

Penney, C. G. (1978). Suffix effect in lists of temporally grouped words. *The Quarterly Journal of Experimental Psychology, 30*, 235-250.

Penney, C. G. (1985). Elimination of the suffix effect on preterminal items with unpredictable list length: evidence for a dual model of suffix effects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11*, 229-247.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition, 17*, 398-422.

Peressotti, F., & Grainger, J. (1995). Letter-position coding in random consonant arrays. *Perception & Psychophysics, 57*, 875-890.

Peterson, L. R., & Johnson, S. T. (1971). Some effects of minimizing articulation on short-term retention. *Journal of Verbal Learning and Verbal Behavior, 10*, 346-354.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193-198.

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology, 48*, 384-404.

Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology, 57*, 137-146.

Posner, M. I., & Konick, A. F. (1966). On the role of interference in short-term retention. *Journal of Experimental Psychology, 72*, 221-231.

Potts, G. R., & Shiffrin, R. (1970). Repetitions, blank trials, and the Von Restorff effect in free recall memory. *Journal of Experimental Psychology, 86*, 128-130.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93-134.

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review, 88*, 552-572.

Richman, H. B., & Simon, H. A. (1994). *EPAM simulations of short-term memory.* (Complex Information Processing Working Paper 514). Carnegie-Mellon University, Department of Psychology.

Robinson, G. M. (1977). Rhythmic organisation in speech processing. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 83-91.

Routh, D. A., & Frosdick, R. M. (1978). The basis and implications of the restoration of a recency effect in immediate serial recall. *The Quarterly Journal of Experimental Psychology, 30,* 201-220.

Rumelhart, D. E., & Norman, D. (1982). Simulating a skilled typist: a study of skilled cognitive-motor performance. *Cognitive Science, 6*, 1-36.

Ryan, J. (1969a). Grouping and short-term memory: different means and patterns of grouping. *The Quarterly Journal of Experimental Psychology, 21*, 137-147.

Ryan, J. (1969b). Temporal grouping, rehearsal and short-term memory. *The Quarterly Journal of Experimental Psychology, 21*, 148-155.

Salame, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 150-164.

Salame, P., & Baddeley, A. (1986). Phonological factors in STM: similarity and the unattended speech effect. *Bulletin of the Psychonomic Society, 24*, 263-265.

Salame, P., & Baddeley, A. (1990). The effects of irrelevant speech on immediate free recall. *Bulletin of the Psychonomic Society, 28*, 540-542.

Sanders, A. F. (1975). Some remarks on short-term memory. In P. M. A. Rabbit & S. Dornic (Eds.), *Attention and Performance V* (pp. 241-267). Hillside, NJ: LEA.

Sanders, A. F., & Willemsen, E. M. (1978a). Proactive interference in immediate serial recall, *Acta Psychologica, 42*, 29-38.

Sanders, A. F., & Willemsen, E. M. (1978b). The course of proactive interference in immediate probed recall, *Acta Psychologica, 42*, 133-144.

Schacter, D. L., & Tulving, E. (Eds.). (1994). *Memory Systems*. London: MIT Press.

Scharroo, J. (1994). Modeling alphabetic retrieval: rejoinder to Klahr (1994). *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 492-495.

Schneider, W., & Detweiler, M. (1988). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation, 22* (pp. 53-120). New York: Academic Press.

Schweikert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21,* 168-175.

Schweikert, R., & Boruff, B. (1986). Short-term memory capacity: magic number seven or magic spell? *Journal of Experimental Psychology: Learning, Memory and Cognition, 1*6, 419-425.

Seamon, J. G., & Chumbley, J. I. (1977). Retrieval process for serial order information. *Memory & Cognition, 5,* 709-715.

Sechler, E. S., & Watkins, M. J. (1991). Learning to reproduce a list and memory for the learning. *American Journal of Psychology, 104,* 367-394.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioural and Brain Sciences*, *17*, 367-448.

Shiffrin, R., & Cook, J. (1978). Short-term forgetting of item and order information. *Journal of Verbal Learning and Verbal Behavior, 17,* 189-218.

Slamecka, N. (1964). An inquiry into the doctrine of remote associations. *Psychological Review, 71*, 61-76.

Slamecka, N. (1967). Serial learning and order information. *Journal of Experimental Psychology, 74*, 62-66.

Slamecka, N. (1985). Ebbinghaus: some associations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11,* 414-435.

Smyth, M. M., & Scholey, K. A. (1996). Serial order in spatial immediate memory. *The Quarterly Journal of Experimental Psychology, 49A*, 159-177.

Sperling, G., & Melchner, M. J. (1976). Estimating item and order information. *Journal of Mathematical Psychology, 13*, 192-213.

Squire, L. R. (1994). Declarative and nondeclarative memory: multiple brain systems supporting learning and memory. In D. L. Schacter & E. Tulving (Eds.), *Memory Systems* (pp. 203-231). London: MIT Press.

Stadler, M. A. (1993). Implicit serial learning: questions inspired by Hebb (1961). *Memory & Cognition, 12*, 819-827.

Sternberg, S. (1967). Retrieval of contextual information from memory. *Psychonomic Science, 8*, 55-56.

Sternberg, S. (1969). Memory scanning: mental processes revealed by reaction time experiments. *American Scientist, 57*, 421-457.

Surprenant, A. M., Pitt, M. A., & Crowder, R. G. (1993). Auditory recency in immediate memory. *The Quarterly Journal of Experimental Psychology, 46A*, 193-223.

Tehan, G., & Humphreys, M. S. (1995). Transient phonemic codes and immunity to proactive interference. *Memory & Cognition, 23*, 181-191.

Tell, P. M. (1971). Influence of vocalisation on short-term memory. *Journal of Verbal Learning and Verbal Behavior, 10*, 149-156.

Toglia, M. P., & Kimble, G. A. (1976). Recall and use of serial position information. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 431-445.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 145-152.

Treisman, M., Cook, N., Naish, P. L. N., & MacCrone, J. K. (1994). The internal clock: electroencephalographic evidence for oscillatory processes underlying time perception. *The Quarterly Journal of Experimental Psychology, 47A*, 241-289.

Tulving, E. (1983). *Elements of episodic memory.* Oxford: Oxford University Press.

Turvey, M. T., Brick, P., & Osborn, J. (1970). Proactive interference in short-term memory as a function of prior item retention interval. *The Quarterly Journal of Experimental Psychology, 22*, 142-147.

Tzeng, O. J. L., & Cotton, B. (1980). A study-phase retrieval model of temporal coding. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 705-716.

Tzeng, O. J. L., Lee, A. T., & Wetzel, C. D. (1979). Temporal coding in verbal information processing. *Journal of Experimental Psychology:Human Learning and Memory, 5,* 52-64.

Walsh, M. F., & Schwartz, M. (1977). The Ranschburg effect: tests of the guessing-bias and proactive interference hypothesis. *Journal of Verbal Learning and Verbal Behavior, 16,* 55-68.

Watkins, M. J. (1977). The intricacy of memory span. *Memory & Cognition*, *5*, 529-534.

Watkins, M. J., Watkins, O. C., & Crowder, R. (1974). The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behavior, 13*, 430-447.

Waugh, N., & Norman, D. (1965). Primary memory. *Psychological Review, 72,* 89-104.

Whittlesea, B. W. A., Dorken, M. D., & Podrouzek, K. W. (1995). Repeated events in rapid lists: Part 1. Encoding and representation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 1670-1688.

Whittlesea, B. W. A., & Podrouzek, K. W. (1995). Repeated events in rapid lists: Part 2. Remembering repetitions. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 1689-1697.

Wickelgren, W. A. (1965a). Acoustic similarity and retroactive interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 4,* 53-61.

Wickelgren, W. A. (1965b). Short-term memory for phonemically similar lists. *American Journal of Psychology, 78,* 567-574.

Wickelgren, W. A. (1965c). Short-term memory for repeated and non-repeated items. *The Quarterly Journal of Experimental Psychology, 17,* 14-25.

Wickelgren, W. A. (1966). Associative intrusions in short-term recall. *Journal of Experimental Psychology, 72,* 853-858.

Wickelgren, W. A. (1967). Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology, 19,* 97-102.

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review, 76,* 1-15.

Wolf, T. M., & Jahnke, J. C. (1968). Effects of intraserial repetition on short-term recognition and recall. *Journal of Experimental Psychology, 77,* 572-580.

Young, R. (1962). Tests of the hypothesis about the effective stimulus in serial learning. *Journal of Experimental Psychology, 63,* 307-313.

Young, R. (1968). Serial learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal Behavior and General Behavior Theory* (pp. 122-148). Englewood Cliffs, NJ: Prentice-Hall.

Yntema, D. B., & Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior, 2,* 65-74.

# Appendix 1: Statistical Techniques

**Logit or "Log-odds" Transform**

If subject *i* of *N* makes $r_i$ errors in a sample of $n_i$ responses, giving a basic proportional score of $p_i = r_i / n_i$, then the log-odds score is defined as:

$$L_i = ln\left(\frac{p_i}{1 - p_i}\right) = ln\left(\frac{r_i}{n_i - r_i}\right) \qquad \text{Equation 9-1}$$

This logarithmic transform "stretches" proportional scores just above zero and just below one, making some allowance for floor and ceiling effects. However, the transform is not defined at these extremina exactly, i.e., when $r_i = 0$ or $r_i = n_i$.

**Empirical Log-odds**

To handle situations when $r_i = 0$ or $r_i = n_i$ $(n_i > 0)$, an empirical log-odds score can be defined (Cox & Snell, 1989):

$$L_i = ln\left(\frac{r_i + 0.5}{n_i - r_i + 0.5}\right) \qquad \text{Equation 9-2}$$

This caters for measured proportions of zero or one. However, it makes no allowance for the fact that one proportional measurement may be based on a large number of observations (when $n_i$ is large), and hence likely to be more accurate than one based on only a small number of observations (when $n_i$ is small).

**Weighted Log-odds**

The empirical log-odds score defined above has an associated variance:

$$U_i = \frac{(n_i + 1)\ (n_i + 2)}{n_i\,(r_i + 1)\ (n_i - r_i + 1)} \qquad \text{Equation 9-3}$$

Empirical log-odds scores can be weighted by the inverse of their variances:

$$W_i = \frac{1}{U_i} \qquad \text{Equation 9-4}$$

to give a weighted mean across subjects:

$$\bar{L} = \frac{\sum W_i L_i}{\sum W_i} \qquad \text{Equation 9-5}$$

where the summand is from subject *i=1* to *i=N*.

Transforming back into the original coordinates, the estimated mean proportion over subjects, $\bar{p}$, has variance, $\bar{V}$, given by:

$$\bar{p} = \frac{e^{\bar{L}}}{1 + e^{\bar{L}}} \qquad \bar{V} = \bar{U}\bar{p}^2(1 - \bar{p})^2 \qquad \text{Equation 9-6}$$

**Testing Related, Weighted, Log-odds**

To test a difference in means of two, related log-odds, $L_i$ and $L'_i$, let:

$$d_i = L_i - L'_i \qquad \text{Equation 9-7}$$

A combined weight, $w_i$, can be determined from:

$$w_i = \left(\frac{1}{W_i} + \frac{1}{W'_i}\right)^{-1} \qquad \text{Equation 9-8}$$

The weighted mean difference score is then:

$$\bar{d} = \frac{\sum w_i d_i}{\sum w_i} \qquad \text{Equation 9-9}$$

and the standard error of the difference scores is:

$$\sqrt{\frac{1}{\sum w_i}} \qquad \text{Equation 9-10}$$

which enables testing of the standardised score:

$$Z(N) = \frac{\sum w_i d_i}{\sqrt{\sum w_i}} \qquad \text{Equation 9-11}$$

To get a measure of the weighted mean difference in terms of proportions, $\bar{q}$:

$$\bar{q} = \frac{\bar{p}(1-\bar{p})(e^{\bar{d}}-1)}{1+\bar{p}(e^{\bar{d}}-1)} \qquad \text{Equation 9-12}$$

given a particular baseline proportion $\bar{p}$.

**Testing Unrelated, Weighted, Log-odds**

To test the difference between log-odds of two, unrelated groups (with $N$ subjects):

$$Z(N) = \frac{\bar{L} - \bar{L'}}{\sqrt{\frac{1}{\sum W_i} + \frac{1}{\sum W_i'}}} \qquad \text{Equation 9-13}$$

**Testing 2x2 Contingency Tables**

Given $r_i$ errors and $(n_i\text{-}r_i)$ correct responses on one measure, and $s_i$ errors and $(n_i\text{-}s_i)$ correct responses on another, such that the 2x2 contingency table for each subjects is:

| | | |
|---|---|---|
| $x_i$ | .. | $r_i$ |
| .. | .. | $n_i\text{-}r_i$ |
| $s_i$ | $n_i\text{-}s_i$ | $n_i$ |

Then, under the null hypothesis that the two measures are uncorrelated, the expected number of cases where they are in agreement, *E(x),* is:

$$x = \sum x_i \qquad E(x) = \sum \frac{r_i s_i}{n_i} \qquad \text{Equation 9-14}$$

and the variance of *x* is:

$$V(x) = \sum \frac{r_i s_i (n_i - r_i)(n_i - s_i)}{n_i n_i (n_i - 1)} \qquad \text{Equation 9-15}$$

which allows a combined test of significance of individual subjects' two-by-two contingency tables by a Z-score:

$$Z(N) = \frac{x - E(x)}{\sqrt{V(x)}} \qquad \text{Equation 9-16}$$

giving a measure of the association or correlation between the two measures. Note that this assumes homogeneity across subjects, such that they all show a similar association in their individual contingency tables.

**Testing Conditional Error Probabilities**

Given $m_j$ reports which have no errors on positions *1...j-1* and assuming $r_j$ errors are made on position *j*, then:

$$m_{j+1} = m_j - r_j \qquad \text{Equation 9-17}$$

Let the conditional probability of an error on position *j* be $q_j$. The maximum likelihood estimate of $q_j$ is:

$$\hat{q}_j = \frac{r_j}{m_j} \qquad \text{Equation 9-18}$$

If there is no change in conditional probability of an error across positions $j$ and $j+1$, then the common maximum likelihood estimator is:

$$\hat{q} = \frac{r_{j+1} + r_j}{m_{j+1} + m_j}$$ Equation 9-19

To test the hypothesis that $q_j = q_{j+1}$, the goodness of fit of the following model is tested:

$$E[r_j] = m_j\hat{q} \qquad E[r_{j+1}] = m_j(1-\hat{q})\hat{q}$$

$$E[m_{j+2}] = m_j(1-\hat{q})(1-\hat{q})$$ Equation 9-20

The corresponding $X^2$ statistic has (approximately) a Chi-squared distribution on one degree of freedom under the null hypothesis. A combined $\chi^2$ across subjects can be obtained by summing individual, signed Z scores, squaring and dividing by $N$.

**Significance of Multiple Pairwise Comparisons**

Given $N$ pairwise, a priori comparisons with individual significance levels of $\alpha$, the appropriate familywise significance level, $\alpha_F$, according to a Bonferroni correction is:

$$\alpha_F = \frac{\alpha}{N}$$ Equation 9-21

This correction is very conservative. A more powerful approach is Holm's method (Howell, 1992), for which the Bonferroni correction is applied iteratively to each individual comparison, testing the largest absolute difference against $\alpha_F$ above, and testing $i$th next largest difference ($i=1..N-1$) against $\alpha_i$, where:

$$\alpha_i = \frac{\alpha}{N-i}$$ Equation 9-22

The iteration continues until the $i$th comparison is nonsignificant, whence all remaining comparisons are also deemed nonsignificant.

**Hotelling's T-squared Test**

Hotelling's one-sample $T^2$-test can be used to test $p$ means of samples taken from $n$ subjects against hypothesised values (Mardia, Kent & Bibby, 1979). Let **h** be the vector of hypothesised means, **d** be the vector of actual means, and **S** be the matrix of the sums of squares of data values. Then the vector **t** of the differences of hypothesised and actual means:

$$\boldsymbol{t} = \boldsymbol{d} - \boldsymbol{h} \qquad \text{Equation 9-23}$$

gives the $T^2$ statistic:

$$T^2 = n\,(\boldsymbol{t}^T \boldsymbol{S}^{-1} \boldsymbol{t}) \qquad \text{Equation 9-24}$$

which can be tested by the $F$-ratio:

$$F\,(p, n-p) = \frac{(n-p)}{p\,(n-1)}\,T^2 \qquad \text{Equation 9-25}$$

Note that $T^2$ statistic, by taking into account the variances and covariances of the $p$ data samples, does not have to assume independence of the $p$ means.

# Appendix 2: Details of Meta-Analyses

The details of the experiments in the meta-analyses of Chapter 4 are given below.

### Meta-analysis 1

Immediate serial recall of ungrouped lists of phonologically dissimilar items.

| Condition | List Length | Items | Rate (item / s) | Present. Modality | Recall Method | No. Lists | No. Subjects. |
|---|---|---|---|---|---|---|---|
| 1 | 6 | letters | 0.50 | visual | written | 12 | 48 |
| 2 | 6 | letters | 0.75 | vocalised | spoken | 21 | 13 |
| 3 | 7 | letters | 0.75 | vocalised | spoken | 21 | 14 |
| 4 | 6 | letters | 0.75 | visual | written | 18 | 13 |
| 5 | 7 | letters | 0.75 | visual | written | 21 | 11 |
| 6 | 7 | digits | 0.60 | visual | written | 20 | 18 |
| 7 | 8 | digits | 0.60 | visual | written | 20 | 18 |
| 8 | 9 | digits | 0.60 | visual | written | 20 | 18 |
| 9 | 5 | letters | 0.75 | visual | written | 11 | 10 |
| 10 | 6 | letters | 0.75 | visual | written | 11 | 10 |
| 11 | 7 | letters | 0.75 | visual | written | 11 | 10 |
| 12 | 8 | letters | 0.75 | visual | written | 11 | 10 |
| 13 | 9 | letters | 0.75 | visual | written | 30 | 25 |
| 14 | 8 | letters | 1.00 | visual | written | 30 | 36 |
| 15 | 8 | letters | 1.00 | auditory | written | 30 | 36 |
| 16 | 5 | words | 1.00 | visual | spoken | 12 | 16 |
| 17 | 6 | words | 1.00 | visual | spoken | 12 | 16 |
| 18 | 5 | words | 1.00 | visual | spoken | 12 | 16 |
| 19 | 6 | words | 1.00 | visual | spoken | 12 | 16 |
| 20 | 5 | words | 1.00 | visual | written | 12 | 14 |
| 21 | 6 | words | 1.00 | visual | written | 12 | 14 |
| 22 | 5 | words | 1.00 | visual | written | 12 | 14 |

| Condition | List Length | Items | Rate (item/s) | Present. Modality | Recall Method | No. Lists | No. Subjects. |
|---|---|---|---|---|---|---|---|
| 23 | 6 | words | 1.00 | visual | written | 12 | 14 |
| 24 | 5 | words | 1.00 | visual | written | 12 | 16 |
| 25 | 6 | words | 1.00 | visual | written | 12 | 16 |
| 26 | 5 | words | 1.00 | visual | written | 12 | 16 |
| 27 | 6 | words | 1.00 | visual | written | 12 | 16 |
| 28 | 5 | words | 1.00 | visual | written | 6 | 16 |
| 29 | 6 | words | 1.00 | visual | written | 6 | 16 |
| 30 | 5 | words | 1.00 | visual | written | 6 | 16 |
| 31 | 6 | words | 1.00 | visual | written | 6 | 16 |
| 32 | 5 | words | 1.00 | visual | written | 6 | 16 |
| 33 | 6 | words | 1.00 | visual | written | 6 | 16 |
| 34 | 5 | words | 1.00 | visual | written | 6 | 16 |
| 35 | 6 | words | 1.00 | visual | written | 6 | 16 |
| 36 | 9 | digits | 1.00 | visual | written | 26 | 12 |
| 37 | 9 | letters | 1.00 | visual | written | 20 | 12 |

Condition 1 corresponds to the PN condition of Experiment 1.

Conditions 2 and 3 correspond to the PN conditions of high- and low-span groups of Experiment 2 in Henson et al. (1996).

Conditions 4 and 5 correspond to the PN conditions of high- and low-span groups of Experiment 3 in Henson et al. (1996).

Conditions 6, 7 and 8 correspond to the U7, U8 and U9 conditions of Experiment 2.

Conditions 9, 10, 11 and 12 correspond to the fixed length conditions of an unpublished study by Page and Norris (1996a) looking at list length effects.

Condition 13 corresponds to the ungrouped condition of an unpublished study by Page and Norris (1996a) looking at grouping.

Conditions 14 and 15 correspond to the visual and auditory conditions of an unpublished study by Page and Norris (1996a) looking at modality effects.

Conditions 16 to 35 correspond to four control conditions in an unpublished study of five experiments by Page and Norris (1996a) looking at word-length effects. Conditions 18, 19, 22, 23, 26, 27, 30, 31, 34, 35 used five-syllable words; others used one-syllable words.

Conditions 36 and 37 are the ungrouped control conditions in a series of two experiments looking at irrelevant tones and grouping in Henson (1996a).

## Meta-analysis 2

Immediate serial recall of grouped lists of phonologically dissimilar items. All lists were grouped temporally in the manner indicated after the list length below.

| Condition | List Length | Items | Rate (item/s) | Present. Modality | Recall Method | No. Lists | No. Subjects. |
|---|---|---|---|---|---|---|---|
| 1 | 9 (333) | digits | 0.60 | visual | written | 20 | 18 |
| 2 | 9 (333) | letters | 0.75 | visual | written | 30 | 25 |
| 3 | 9 (333) | letters | 1.00 | visual | written | 9 | 9 |
| 4 | 9 (333) | digits | 1.00 | visual | written | 26 | 12 |
| 5 | 9 (333) | letters | 1.00 | visual | written | 20 | 12 |
| 6 | 8 (44) | digits | 0.60 | vocalised | spoken | 24 | 18 |
| 7 | 8 (44) | digits | 1.00 | visual | written | 33 | 30 |
| 8 | 8 (44) | digits | 1.00 | visual | written | 57 | 30 |
| 9 | 8 (44) | digits | 1.00 | visual | written | 140 | 45 |

Condition 1 corresponds to the G9 condition of Experiment 2.

Condition 2 corresponds to the grouped condition of an unpublished study by Page and Norris (1996a) looking at grouping.

Condition 3 corresponds to a grouped condition of an unpublished study by Frankish (personal communication, 1995).

Conditions 4 and 5 correspond to the grouped control conditions in a study of two experiments looking at irrelevant tones and grouping in Henson (1996a).

Condition 6 corresponds to the forward recall condition in Henson (1995).

Conditions 7, 8 and 9 correspond to an unpublished study of three experiments by Page and Norris (1996a) looking at proactive interference in grouped lists.

# Meta-analysis 3

Immediate serial recall of ungrouped lists of alternating phonologically similar and phonologically dissimilar items.

| Condition | List Length | Items | Rate (item/s) | Present. Modality | Recall Method | No. Lists | No. Subjects. |
|---|---|---|---|---|---|---|---|
| 1 | 6 | letters | 0.60 | visual | written | 12 | 48 |
| 2 | 6 | letters | 0.60 | visual | written | 12 | 48 |
| 3 | 6 | letters | 0.75 | visual | written | 18 | 13 |
| 4 | 6 | letters | 0.75 | visual | written | 18 | 13 |
| 5 | 7 | letters | 0.75 | visual | written | 21 | 11 |
| 6 | 7 | letters | 0.75 | visual | written | 21 | 11 |
| 7 | 6 | letters | 0.75 | vocalised | spoken | 18 | 13 |
| 8 | 6 | letters | 0.75 | vocalised | spoken | 18 | 13 |
| 9 | 7 | letters | 0.75 | vocalised | spoken | 21 | 11 |
| 10 | 7 | letters | 0.75 | vocalised | spoken | 21 | 11 |

Conditions 1 and 2 correspond to the AC and AN conditions of Experiment 1.

Conditions 3 and 4 correspond to the AC and AN conditions of the low-span group in Experiment 2 of Henson et al. (1996).

Conditions 5 and 6 correspond to the AC and AN conditions of the high-span group in Experiment 2 of Henson et al. (1996).

Conditions 7 and 8 correspond to the AC and AN conditions of the low-span group in Experiment 3 of Henson et al. (1996).

Conditions 9 and 10 correspond to the AC and AN conditions of the low-span group in Experiment 3 of Henson et al. (1996).

**Experimental Procedure for Serial Recall**

These meta-analyses have collapsed over differences in list-length, items, presentation rate, presentation modality and recall method. Such differences were of secondary concern to the models described in Chapters 4 and 5. Nonetheless, they may have subtle effects on error patterns in serial recall. For example, longer lists are more likely to be grouped subjectively, affecting the pattern of transpositions (Chapter 3); digits, letters and words come from vocabularies of different sizes, affecting the incidence of intrusions; slower presentation rates allow more time-based decay of phonological representations, but greater opportunity for rehearsal (Chapter 5); auditory presentation may introduce additional effects of echoic storage (Chapter 5); written recall may allow more scope for reordering and editing responses.

In the author's opinion, the experimental design most suitable for examining short-term memory for serial order (in the absence of other constraints) is the following:

Lists of between 4-7 items, minimising the risk of subjective grouping and producing performance levels close to span (performance too good will suffer from ceiling effects; performance too low is likely to produce a large proportion of omissions and random guesses).

Lists of consonants, balanced and low in predictability (Henson et al., 1996), and with obvious acronyms removed (digits lead to too many erroneous runs, such as *5678...*, and have little scope for phonological similarity and intrusions, while words can be semantically recoded, though they are of course necessary to study effects of word-length, familiarity, etc.).

Visual, sequential presentation of items with vocalisation (to aid concentration and allow monitoring by the experimenter) in a regular, monotone voice (to reduce grouping).

Presentation rates of about 2 items per second, which gives little time for rehearsal (Baddeley & Lewis, 1984), but ensures few errors in encoding (Aaronson, 1968), given concurrent vocalisation.

A short delay of shadowing irrelevant distractors (e.g., digits) to prevent rehearsal, minimise potentially confounding effects of auditory information from vocalisation of list items (Tell, 1971) and possibly allow titration of performance to appropriate levels.

Spoken recall to enforce forward recall (spoken responses being harder to reorder than written responses), prevent reperception of previous responses (as in written recall, which allows editing of responses such as those causing repetitions) and possibly allow measurement of response times.

# Appendix 3: Formal Definition of SEM

There are two versions of SEM: a single-trial version and a multiple-trial version. The single-trial version does not model intertrial effects: The tokens in short-term memory are restricted to those from the most recent list. The multiple-trial version is more general, including tokens from previous trials, together with general context, phonological decay and rehearsal. These versions are formalised below.

Note that the formalism can obscure the relatively simple mechanisms underlying the model, which are described verbally in Chapters 5 and 6. In particular, some of the parameters and variables introduced in Chapter 5 require additional suffices for clarification below. In general, parameters are in upper-case and variables are in lower-case. The large number of parameters reflects the generality of the model. In most cases, these parameters are either fixed, constrained by the experimental design, or constrained by values of other parameters.

## Single-trial Version of SEM

The single-trial version takes a single list of items, and simulates $N_L$ independent trials at serial recall of that list. Each trial can be split into two stages of presentation and recall.

### Presentation

A token is created for each item at position $p=1..N_P$ of the list. Specifically, for each group $g=1..N_G$ and each item in group $g$, $i=1..N_I(g)$, a token $t$ is created with positional codes $\boldsymbol{p}_I^{(t)}$ and $\boldsymbol{p}_G^{(t)}$.[1]

The vector $\boldsymbol{p}_I^{(t)} = < x_I(i) \;\; y_I(i) >$ is a positional code for the position of item $i$ within group $g$, where $x_I(i)$ and $y_I(i)$ are the strengths of markers for the start and end of that group:

$$x_I(i) = S_{0,I}S^{i-1} \qquad y_I(i) = E_{0,I}E_I^{N_I-i} \qquad \text{Equation 10-1}$$

1. Ungrouped lists can be modelled either with $N_G=1$ and $N_I(1)=N_P$, or with $N_G=0$, in which case there are no group start and end markers and $i=1..N_P$.

where $S_{0,I}$ and $S_I$, are parameters reflecting the initial strength of the start marker and the rate of change of its strength, and $E_{0,I}$ and $E_I$ are parameters reflecting the initial strength of the end marker and the rate of change of its strength.

Associated with each positional code $\boldsymbol{p}_I^{(t)}$ is a quantity $d_I^{(t)}$, reflecting the noise in the encoding of that position. The value of $d_I^{(t)}$ is drawn from a zero-mean Gaussian distribution with standard deviation $D_I$ for each position $p$.

The vector $\boldsymbol{p}_G^{(t)} = < x_G(g)\ y_G(g) >$ is a positional code for the position of group $g$ in the list, where $x_G(g)$ and $y_G(g)$ are the strengths of markers for the start and end of that list:

$$x_G(g) = S_{0,G} S_G^{g-1} \qquad y_G(g) = E_{0,G} E_G^{N_G - g}$$ Equation 10-2

where $S_{0,G}$ and $S_G$ are parameters reflecting the initial strength of the start marker and the rate of change of its strength, and $E_{0,G}$ and $E_G$ are parameters reflecting the initial strength of the end marker and the rate of change of its strength.

Associated with each positional code $\boldsymbol{p}_G^{(t)}$ is a quantity $d_G^{(t)}$, reflecting the noise in the encoding of that position. The value of $d_G^{(t)}$ is drawn from a zero-mean Gaussian distribution with standard deviation $D_G$ for each group $g$.

**Recall**

For each response $r=1..N_P$, a cue is generated with positional codes $\boldsymbol{p}_I^{(r)}$ and $\boldsymbol{p}_G^{(r)}$, as defined in Equation 10-1 and Equation 10-2. The noise associated with reinstating these positional codes is given by $d_I^{(r)}$ and $d_G^{(r)}$, again drawn from zero-mean Gaussian distributions with standard deviations $D_I$ and $D_G$ respectively. The variable $d_I^{(r)}$ is drawn for each response; the variable $d_G^{(r)}$ is drawn for each new group recalled.

The retrieval of an item as response $r$ can be divided into six stages:

**Stage 1: Cuing**

The positional codes $\boldsymbol{p}_I^{(r)}$ and $\boldsymbol{p}_G^{(r)}$are matched against the positional codes, $\boldsymbol{p}_I^{(t)}$ and $\boldsymbol{p}_G^{(t)}$, of the $t=1..N_T$ tokens in short-term memory[2], cuing each with strength $q^{(t)}(r)$:

---

2. ($N_T=N_P$ in the single-trial version, but not the multiple-trial version)

$$q^{(t)}(r) = \bar{m}(\boldsymbol{p}_I^{(t)}, \boldsymbol{p}_I^{(r)}, d_I, M_I)\,\bar{m}(\boldsymbol{p}_G^{(t)}, \boldsymbol{p}_G^{(r)}, d_G, M_G)$$ Equation 10-3

where $d_I = d_I^{(t)} + d_I^{(r)}$, $d_G = d_G^{(t)} + d_G^{(r)}$, and the parameters $M_I$ and $M_G$ are match criteria, representing the degree to which positional codes must match in order to be cued, as defined by the linear thresholding function, $\bar{m}$:

$$\bar{m}(\boldsymbol{p}, \boldsymbol{q}, d, M) = \begin{cases} 0 & m(\boldsymbol{p}, \boldsymbol{q}, d) < M \\ m(\boldsymbol{p}, \boldsymbol{q}, d) & m(\boldsymbol{p}, \boldsymbol{q}, d) \geq M \end{cases}$$ Equation 10-4

where $m$ is the noisy match between positional codes:

$$m(\boldsymbol{p}, \boldsymbol{q}, d) = o(\boldsymbol{p}, \boldsymbol{q}) + d$$ Equation 10-5

and $o(\boldsymbol{p},\boldsymbol{q})$ is the overlap between position codes $\boldsymbol{p}$ and $\boldsymbol{q}$:

$$o(\boldsymbol{p}, \boldsymbol{q}) = \sqrt{\boldsymbol{p} \cdot \boldsymbol{q}} \times exp\left(-\sqrt{\sum_k (p_k - q_k)^2}\right)$$ Equation 10-6

where the summand $k$ is over the (two) components of vectors $\boldsymbol{p}$ and $\boldsymbol{q}$.

Note that the positional uncertainty functions, $f(i,j)$, representing the overlap between all $i,j=1..N_p$ positions of a sequence, are given by:

$$f(i, j) = o(\boldsymbol{p}_I^{(i)}, \boldsymbol{p}_I^{(j)})\,o(\boldsymbol{p}_G^{(i)}, \boldsymbol{p}_G^{(j)})$$ Equation 10-7

These functions are also the average, unthresholded cued strength of items at each position $i$ during recall of each response $j$.

**Stage 2: Categorical Selection**

Items compete for selection with a strength proportional to their most strongly cued token. Specifically, the categorical (type) representations of all items $u=1..N_V$ in the vocabulary compete with strength $c_C^{(u)}$, where:

$$c_C^{(u)} = max\{q^{(t)}|_{i(t) = u}\}\,(1 - s_C^{(u)}) + n_C \qquad \text{Equation 10-8}$$

where *i(t)* is the identity of (the item corresponding to) token *t*, $s_C^{(u)}$ is the suppression of the categorical representation of item *u*, and $n_C$ is random noise drawn from a zero-mean Gaussian distribution with standard deviation $G_C$ for each item *u*. The strongest item *u** is selected and passed to Stage 3.

Note that the "max" function in Equation 10-8 could be changed for another function, such as the "sum" of cued strengths of all tokens of a particular item. This choice really depends on empirical data concerning the effect of repeated items in recall (Chapter 7).

**Stage 3. Suppression**

The categorical representation of the item selected at Stage 2 is suppressed, such that $s_C^{(u*)}=1$. Meanwhile, the suppression of all other items *u*, except *u**, wears off according to the update rule:

$$s_C^{(u)} \rightarrow s_C^{(u)}\,exp(-R_S) \qquad \text{Equation 10-9}$$

where $R_S$ is the rate of decay of suppression. This decay is assumed to operate in real-time, though for convenience, suppression is only updated during each response and is assumed to have worn off completely between trials.

**Stage 4. Phonological Retrieval**

The item *u** selected from Stage 2 is matched against a second set of phonological representations in order to articulate a response. The possibility of phonological confusions arises at this stage. Specifically, competition is held over a set of phonological item representations v=$1..N_V$, each of which competes with strength, $c_P^{(v)}$:

$$c_P^{(v)} = c_C^{(v)} + p(v, u^*)\,a_P^{(v)}\,(1 - s_P^{(v)}) + n_P \qquad \text{Equation 10-10}$$

where *p(v,u)* is the phonological similarity between items *v* and *u*, $a_P^{(v)}$ is the activation of the phonological representation of item *v*, $s_P^{(v)}$ is the suppression of the phonological representation of item *v*, and $n_P$ is random noise drawn from a zero-mean Gaussian distribution with standard deviation $G_P$ for each item *v*.

The value of $p(v,u)$ is such that $p(v,u)=1$ if $v=u$, $p(v,u)=P_S$ if item *v* and item *u* are phonologically similar (i.e. confusable), and $p(v,u)=P_D$ if they are dissimilar (i.e., if one is nonconfusable). The value of $a_P^{(v)}$ is such that $a_P^{(v)}=A_P$ if item *v* was in the most recent list, and $a_P^{(v)}=0$ otherwise. The strongest item $v^*$ is passed on to Stage 5.

**Stage 5. Thresholding and Guessing**

If the strength of the item retrieved from Stage 4 is above a guessing threshold $T_G$, such that $c_P^{(v^*)}>T_G$, it is passed directly to Stage 6.

If the strength of the item selected from Stage 4 is below the guessing threshold, but above an omission threshold $T_O$, such that $T_O<c_P^{(v^*)}<T_G$, then an item is guessed instead. This guessing is over the $v=1..N_V$ phonological representations, which compete with strengths $c_G^{(v)}$, given by:

$$c_G^{(v)} = a_P^{(v)} (1 - s_P^{(v)}) + n_G \qquad \text{Equation 10-11}$$

where $n_G$ is a random noise drawn from a zero-mean Gaussian distribution with standard deviation $G_G$ for each item *v*. The item winning this competition is passed to Stage 6.

If the strength of the item retrieved from Stage 4 is below the omission threshold, such that $c_P^{(v^*)}<T_O$, then no item is recalled and an omission is indicated instead. The next response is then cued (returning to Stage 1).

**Stage 6. Output**

The item $v^*$ selected or guessed after Stage 4 is output as response *r*. Its phonological representation is suppressed, such that $s_P^{(v^*)}=0$, and the suppression of other phonological representations decays in the same manner as Equation 10-9. Note that the value of $s_P^{(v)}$ is independent of the value of $s_C^{(u)}$ (i.e., the categorical and phonological representations of the same item represent distinct loci of suppression).

The above process then repeats for response $r+1$, returning to Stage 1.

## Multiple-trial Version

In this version of SEM, short-term memory is assumed to contain tokens from previous trials as well as the most recent trial (i.e., $N_T > N_P$). These tokens include a new component which represents general (nonpositional) context, which cannot be reinstated at recall. In addition, each item recalled is recoded as a new token (coded with its recall position, irrespective of whether that is correct), a process which also reactivates its phonological representation. Finally, the activation of phonological representations is assume to decay over time, to reflect transient nature of phonological information in short-term memory.

The multiple-trial version takes $N_L$ different lists and recalls each one once. Recall of each list $l=1..N_L$, with positions $p=1..N_P(l)$, can be split into presentation, retention, recall and intertrial intervals. Only the differences between the multiple-trial version and the single-trial version are formalised below.

### Presentation

Each token $t$ has three components $\boldsymbol{p}_I^{(t)}$, $\boldsymbol{p}_G^{(t)}$and $\boldsymbol{p}_C^{(t)}$, where $\boldsymbol{p}_I^{(t)}$, $\boldsymbol{p}_G^{(t)}$ are the positional contexts defined in Equation 10-1 and Equation 10-2, and $\boldsymbol{p}_C^{(t)}$ is a one-dimensional vector representing the general (nonpositional) context when token $t$ was created. For mathematical convenience, the current general context is represented by the constant value $E_{0,C}$, and the general context of all tokens in memory is updated each time the general context changes. Thus, each time an item is presented, its token is created with $\boldsymbol{p}_C^{(t)}=\langle E_{0,C}\rangle$. During subsequent contextual changes (e.g., presentation of other items), the general context of all tokens is updated according to:

$$\boldsymbol{p}_C^{(t)} \rightarrow \boldsymbol{p}_C^{(t)} E^c{}_{0,C} \qquad \text{Equation 10-12}$$

where $E_C$ represents the rate of contextual change, and $c$ represents the number of contextual changes (episodes). During presentation of each item, $c$ is parameterised by $c=C_P$.

Presentation of an item $v$ also activates its phonological representation by an amount $a_P^{(v)}=A_P$, while the activation of other phonological representations decays as follows:

$$a_P^{(v)} \rightarrow a_P^{(v)} exp\,(-cR_P) \qquad \text{Equation 10-13}$$

where $R_P$ is the rate of decay of phonological activations, and $c$ is the number of episodes (again, $c=C_P$ during presentation). All items have a baseline activation of $a_P=0$ at the start of the first trial.

Note that when the list length is unpredictable from trial to trial (as in Experiment 5), the behaviour of end markers during presentation differs to that in Equation 10-1. The strength of the end marker coding position in group (or position in list for ungrouped lists) becomes a function of the minimum expected list length, $N_M$, such that:

$$y_I(i) = \begin{matrix} E_{0,I}E_I^{N_M-i} & i < N_M \\ E_{0,I} & i \geq N_M \end{matrix} \qquad \text{Equation 10-14}$$

During recall, when the length is known, the end marker behaves as before (Equation 10-1).

**Retention Interval**

During the retention interval, the general context of all tokens is updated according to Equation 10-12, and the phonological activations of items decay according to Equation 10-13, where $c=C_D$ represents the number of episodes during the (filled) delay before recall.

**Recall**

For each response $r=1..N_P(l)$, a cue is generated with positional context $\boldsymbol{p}_I^{(r)}$, $\boldsymbol{p}_G^{(r)}$ and general context $\boldsymbol{p}_C^{(r)}$, where $\boldsymbol{p}_C^{(r)}$ is always the current context $< E_{0,C} >$. The noise associated with reinstating the positional codes is given by $d_I^{(r)}$ and $d_G^{(r)}$, as before.

The multiple-trial and single-trial versions differ in Stages 1 and 5 of recall:

**Stage 1: Cuing**

The positional and general context of the cue is matched against that of the $t=1..N_T$ tokens in short-term memory, cuing each with strength $q^{(t)}(r)$:

$$q^{(t)}(r) = \bar{m}(\boldsymbol{p}_I^{(t)}, \boldsymbol{p}_I^{(r)}, d_I, M_I)\,\bar{m}(\boldsymbol{p}_G^{(t)}, \boldsymbol{p}_G^{(r)}, d_G, M_G)\,o(\boldsymbol{p}_C^{(t)}, \boldsymbol{p}_C^{(r)}) \qquad \text{Equation 10-15}$$

where $\overline{m}$ is the thresholded overlap as in Equation 10-4. (For simplicity, no noise or match criterion is assumed for the general context.)

**Stage 6: Output**

The item $v*$ selected after thresholding in Stage 5 is output as response $r$, as before. In addition however, its phonological representation is reactivated, such that $a_P^{(v*)}=A_P$, and it is recoded as a new token in short term memory, with positional and general context given by $\boldsymbol{p}_I^{(r)}$, $\boldsymbol{p}_G^{(r)}$ and $\boldsymbol{p}_C^{(r)}$ (i.e., that of the cue for response $r$), together with the noise associated with the encoding process, as before.

Finally, the general context of all tokens is updated according to Equation 10-12, and the phonological activations of items decay according to Equation 10-13, where $c=C_R$ represents the number of episodes during the recall of each item.

Note that, in order to handle uncertain responses in Experiments 4 and 5, the multiple-trial version also includes an uncertainty threshold, $T_U$. This threshold functions much like the omission threshold $T_O$, but is applied at the output rather than thresholding stage. Setting $T_U>T_G$ allows SEM to simulate the removal of uncertain responses from subjects' reports.

**Intertrial interval**

Between trials, the general context of all tokens is updated according to Equation 10-12, where $c=C_I+C_A$. The parameter $C_I$ represents the number of episodes during the intertrial interval and the parameter $C_A$ represents the number of contextual changes owing to attentional shifts during the intertrial interval (Chapter 5). The activations of phonological representations also decay according to Equation 10-13, with $c=C_I$.

Finally, the suppression of categorical and phonological representations of item $u$, $s_C^{(u)}$ and $s_P^{(u)}$, are reset to zero (with the assumption that the length of the intertrial interval and rate of decay of suppression make this a reasonable approximation).

Note that, in general, the real-time decay of suppression and activation might be uncoupled further from contextual change (Chapter 5) by parameterising presentation rates, length of retention interval, etc. (i.e., introducing new parameters in addition to $C_P$, $C_D$, $C_R$ and $C_I$). This is beyond the scope of the present model.

## Extension to Groups of Groups

SEM is readily extendible to any number of subgroupings of a sequence, by assuming that each boundary between groupings can be marked by start and end markers. With $L$ levels of grouping, the positional codes are given by $\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3 ... \boldsymbol{p}_L$, the general context by $\boldsymbol{p}_{L+1}$ and the strength with which token $t$ is cued by the positional cue for response $r$ is:

$$q^{(t)}(r) = \prod_{k=1}^{L+1} \overline{m}(\boldsymbol{p}_k^{(t)}, \boldsymbol{q}_k^{(r)}, d_k, M_k) \qquad \text{Equation 10-16}$$

The positional uncertainty functions for a sequence of $i,j=1..N_P$ positions coded by $k=1..L$ start and end markers is given by:

$$f = f_1 f_2 \ldots f_L \qquad f_k(i,j) = o(\boldsymbol{p}_k^{(i)}, \boldsymbol{p}_k^{(j)}) \qquad \text{Equation 10-17}$$

## Implementational Details

The single-trial and multiple-trial versions of SEM have been written as computer programs in C to run on Unix. Both are available from the author on request (as is the program used to analyse the reports produced by SEM and by subjects), though the above formalism should be sufficient for one to implement their own version.

In addition to specifying parameter values, the single-trial program requires three further arguments, one representing the list, one representing the vocabulary and one representing the set of phonologically confusable items in the vocabulary (each item is represented by a single character). The program outputs two files, one with $N_L$ copies of the specified list and one with the corresponding $N_L$ reports. The multiple-trial program on the other hand reads the $N_L$ lists from a file (often the same lists given to subjects), and outputs a file with one report of each list.

In theory, SEM does not assume a limit on the number of tokens in short-term memory. In practice however, only the most recent tokens can ever be retrieved, assuming continual context drift. The multiple-trial program therefore stores a finite number of tokens, specified by the parameter $N_T$, and functions as a FIFO stack in which the oldest token is overwritten by the newest. In practice, ensuring $N_T$>$4N_P$ tokens is sufficient to give reasonable levels of proactive interference.

Finally, both versions of SEM require an additional “random seed”, which determines the exact values selected from the model’s random generation function (algorithm AS 183 from Applied Statistics). This seed of course has negligible effect the asymptotic behaviour of the model, when $N_L$ is large, but can produce different fits when $N_L$ is small. In all fits herein, this seed was constant at *0*.

## Summary of Fits

A table with the complete set of parameter values for each fit in Chapters 5 and 6 is given below, together with four additional fits illustrating further properties of SEM. Parameter values indicated with a hyphen are irrelevant to a fit (e.g., the value of $P_S$ when no confusable items are specified); parameter values indicated with an asterix vary between simulations within a fit (e.g., to simulate different experimental conditions), and their values are given in the text above the tables.

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

The single-trial version was fitted to the error position curve in the Long condition of Experiment 2 (one simulation). Setting $S_{0,I}$=*1.00*, $S_I$=*0.80*, this fit had three effective free parameters $E_{0,I}$, $E_I$ and $G_C$. Remaining parameters were fixed at *0.00*.

Given the list *12345* and vocabulary *12345*, parameter values were:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_P$ | $N_T$ | $N_V$ | $N_G$ | | | | | $N_L$ |
| *5* | *5* | *5* | *0* | | | | | $10^5$ |
| $S_{0,I}$ | $S_I$ | $E_{0,I}$ | $E_I$ | | $S_{0,G}$ | $S_G$ | $E_{0,G}$ | $E_G$ |
| *1.00* | *0.80* | *0.60* | *0.48* | | - | - | - | - |
| $D_I$ | $D_G$ | | $M_I$ | $M_G$ | | $G_C$ | $G_P$ | $G_G$ |
| *0.00* | - | | *0.00* | - | | *0.14* | *0.00* | *0.00* |
| $T_O$ | $T_G$ | | $P_S$ | $P_D$ | | $A_P$ | | $R_S$ |
| *0.00* | *0.00* | | - | - | | *0.00* | | *0.00* |

**Fit 2. Omissions**

The single-trial version was fitted to transpositions and omissions (including intrusions) in the Long condition of Experiment 2 (one simulation). All parameters were fixed from Fit 1, except the new free parameter $T_O$.

Given the list *12345* and vocabulary *12345*, parameter values were:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_P$ | $N_T$ | $N_V$ | $N_G$ | | | | | $N_L$ |
| *5* | *5* | *5* | *0* | | | | | $10^5$ |
| $S_{0,I}$ | $S_I$ | $E_{0,I}$ | $E_I$ | | $S_{0,G}$ | $S_G$ | $E_{0,G}$ | $E_G$ |
| *1.00* | *0.80* | *0.60* | *0.48* | | - | - | - | - |
| $D_I$ | $D_G$ | | $M_I$ | $M_G$ | | $G_C$ | $G_P$ | $G_G$ |
| *0.00* | - | | *0.00* | - | | *0.14* | *0.00* | *0.00* |
| $T_O$ | $T_G$ | | $P_S$ | $P_D$ | | $A_P$ | | $R_S$ |
| *0.48* | *0.00* | | - | - | | *0.00* | | *0.00* |

**Fit 3. Repetitions**

The single-trial version was fitted to transpositions, omissions and repetitions in the PN condition of Experiment 1 (one simulation). All parameters were fixed from Fit 2, except $N_P$, $N_T$ and $N_V$, for the longer lists, and $G_C$, $T_O$ and $R_S$ as the three free parameters.

Given the list *123456* and vocabulary *123456*, parameter values were:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_P$ | $N_T$ | $N_V$ | $N_G$ | | | | | $N_L$ |
| *6* | *6* | *6* | *0* | | | | | $10^5$ |
| $S_{0,I}$ | $S_I$ | $E_{0,I}$ | $E_I$ | | $S_{0,G}$ | $S_G$ | $E_{0,G}$ | $E_G$ |
| *1.00* | *0.80* | *0.60* | *0.48* | | - | - | - | - |
| $D_I$ | $D_G$ | | $M_I$ | $M_G$ | | $G_C$ | $G_P$ | $G_G$ |
| *0.00* | - | | *0.00* | - | | *0.08* | *0.00* | *0.00* |
| $T_O$ | $T_G$ | | $P_S$ | $P_D$ | | $A_P$ | | $R_S$ |
| *0.32* | *0.00* | | - | - | | *0.00* | | *0.50* |

**Fit 4. Phonological Confusions**

The single-trial version was fitted to transpositions, omissions and confusions in all four conditions of Experiment 1 (four simulations). Parameters were maintained from Fit 3, except the fixed parameters $N_V=12$, $A_P=1.00$, $P_D=0.00$ and three free parameters $G_P$, $T_O$, $P_S$.

Given the lists *RHKYMQ* (condition PN), *BMGQVK* (condition AC), *KGQVMB* (condition AN), and *VBGDPT* (condition PC), a vocabulary of *RHKYMQVBGDPT* and confusable set *VBGDPT,* parameter values were:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_P$ | $N_T$ | $N_V$ | $N_G$ | | | | | $N_L$ |
| 6 | 6 | 12 | 0 | | | | | $10^5$ |
| $S_{0,I}$ | $S_I$ | $E_{0,I}$ | $E_I$ | | $S_{0,G}$ | $S_G$ | $E_{0,G}$ | $E_G$ |
| 1.00 | 0.80 | 0.60 | 0.48 | | - | - | - | - |
| $D_I$ | $D_G$ | | $M_I$ | $M_G$ | | $G_C$ | $G_P$ | $G_G$ |
| 0.00 | - | | 0.00 | - | | 0.08 | 0.30 | 0.00 |
| $T_O$ | $T_G$ | | $P_S$ | $P_D$ | | $A_P$ | | $R_S$ |
| 0.90 | 0.00 | | 0.75 | 0.00 | | 1.00 | | 0.50 |

**Fit 5. List Length, Grouping and Interpositions**

The single-trial version was fitted to the conditions of Experiment 2 (four simulations). Parameters were maintained from Fit 4, except the fixed parameters $S_{0,G}=1.00$, $S_G=0.80$, $N_P$, $N_T$, $N_G$, $N_I$, and the eight free parameters $E_{0,I}$, $E_I$, $E_{0,G}$, $E_G$, $D_I$, $M_I$, $D_G$ and $M_G$.

Parameters $N_P$, $N_T$, $N_G$, $N_I$, $E_{0,I}$, $E_I$ and $D_I$ varied between the ungrouped and grouped conditions. In the ungrouped conditions $N_P$, $N_T$, $N_I(1)$ were equal to the list length, and $N_G=1$, $N_I(2)=N_I(3)=0$, $E_{0,I}=0.60$, $E_I=0.60$, $D_I=0.04$. In the grouped condition, $N_P=N_T=9$, $N_G=3$, $N_I(1)=N_I(2)=N_I(3)=3$, $E_{0,I}=1.00$, $F_I=0.20$, $D_I=0.16$.

Given lists *1234567* (condition U7), *12345678* (condition U8), *123456789* (conditions U9 and G9), and a vocabulary of *0123456789*, parameter values were:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_P$ | $N_T$ | $N_V$ | $N_G$ | $N_I(1)$ | $N_I(2)$ | $N_I(3)$ | | $N_L$ |
| * | * | *10* | * | * | * | * | | $10^5$ |
| $S_{0,I}$ | $S_I$ | $E_{0,I}$ | $E_I$ | | $S_{0,G}$ | $S_G$ | $E_{0,G}$ | $E_G$ |
| *1.00* | *0.80* | * | * | | *1.00* | *0.80* | *0.60* | *0.80* |
| $D_I$ | $D_G$ | | $M_I$ | $M_G$ | | $G_C$ | $G_P$ | $G_G$ |
| * | *0.08* | | *0.40* | *0.85* | | *0.08* | *0.30* | *0.00* |
| $T_O$ | $T_G$ | | $P_S$ | $P_D$ | | $A_P$ | | $R_S$ |
| *0.90* | *0.00* | | *0.75* | *0.00* | | *1.00* | | *0.50* |

**Fit 6. Intertrial Interval and Protrusions**

The multiple-trial version was fitted to both conditions of Experiment 3 (two simulations). All parameter values were maintained from Fit 5, except the new fixed parameters $E_{0,C}=1.00$, $C_P=C_R=1$, $C_D=3$, and the five free parameters $E_C$, $R_P$, $C_A$, $G_C$ and $T_O$.

The parameter $C_I$ varied between conditions (according to the experimental design). In the Long condition $C_I=20$; in the Short condition $C_I=2$.

Given 100,000 copies of the lists given to subjects, and a vocabulary of *YGVPWHCK MLSBFT* (representing words used; none of which were confusable), parameter values were:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_P$ | $N_T$ | $N_V$ | $N_G$ | $N_I(1)$ | | | $N_M$ | $N_L$ |
| *5* | *20* | *14* | *1* | *5* | | | - | $10^5$ |
| $S_{0,I}$ | $S_I$ | $E_{0,I}$ | $E_I$ | | $S_{0,G}$ | $S_G$ | $E_{0,G}$ | $E_G$ |
| *1.00* | *0.80* | *0.60* | *0.60* | | *1.00* | *0.80* | *0.60* | *0.80* |
| $E_{0,C}$ | $E_C$ | | | $C_P$ | $C_D$ | $C_R$ | $C_I$ | $C_A$ |
| *1.00* | *0.98* | | | *1* | *3* | *1* | * | *20* |
| $D_I$ | $D_G$ | | $M_I$ | $M_G$ | | $G_C$ | $G_P$ | $G_G$ |
| *0.04* | *0.08* | | *0.40* | *0.85* | | *0.10* | *0.30* | *0.00* |
| $T_O$ | $T_G$ | $T_U$ | $P_S$ | $P_D$ | | $A_P$ | $R_P$ | $R_S$ |
| *0.70* | *0.00* | *0.00* | *0.75* | *0.00* | | *1.00* | *0.05* | *0.50* |

**Fit 7. Interpositions in Variable Groups**

The multiple-trial version was fitted to all three conditions of Experiment 4, with and without guesses (six simulations). Parameters were maintained from Fit 5, except the fixed parameters $N_G$, $N_I$, $G_G=0.30$, $T_O=0.00$, $C_P=C_R=1$, $C_D=C_I=0$, and the five free parameters $D_G$, $M_G$, $G_C$, $T_G$ and $T_U$.

The parameters $N_G$, $N_I$, $E_{0,I}$ and $E_I$ varied between conditions (according to experimental design and Fit 5). In the Ungrouped condition $N_G=1$, $N_I(1)=7$, $E_{0,I}=0.60$, $E_I=0.60$; in the Grouped 3-4 condition, $N_G=2$, $N_I(1)=3$, $N_I(2)=4$; $E_{0,I}=1.00$, $E_I=0.20$, and in the Grouped 4-3 condition, $N_G=2$, $N_I(1)=4$, $N_I(2)=3$, $E_{0,I}=1.00$, $E_I=0.20$. The parameter $T_U$ varied to simulate the removal of guesses. with guesses, $T_U=0.00$; without, $T_U=1.10$.

Given 100,000 copies of the lists given to subjects, and a vocabulary of *GVCLBFT* (representing words used; none of which were confusable), parameter values were:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_P$ | $N_T$ | $N_V$ | $N_G$ | $N_I(1)$ | $N_I(2)$ | | $N_M$ | $N_L$ |
| 7 | 28 | 7 | * | * | * | | - | $10^5$ |
| $S_{0,I}$ | $S_I$ | $E_{0,I}$ | $E_I$ | | $S_{0,G}$ | $S_G$ | $E_{0,G}$ | $E_G$ |
| 1.00 | 0.80 | * | * | | 1.00 | 0.80 | 0.60 | 0.80 |
| $E_{0,C}$ | $E_C$ | | | $C_P$ | $C_D$ | $C_R$ | $C_I$ | $C_A$ |
| 1.00 | 0.98 | | | 1 | 0 | 1 | 0 | 20 |
| $D_I$ | $D_G$ | | $M_I$ | $M_G$ | | $G_C$ | $G_P$ | $G_G$ |
| 0.04 | 0.10 | | 0.40 | 0.95 | | 0.06 | 0.30 | 0.30 |
| $T_O$ | $T_G$ | $T_U$ | $P_S$ | $P_D$ | | $A_P$ | $R_P$ | $R_S$ |
| 0.00 | 0.90 | * | 0.75 | 0.00 | | 1.00 | 0.05 | 0.50 |

**Fit 8. Protrusions in Variable Lists**

The multiple-trial version was fitted to both conditions of Experiment 5, with and without guesses (four simulations). Parameters were maintained from Fit 7, except the one free parameter $G_C$.

No parameters changed across conditions, except for $N_P$ which depended on the list length. Note that the value of $N_M$ was only relevant to the Variable condition, were list lengths varied unpredictably. This was fixed as $N_M$=5, reflecting in the minimum expected list length in that condition. The parameter $T_U$ varied to simulate the removal of guesses as in Fit 7: with guesses, $T_U=0.00$; without, $T_U=1.10$.

Given 100,000 copies of lists given to subjects, and a vocabulary of *YGVPWHCK MLSBFT* (representing words used; none of which were confusable), parameter values were:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_P$ | $N_T$ | $N_V$ | $N_G$ | $N_I(1)$ | | | $N_M$ | $N_L$ |
| * | *24* | *14* | *1* | * | | | *5* | $10^5$ |
| $S_{0,I}$ | $S_I$ | $E_{0,I}$ | $E_I$ | | $S_{0,G}$ | $S_G$ | $E_{0,G}$ | $E_G$ |
| *1.00* | *0.80* | *0.60* | *0.60* | | *1.00* | *0.80* | *0.60* | *0.80* |
| $E_{0,C}$ | $E_C$ | | | $C_P$ | $C_D$ | $C_R$ | $C_I$ | $C_A$ |
| *1.00* | *0.98* | | | *1* | *0* | *1* | *0* | *20* |
| $D_I$ | $D_G$ | | $M_I$ | $M_G$ | | $G_C$ | $G_P$ | $G_G$ |
| *0.04* | *0.10* | | *0.40* | *0.95* | | *0.01* | *0.30* | *0.30* |
| $T_O$ | $T_G$ | $T_U$ | $P_S$ | $P_D$ | | $A_P$ | $R_P$ | $R_S$ |
| *0.00* | *0.90* | * | *0.75* | *0.00* | | *1.00* | *0.05* | *0.50* |

## Additional Fits not Reported in Chapters 5 and 6

### Fit 9. Retention Interval and Phonological Confusions

This was a qualitative fit of the multiple-trial version showing an interaction between phonological similarity and delay for each list-type in Experiment 1 (sixteen simulations). Parameters were fixed from Fit 8, except $N_P$=$N_I$=6, $N_V=12$, $T_U=0.00$ and $T_O=0.70$ (i.e., uncertain responses were included but omissions were added). The value of $C_D$ varied from *0* to *5* to *10* to *20,* to simulate the length of a filled retention interval (in seconds).

Given 100,000 copies of the lists *RHKYMQ* (condition PN), *BMGQVK* (condition AC), *KGQVMB* (condition AN), *VBGDPT* (condition PC), a vocabulary *RHKYMQVBGDPT* and confusable set *VBGDPT,* parameter values were:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_P$ | $N_T$ | $N_V$ | $N_G$ | $N_I(1)$ | | | $N_M$ | $N_L$ |
| 6 | 24 | 24 | 1 | 6 | | | - | $10^5$ |
| $S_{0,I}$ | $S_I$ | $E_{0,I}$ | $E_I$ | | $S_{0,G}$ | $S_G$ | $E_{0,G}$ | $E_G$ |
| 1.00 | 0.80 | 0.60 | 0.60 | | 1.00 | 0.80 | 0.60 | 0.80 |
| $E_{0,C}$ | $E_C$ | | | $C_P$ | $C_D$ | $C_R$ | $C_I$ | $C_A$ |
| 1.00 | 0.98 | | | 1 | * | 1 | 0 | 20 |
| $D_I$ | $D_G$ | | $M_I$ | $M_G$ | | $G_C$ | $G_P$ | $G_G$ |
| 0.04 | 0.10 | | 0.40 | 0.95 | | 0.01 | 0.30 | 0.30 |
| $T_O$ | $T_G$ | $T_U$ | $P_S$ | $P_D$ | | $A_P$ | $R_P$ | $R_S$ |
| 0.70 | 0.90 | 0.00 | 0.75 | 0.00 | | 1.00 | 0.05 | 0.50 |

The sawteeth shape of error position curves for alternating lists disappeared as the retention interval increased (upper panel of Figure A-1). In other words, the phonological similarity effect disappeared and confusions fell to chance levels. Indeed, when $C_D=20$, performance on PC lists was almost identical to PN lists. This arises because the phonological activations decay to zero as the delay increases. This is consistent with previous results, but has not been demonstrated empirically in the striking fashion shown here with alternating lists.

The retention interval also affected the pattern of errors. As the phonological activations decayed, the incidence of omissions and intrusions increased, faster than that of transpositions (lower panel of Figure A-1). No study has shown this specific pattern: The increase in omissions and intrusions (and decrease in confusions) are predictions of SEM.

Unlike the single-trial version of SEM, the feedback of responses in the multiple-trial version means that errors on confusable items do impair recall of subsequent nonconfusable items to a small extent (there were small differences of about 5% between nonconfusables in alternating and nonconfusable curves). This is not necessarily a problem, given that such a small effect was suggested by the third meta-analysis of Chapter 4.

Finally, the proportion of intrusions that were protrusions decreased as retention interval increases, reflecting the increasing effects of guessing (see Fit 6 in Chapter 5).

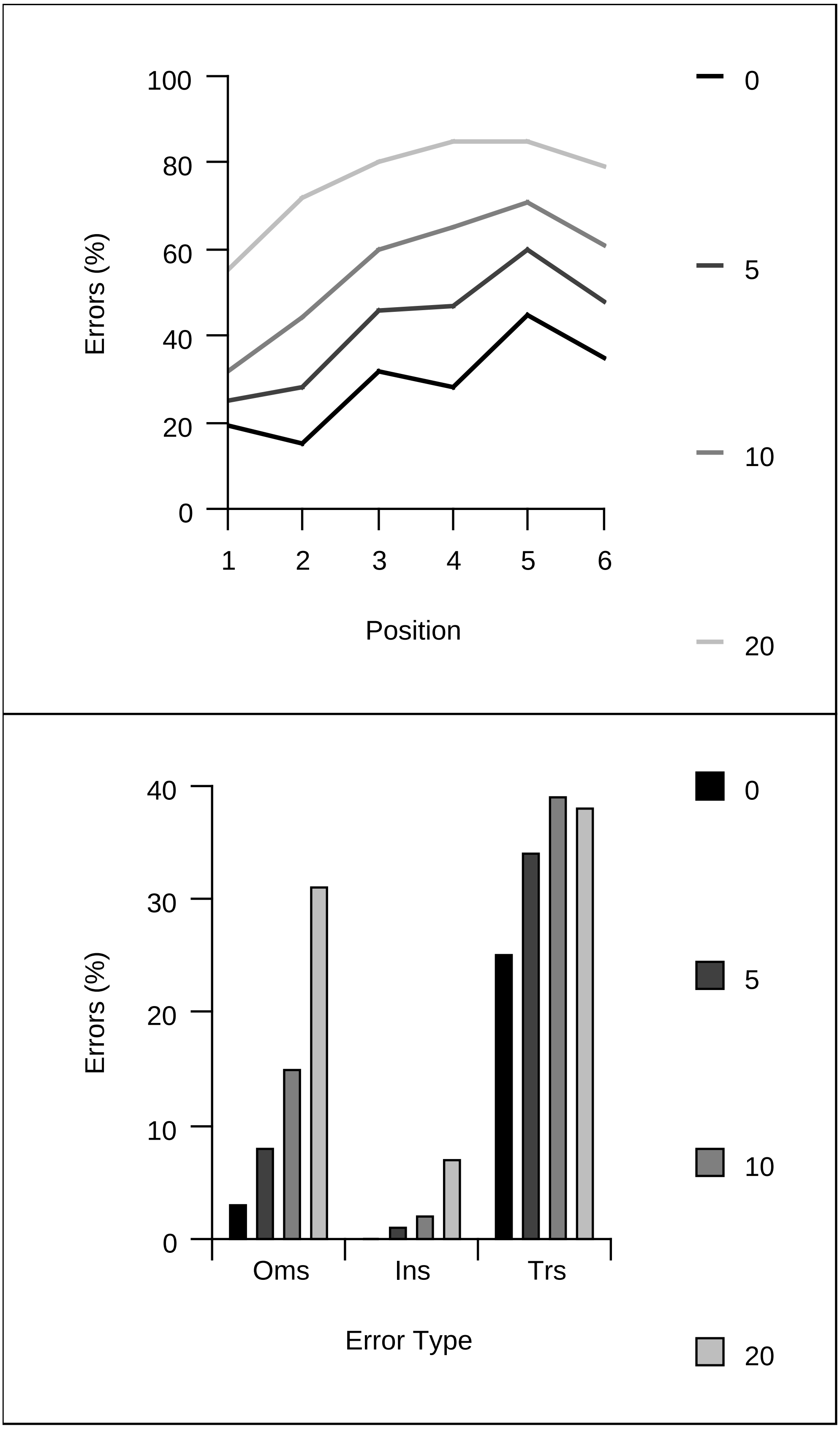



Figure A-1: Effects of retention interval on error position curves (upper panel) and error types (lower panel) for condition AC in Fit 9.
(Numbers refer to parameter $C_D$; Oms=omissions, Ins=intrusions and Trs=transpositions.)

**Fit 10. List Length and Span**

This was a qualitative fit of the multiple-trial version to list length effects (eight simulations). Parameters were fixed from Fit 9, except $C_D=0$, $G_C=0.06$, $N_T=36$, $N_T=10$, and $N_P$, which increased from 2-9, reflecting increases in list length.

Given 10,000 copies of lists drawn from *0123456789*, parameter values were:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_P$ | $N_T$ | $N_V$ | $N_G$ | $N_I(1)$ | | | $N_M$ | $N_L$ |
| * | 36 | 10 | 1 | * | | | - | $10^4$ |
| $S_{0,I}$ | $S_I$ | $E_{0,I}$ | $E_I$ | | $S_{0,G}$ | $S_G$ | $E_{0,G}$ | $E_G$ |
| 1.00 | 0.80 | 0.60 | 0.60 | | 1.00 | 0.80 | 0.60 | 0.80 |
| $E_{0,C}$ | $E_C$ | | | $C_P$ | $C_D$ | $C_R$ | $C_I$ | $C_A$ |
| 1.00 | 0.98 | | | 1 | 0 | 1 | 0 | 20 |
| $D_I$ | $D_G$ | | $M_I$ | $M_G$ | | $G_C$ | $G_P$ | $G_G$ |
| 0.04 | 0.10 | | 0.40 | 0.95 | | 0.06 | 0.30 | 0.30 |
| $T_O$ | $T_G$ | $T_U$ | $P_S$ | $P_D$ | | $A_P$ | $R_P$ | $R_S$ |
| 0.70 | 0.90 | 0.00 | 0.75 | 0.00 | | 1.00 | 0.05 | 0.50 |

Longer lists increased errors on all positions, including the first (upper panel of Figure A-2). Note that error position curves for lists of more than seven items may not resemble those found empirically, because of the tendency for subjects to spontaneously group such lists (Experiment 2). Importantly, SEM produced the characteristic inverse-S shaped curves of lists correct against list length (lower panel of Figure A-2).

**Fit 11. Word Length and Articulation Rate**

This was a qualitative fit of the multiple-trial versions to effects of word-length and articulation rate (forty simulations). Parameters were fixed from Fit 9, except for a factorial combination of $C_P=C_R=1...5$, reflecting word-length, and $N_P=2...9$, reflecting list length. The values of $C_P$ and $C_R$ were greater for long words than short words, with the assumption that they allow a greater opportunity for decay during presentation (i.e., ignoring rehearsal during presentation for simplicity; Chapter 5).

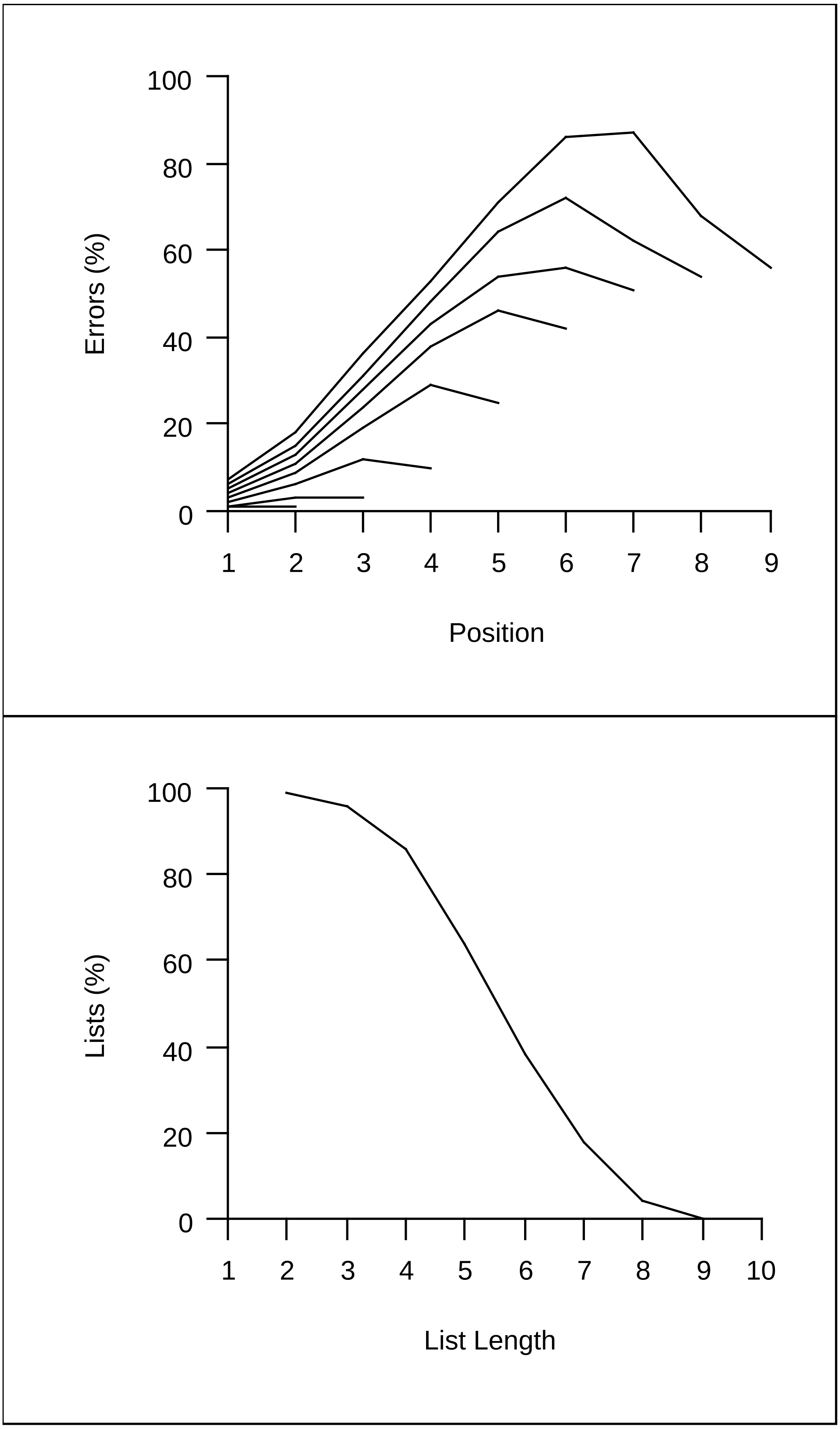



Figure A-2: Effect of list length on error position curves (upper panel) and lists correct (lower panel) in Fit 10.

Given 10,000 copies of lists drawn from *0123456789*, parameter values were:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_P$ | $N_T$ | $N_V$ | $N_G$ | $N_I(1)$ | | | $N_M$ | $N_L$ |
| * | *36* | *10* | *1* | * | | | - | $10^4$ |
| $S_{0,I}$ | $S_I$ | $E_{0,I}$ | $E_I$ | | $S_{0,G}$ | $S_G$ | $E_{0,G}$ | $E_G$ |
| *1.00* | *0.80* | *0.60* | *0.60* | | *1.00* | *0.80* | *0.60* | *0.80* |
| $E_{0,C}$ | $E_C$ | | | $C_P$ | $C_D$ | $C_R$ | $C_I$ | $C_A$ |
| *1.00* | *0.98* | | | * | *0* | * | *0* | *20* |
| $D_I$ | $D_G$ | | $M_I$ | $M_G$ | | $G_C$ | $G_P$ | $G_G$ |
| *0.04* | *0.10* | | *0.40* | *0.95* | | *0.06* | *0.30* | *0.30* |
| $T_O$ | $T_G$ | $T_U$ | $P_S$ | $P_D$ | | $A_P$ | $R_P$ | $R_S$ |
| *0.70* | *0.90* | *0.00* | *0.75* | *0.00* | | *1.00* | *0.05* | *0.50* |

Longer words decreased spans (the 50% correct level in upper panel of Figure A-3). Closer analysis of errors showed that longer words increased both transpositions and omissions, particularly towards the end of recall, in agreement with unpublished data (Page & Norris, 1996a).

The parameters $C_P=C_R=C$ were assumed to be related to the number of syllables in words, *W*, by the formula $W=(C+1)/2$ (i.e., an extra syllable corresponded to an increase of 2 in $C_P$ and $C_R$). They were also converted into speeded articulation rate, *R*, by the formula $R=3.1-0.6C$. The latter formula gives a good approximation to the rates determined by Page & Norris (1996b) in their fit to Hulme et al (1991). The resulting relationship between span and articulation rate is shown in the lower panel of Figure A-3. The relationship is near-linear ($R^2=.93$), with an approximate slope of *0.88* and intercept of *3.15*, in reasonable agreement with the data of Hulme et al (1991). The relationship between span and rate departs most from linearity for the slowest and fastest rates. It is noteworthy that these rates are at or beyond the limits normally achieved experimentally, and thus the quadratic component of the span-rate curve reflects a testable prediction of SEM (though other relationships between span and rate may be possible with different parameter values).

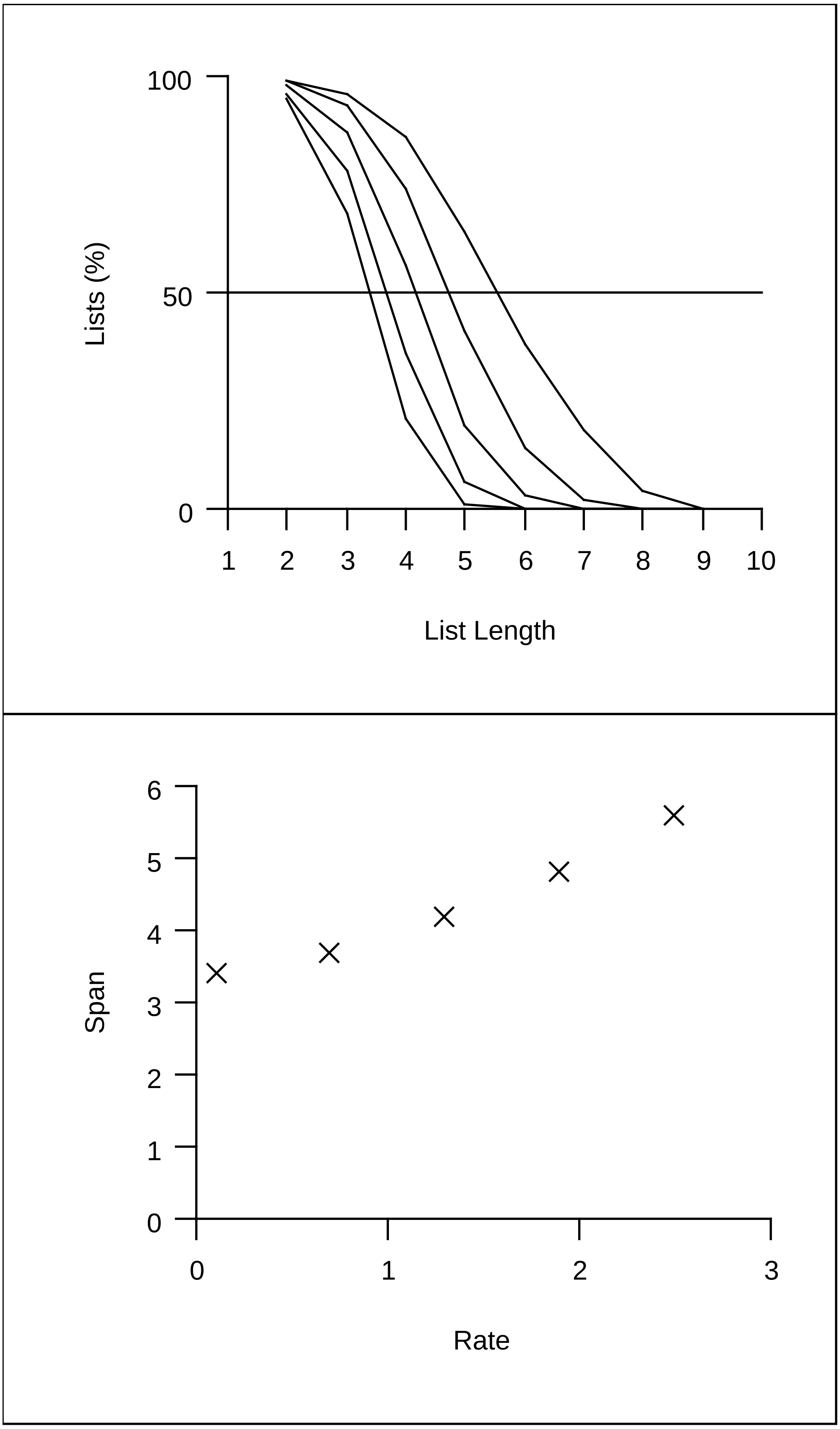



Figure A-3: Effect of word-length on lists correct (upper panel; curves to the left represent longer words) and effect of articulation rate on span (lower panel) in Fit 11.

**Fit 12. Modality and Suffix Effects**

This was a qualitative fit of the multiple-trial version to modality and suffix effects in recall of eight items (three simulations). Parameters were fixed from Fit 11, except for the values $N_P=8$, $N_T=32$, $E_I=0.40$ and value of $E_{0,I}$, which changed between conditions.

The parameter $E_{0,I}=0.60$ for auditory lists and $E_{0,I}=0.20$ for visual lists. The latter reflected the assumption that the end marker for auditory lists is stronger than for visual lists. The value $E_{0,I}=0.24$ $(=0.60.E_I)$ for auditory lists with a suffix reflected the assumption that the suffix was marked in the last position, rather than the last item (i.e., the suffix was unavoidably grouped together with the list items, so that position was coded as if there were nine positions, though only the eight list items competed for recall). This fit did not take into account the additional delay caused by a suffix, though this could be modelled simply by increasing $C_R$, which would affect mainly middle items (Baddeley & Hull, 1979).

Given 100,000 copies of lists drawn from *0123456789*, parameter values were:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_P$ | $N_T$ | $N_V$ | $N_G$ | $N_I(1)$ | | | $N_M$ | $N_L$ |
| *(8)* | *32* | *10* | *1* | *8* | | | - | $10^5$ |
| $S_{0,I}$ | $S_I$ | $E_{0,I}$ | $E_I$ | | $S_{0,G}$ | $S_G$ | $E_{0,G}$ | $E_G$ |
| *1.00* | *0.80* | * | *0.40* | | *1.00* | *0.80* | *0.60* | *0.80* |
| $E_{0,C}$ | $E_C$ | | | $C_P$ | $C_D$ | $C_R$ | $C_I$ | $C_A$ |
| *1.00* | *0.98* | | | *1* | *0* | *1* | *0* | *20* |
| $D_I$ | $D_G$ | | $M_I$ | $M_G$ | | $G_C$ | $G_P$ | $G_G$ |
| *0.04* | *0.10* | | *0.40* | *0.95* | | *0.06* | *0.30* | *0.30* |
| $T_O$ | $T_G$ | $T_U$ | $P_S$ | $P_D$ | | $A_P$ | $R_P$ | $R_S$ |
| *0.70* | *0.90* | *0.00* | *0.75* | *0.00* | | *1.00* | *0.05* | *0.50* |

The modality advantage for auditory presentation extended over the last two or three positions, but was removed by an additional suffix (Figure A-4). The auditory advantage reflected a decrease in both omissions and transpositions, in agreement with unpublished data (Page & Norris, 1996a). Note that the exact values of $E_{0,I}$ and $E_I$ needed to produce this pattern do depend on the values of other parameters (such as the noise $G_C$). The purpose of this fit is simply to show how SEM's assumption of marking the end of a list can in principle reproduce modality and suffix effects; further work is needed on how exactly auditory presentation or additional suffixes affect such marking.

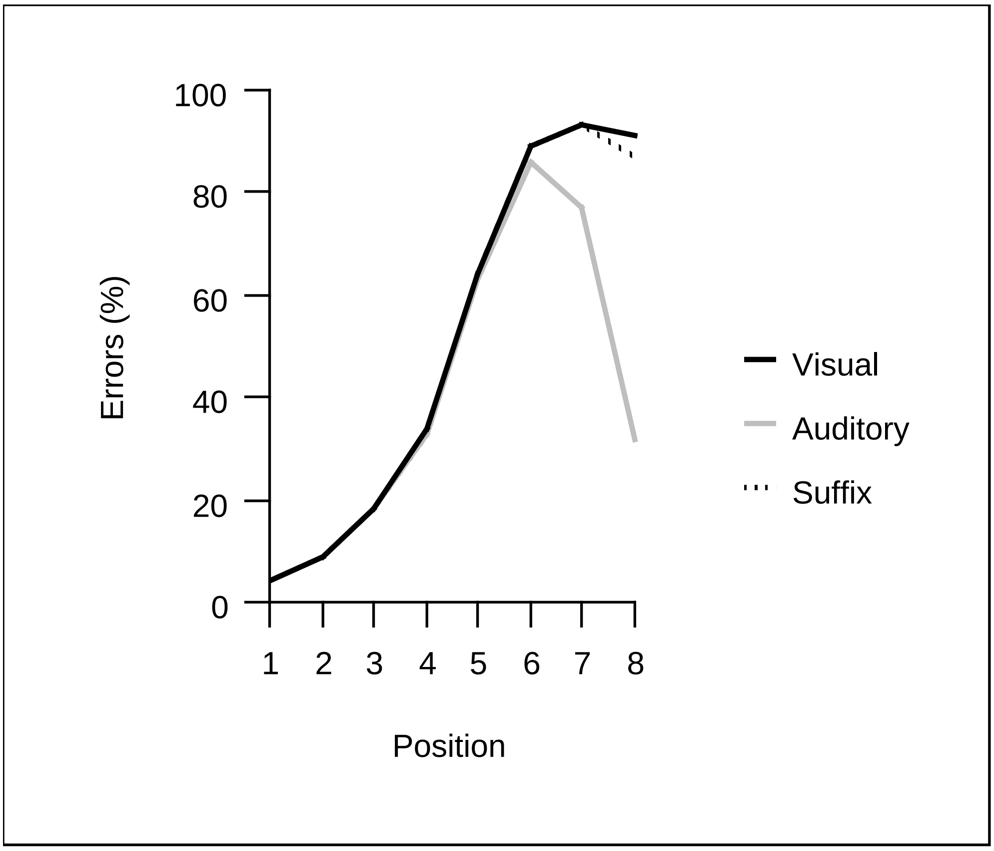



Figure A-4: Modality and suffix effects in Fit 12.

# Glossary of Terms in SEM

*Parameters*

| | |
|---|---|
| $N_T$ | Number of tokens in short-term memory |
| $N_V$ | Number of items in vocabulary |
| $N_L$ | Number of lists |
| $N_P(l)$ | Number of positions in list *l* (list length) |
| $N_G(l)$ | Number of groups in list *l* |
| $N_I(g)$ | Number of items in group *g* (group size) |
| $N_M$ | Minimum expected list length |
| $S_{0,I}$ | Initial value of start marker for item position in group |
| $S_I$ | Decay rate of start marker for item position in group |
| $E_{0,I}$ | Initial value of end marker for item position in group |
| $E_I$ | Decay rate of end marker for item position in group |
| $S_{0,G}$ | Initial value of start marker for group position in list |
| $S_G$ | Rate of change of start marker for group position in list |
| $E_{0,G}$ | Initial value of end marker for group position in list |
| $E_G$ | Rate of change of end marker for group position in list |
| $E_{0,C}$ | Value for current general context |
| $E_C$ | Rate of change of general context |
| $D_I$ | SD of Gaussian noise in item position codes |
| $D_G$ | SD of Gaussian noise in group position codes |

| | |
|---|---|
| $M_I$ | Positional Match Criterion for item position codes |
| $M_G$ | Positional Match Criterion for group position codes |
| | |
| $G_C$ | SD of Gaussian noise in categorical competition |
| $G_P$ | SD of Gaussian noise in phonological competition |
| $G_G$ | SD of Gaussian noise in guessing |
| | |
| $T_O$ | Omission threshold |
| $T_G$ | Guessing threshold |
| $T_U$ | Uncertainty threshold |
| | |
| $P_S$ | Phonological similarity between similar items |
| $P_D$ | Phonological similarity between dissimilar items |
| | |
| $A_P$ | Baseline Activation of phonological representations |
| $R_P$ | Rate with which phonological activation decays |
| $R_S$ | Rate with which suppression decays |
| | |
| $C_P$ | Effective presentation rate (ignoring rehearsal) |
| $C_D$ | Length of filled delay during retention interval |
| $C_R$ | Effective recall rate (ignoring rehearsal) |
| $C_I$ | Length of intertrial interval |
| $C_A$ | Contextual/Attentional shift during intertrial interval |

*Indices*

| | |
|---|---|
| $l=1..N_L$ | List number |
| $p=1..N_P(l)$ | Item (input) position in list *l* |
| $r=1..N_P(l)$ | Response (output) position in recall of list *l* |

| | |
|---|---|
| $i=1..N_I(g)$ | Item position in group *g* |
| $g=1..N_G(l)$ | Group position in list *l* |
| $c=0..inf$ | Number of episodes for contextual change/decay |
| $u=1..N_V$ | Index for categorical representation of item |
| $v=1..N_V$ | Index for phonological representation of item |

*Variables*

| | |
|---|---|
| $x(i)$ | Strength of start marker at position *i* |
| $y(i)$ | Strength of end marker at position *i* |
| $\boldsymbol{p}_I^{(t)}$ | Positional code for item position in group for token *t* |
| $\boldsymbol{p}_G^{(t)}$ | Positional code for group position in list for token *t* |
| $\boldsymbol{p}_C^{(t)}$ | Positional code for general context for token *t* |
| $\overline{m}(\boldsymbol{p},\boldsymbol{q})$ | Thresholded match between positional codes ***p*** and ***q*** |
| $m(\boldsymbol{p},\boldsymbol{q})$ | Noisy match between positional codes ***p*** and ***q*** |
| $o(\boldsymbol{p},\boldsymbol{q})$ | Overlap between positional codes ***p*** and ***q*** |
| $f(i,j)$ | Positional uncertainty functions over positions *i*,*j* |
| $q^{(t)}(r)$ | Cued strength of token t for response r |
| $c^{(u)}$ | Competition strength of item *u* |
| $a^{(u)}$ | Activation of item *u* |
| $s^{(u)}$ | Suppression of item *u* |
| $d$ | Noise in positional code |
| $n$ | Noise in competition |
| $p(u,v)$ | Phonological similarity between item u and item v |