The problem of serial order in visuospatial short-term memory

Mark John Hurlstone

Thesis submitted for the degree of Doctor of Philosophy

Department of Psychology

University of York

May 2010

Abstract

How do we remember the order of a novel sequence of items? Much research has examined how people remember sequences of verbal stimuli (e.g., digits in a phone number), and several mechanisms of serial order have been proposed to underlie memory for such sequences. Less research has examined how people remember the order of sequences of visuospatial stimuli (e.g., a series of spatial locations), and the mechanisms of serial order underlying such sequences remain unspecified. This thesis explores the extent to which memory for sequences of visuospatial stimuli is explicable in term of mechanisms proposed to underlie memory for verbal sequences.

Contemporary models of verbal short-term memory represent serial order either by: (1) using a *competitive queuing* sequence planning and control mechanism, by (2) *position marking*, by (3) a *primacy gradient* of activation, by (4) incorporating *response suppression*, and by (5) implementing *output interference*, or through some combination of these mechanisms. Empirical evidence suggests that all five mechanisms must coexist in any adequate model of serial order memory for verbal sequences.

In this thesis, I argue that extant data indicating functional similarities between verbal and visuospatial serial order memory support the idea that visuospatial sequences are planned and controlled using a competitive queuing mechanism. However, direct evidence for the role of the four remaining mechanisms of serial order in visuospatial short-term memory is currently lacking. I present a series of twelve experiments examining memory for visuospatial sequences, combined with computational modelling work, which sought direct evidence for the role (or lack thereof) of the different mechanisms of serial order.

The outcomes of the experiments and computational modelling work suggest that the serial order of a visuospatial sequence is represented by a competitive queuing system, equipped with a primacy gradient, positional markers, and response suppression. The results therefore buttress the notion that verbal and visuospatial short-term memory rely on some common mechanisms for the representation and generation of serial order.

Contents

List of figures

List of tables

Acknowledgments

This research was supported by an Economic and Social Research Council (ESRC) 1+3 research studentship.

I am extremely grateful to my supervisors Graham Hitch and Alan Baddeley for their guidance and support, and for providing me with the freedom to choose a research area and to pursue it. Words can not express how fortunate and grateful I am to have benefited from the tutelage of these eminent old hands. I am also grateful to Tom Hartley, the third member of my research committee, for numerous informative and stimulating discussions.

I also owe a great debt of gratitude to Simon Farrell from the University of Bristol who agreed to host a four month study visit so that I could learn about aspects of cognitive modelling. The skills acquired from this visit provided the basis for the computational modelling work that forms an extensive component of this thesis. Thanks additionally go to the Experimental Psychology Society (EPS) of the United Kingdom for funding this important learning opportunity via their study visit grant scheme.

Last, but certainly not least, I thank Su Yin Min for her unfaltering love, encouragement, and support, particularly during the strenuous period of writing up. This thesis is affectionately dedicated to her.

Declaration

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which has been accepted for the award of any other degree or diploma at the University of York or any other educational institution. I also declare that the intellectual content of this thesis is the product of my own work and further that this thesis contains nothing that is the outcome of work performed in collaboration.

1

An overview of the literature and thesis

The problem of serial order

In a seminal article, Karl Lashley (1951) drew attention to the fact that a fundamental requirement for most if not all of behaviour is the ability to process serial order information. The capacity is central to linguistic behaviours ranging from speech perception and generation to vocabulary acquisition and spelling, as well as, non-linguistic behaviours ranging from motor control, planning, and goal-directed action. How serial order information is processed across these and other behavioural domains was dubbed by Lashley (1951) as the *problem of serial order in behaviour*. This thesis is concerned with one instantiation of this problem, the problem of serial order in short-term memory – specifically short-term memory for nonverbal, visuospatial information.

Applied to short-term memory, the problem of serial order is to specify the mechanisms underpinning the encoding and retrieval of novel sequences of briefly presented items. This problem has been studied extensively in verbal short-term memory using the task of *serial recall* in which participants are given short sequences of familiar verbal items that they must subsequently recall in order. Much of this research has been interpreted within the framework of the working memory model of Baddeley and Hitch (1974; see also Baddeley, 1986; Baddeley, 2000), which comprises (amongst other components) a subsystem for the retention of verbal information known as the *phonological loop*, complemented by a subsystem for the retention of visual and spatial information known as the *visuospatial sketchpad*. Although the phonological loop has been successful in explaining a wealth of serial recall data at a qualitative level, as noted by Burgess and Hitch (1992), it lacks any detailed mechanism(s) for the retention of serial order.

In recent years, this omission has served as the stimulus for the development of a wealth of sophisticated computational theories of verbal short-term memory, which explain detailed aspects of the serial recall data at a quantitative level, using well-specified mechanisms for representing serial order. Some of these theories have been couched within the concept of the phonological loop theory (e.g., Burgess & Hitch, 1992, 1999, 2006; Page & Norris, 1998), whilst others have been framed within alternative theoretical approaches (e.g., Botvinick & Plaut, 2006; Brown, Neath, & Chater, 2007; Brown, Preece, & Hulme, 2000; Farrell & Lewandowsky, 2002; Grossberg & Pearson, 2008; Henson, 1998a; Lewandowsky & Farrell, 2008).

Despite differences between theories, some theoretical convergence is apparent, and a number of basic mechanisms for representing serial order have been identified that are widely employed across models. These mechanisms (delineated later in this chapter) include a *competitive queuing* sequence planning and control mechanism, a *primacy gradient* of activation levels, associations between items and temporal, absolute, or relative representations of their sequence positions through *position marking*, and the temporary inhibition of recalled items through *response suppression*. Although computational theories of serial recall are committed to different combinations of these mechanisms, they are each buttressed by direct empirical precedents from the serial recall literature (reviewed later in this chapter).

Like its phonological loop counterpart, the visuospatial sketchpad lacks any detailed mechanism(s) for the retention of serial order. However, in contrast to the wealth of data and theoretical progress relating to verbal short-term memory, there has been comparatively less research exploring the problem of serial order in the visuo-spatial domain. This is perhaps because memory for serial order appears at a cursory glance to be a more obvious requirement for language, given its inherently temporal structure. However, it is also fundamental in the visuospatial domain, most notably in the acquisition of new motor skills, which are often learned by observing and imitating the actions of others. This clearly depends upon some intermediate memory for the serial order of the to-be-performed actions.

In recognition of this shortcoming, a steady wave of recent behavioural studies has examined short-term memory for serial order employing visuospatial stimuli. These studies have shown that short-term memory for sequences of spatial locations (Farrand & Jones, 1996; Farrand, Parmentier,

& Jones, 2001; Smyth & Scholey, 1994, 1996; Jones, Farrand, Stuart, & Morris, 1995), novel visual patterns (Avons, 1998, Avons & Mason, 1999; Smyth, Hay, Hitch, & Horton, 2005; Ward, Avons, & Melling, 2005), and connected linear movements (Agam, Bullock, & Sekuler, 2005; Agam, Galperin, Gold, Sekuler, 2007) exhibit many characteristics reminiscent of those observed with verbal stimuli.

These functional similarities notwithstanding, theoretical accounts of how serial order information is processed in visuospatial short-term memory are currently lacking. Consequently, the mechanisms of serial order in this domain remain unspecified. The purpose of this thesis is to explore the extent to which the commonalities between the two domains can be explained by recourse to the hypothesis that mechanisms of serial order identified as contributing to verbal serial memory play a similar role in visuospatial serial memory. It is of course possible that there are some distinct mechanisms underpinning the representation and generation of serial order in the two domains, and it would be premature to rule out this possibility. However, given the existence of a common set of behavioural features, it is clearly more parsimonious to assume that at least some core sequencing mechanisms exist which apply to both domains. If this is indeed the case then it would not necessarily compromise the assumption of distinct verbal and visuospatial short-term memory subsystems, as specified by the working memory model, but might suggest instead that the problem of serial order has been resolved in similar ways across the two short-term memory systems.

The rest of this chapter is organised as follows. I begin by reviewing the major empirical constraints of verbal serial memory, before describing and evaluating approaches and mechanisms for serial order generation with reference to their ability to meet those constraints. This analysis identifies a combination of five explanatory mechanisms that must coexist in any adequate model of verbal serial memory. I then describe a wave of recent studies that have successfully replicated many of the empirical constraints listed initially employing sequences of visual and spatial items as stimuli. I conclude that the functional similarities are consistent with the hypothesis that at least some of the explanatory mechanisms of verbal serial memory contribute to the representation and generation of serial order in the visuospatial domain. However, the absence of direct evidence for specific explanatory mechanisms prevents the identification of a preferred combination of those mechanisms. The solution pursued in this thesis is to examine the extent to which empirical constraints identified as unique signatures of specific explanatory mechanisms generalise to the study of visuospatial sequences.

Verbal serial memory: Data

This section prevents an overview of verbal serial memory phenomenon documented employing explicit order processing tasks in which novel sequences of verbal items (e.g., consonants, digits, words) must be reported in order immediately (or shortly) following presentation. These tasks include the dominant task of *serial recall* in which participants must recall sequence items in their presentation order without any cues as to the items themselves, and the related task of *serial reconstruction* in which sequence items are re-presented at recall in a random arrangement and participants must sort the items back into their presentation order.

These tasks (serial recall in particular) have been employed in hundreds of studies of verbal short-term memory, generating a wealth of empirical observations that can serve to elucidate the mechanisms underpinning the storage and retrieval of novel sequences of verbal items. However, the sheer volume of available data prevents a description of all the major findings documented. Instead, this section summarizes a sub-set of constraints that are particularly relevant to the problem of serial order, and are sufficiently well replicated that they can be considered benchmarks of the field. These constitute the core explicanda for theories of verbal short-term memory for serial order. Note that in general, the constraints outlined below apply whether the to-be-remembered items are presented visually (as in most studies) or auditorily.

1. Serial position curves

The serial position curve plots recall accuracy as a function of the serial position of items and is characterised by two canonical effects. First, there is a sharp monotonic decrease in recall accuracy extending from the first position onwards, which is known as the *primacy effect*. Second, for visually presented sequences there is a small upturn in performance for the final serial position, which is known as the *recency effect*. The extent and magnitude of the recency effect is greater under auditory sequence presentation.

Serial position curves can also be plotted using mean response latency as the dependent measure. Recall timing studies (Anderson, Bothel, Lebierre, & Matessa, 1998; Farrell & Lewandowsky, 2004; Maybery, Parmentier, & Jones, 2002; Thomas, Milner, & Haberlandt, 2003) have shown that the latency to select the first item in the sequence (measured from recall onset) is considerably longer than the inter-response latencies for subsequent responses. When the latency for the first output position is ignored the latency serial position function exhibits an inverted U shaped trend.

2. Effects of sequence length

Serial recall accuracy deteriorates as the length of the to-be-remembered sequence increases, indicating a limited capacity short-term memory system (Anderson et al., 1998; Crannell & Parish, 1957; Maybery et al., 2002).

3. Basic error patterns

Errors in serial recall can be item errors or order errors. Item errors include omissions (an item is not recalled), repetitions (an item is recalled more than once, despite being presented on only a single occasion), and intrusions (an item is recalled that was not part of the study sequence). Order errors, which are also known as transpositions, can be further divided into *movement errors* whereby an item is recalled in the wrong position and *exchange errors* whereby two items swap positions. Transpositions are more prevalent than item errors, accounting for around 80% of total errors (Aaronson, 1968).

Transpositions can be classified in terms of their displacement, which refers to the numerical difference between an item's presentation and recall positions. Transpositions with negative displacement values are known as *anticipation errors* and correspond to items recalled ahead of their correct positions. For example, a –4 displacement refers to an item recalled four positions before its correct position. Transpositions with positive displacement values are known as *postponement errors* and correspond to items recalled following their correct positions. For example, a +2 displacement corresponds to an item recalled two positions after its correct position. Items recalled in their correct positions are represented by a displacement value of zero.

Transpositions are typically measured in terms of a transposition gradient, which plots the proportion of transpositions as a function of transposition displacement. A major hallmark of transposition gradients is that the proportion of transpositions is greatest for displacements with an absolute value of one, with error proportions decreasing gradually as the absolute displacement value increases. Henson (1996) has dubbed this property of transpositions as the *locality constraint.*

Farrell and Lewandowsky (2004) have shown that the transposition gradient is accompanied by a systematic pattern of recall latencies. When the mean recall latency of transpositions is plotted as a function of transposition displacement, anticipations are consistently slower than postponements. Furthermore, transposition displacement has different effects on the mean latencies for anticipations and postponements. The mean recall latencies for anticipations increase as a function of displacement. Thus, –3 displacements will have a longer mean latency than –1 displacements, but a shorter mean latency than –6 displacements. In contrast, the mean latencies for postponements are generally invariant with respect to transposition displacement. Thus, the mean latencies for +1, +3, and +6 displacements will generally be very similar.

Another feature of transpositions is that they are characterised by a particular pattern of sequential dependency that is not apparent from inspection of the transposition gradients alone. Specifically, if an item *i* is recalled one position too soon then recall of item *i*-1 is more likely at the next output position than item *i*+1. To explain, given the sequence ABC, if B is recalled at the first output position then a *fill-in* error, reflected by the recall of A at the next output position, is more likely than an *infill* error, reflected by the recall of C. Available data on these errors suggests that fill-in errors outweigh infill errors by a ratio of approximately 2:1 (Henson, 1996; Surprenant, Kelley, Farley, & Neath, 2005). The prevalence of fill-in over infill errors is known as the *fill-in constraint* (Henson, 1996).

6. Grouping effects

Recall accuracy is enhanced if sequences are differentiated into sub-groups, for example, by inserting extended temporal pauses after every few items (Frankish, 1985, 1989; Henson, 1996; 1999; Hitch, Burgess, Towse, & Culpin, 1995; Maybery et al., 2002; Ng & Maybery, 2005; Ryan, 1969a, b). Grouping also engenders effects of primacy and recency within-groups (Frankish, 1985, 1989; Hitch et al., 1996), causes a reduction in the frequency of errors between-groups, and fosters scalloping of the response latency serial position function, as reflected by punctuated peaks in the response latencies for the first item of each group (Anderson & Matessa, 1997; Farrell & Lewandowsky, 2004; Maybery et al., 2002; Parmentier & Maybery, 2008).

7. Positional errors

Two classes of errors in serial recall involve the preservation of positional information. The first class of errors, known as *protrusions* (Henson, 1996), manifest when an item presented on trial *n*-1 is recalled in its same position on trial *n*, despite being outside the experimental vocabulary for that trial. For example, recalling the sequence *B* *X* *R F* instead of *B* *K* *R F* on trial *n* following the sequence *F* *X* *L Y* on trial *n*-1, involves a protrusion error at the second recall position on trial *n*. Thus, a protrusion is an intrusion error from the previous trial that preserves its within-sequence position. Henson (1996) has shown that protrusions occur at a level above that expected by chance, and most often preserve their position from the previous recall episode, rather than the previous presentation episode.

The second class of errors, known as *interpositions* (Henson, 1996), occur in grouped sequences, and are between-group transpositions that preserve their within-group positions (Henson, 1996, 1999; Ng & Maybery, 2005; Ryan, 1969a). For example, recalling the sixth item at the third recall position in a sequence of six items organised into two groups of three constitutes an interposition; the sixth item was presented at the third position within the second group, but has been recalled in the third position within the first group. The incidence of such interpositions in grouped sequences is above that expected by chance.

8. Repetition effects

Repetition errors are rare and widely separated in serial recall. For sequences composed of unique items, repetitions account for approximately 2% (Henson, 1996) to 5% (Vousden & Brown, 1998) of all responses, and are separated by an average lag of 3.34 positions (Henson, 1996). The scarcity of erroneous repetitions is referred to as the *repetition constraint* (Henson, 1996). That individuals avoid repetitions in their responding is further indicated by the results of studies employing sequences containing a repeated item (Duncan & Lewandowsky, 2005; Henson, 1998b; Vousden & Brown, 1998) where repetitions are required. These studies have shown that recall of one instance of a repeat renders it unlikely that the second instance will be recalled, a phenomenon dubbed variously as *repetition inhibition* and the *Ranschburg effect* (after its founder: Pablo Ranschburg).

9. *Item similarity effects*

A classic and robust finding is that sequences composed of purely phonologically dissimilar items (e.g., *F K L R X Y*) are recalled with greater accuracy than sequences composed of purely phonologically similar items (e.g., *B D G P T V*) (Baddeley, 1966; Conrad & Hull, 1964). This *phonological similarity effect* (Baddeley, 1986) is also apparent with lists composed of alternating dissimilar and similar items (e.g., *F B K G R T*). Such mixed lists engender a saw-toothed serial position curve with peaks in the recall of dissimilar items and troughs in the recall of similar items (Baddeley, 1968; Farrell, 2006; Farrell & Lewandowsky, 2003; Henson, Norris, Page, & Baddeley, 1996).

Verbal serial memory: Theory

Having introduced some of the major empirical constraints, consideration is now given to theoretical accounts of verbal serial memory. Although several influential theories of the competency to store and recall ordered sequences of verbal items have been developed over the years, the wealth of contemporary models prevents the provision of a detailed historical perspective. However, by way of an introduction, this section begins by describing a once-popular approach to serial order that shaped initial theoretical developments in this field, but has since fallen from prominence, namely associative chaining theory. The core mechanisms underpinning the operation of a new generation of computational theories, that eschew the chaining notion, are then delineated.

Associative chaining

Associative chaining is the oldest approach to serial order in short-term memory (e.g., Ebbinghaus, 1913/1964) and serial behaviour more generally (e.g., Lashley, 1951). It is mentioned for historic purposes, since there is now a consensus that chaining is inadequate as a solution to the problem of serial order in short-term memory (e.g., Burgess & Hitch, 1999; Henson, 1998a; Page & Norris, 1998). The basic premise behind chaining is that serial order is encoded by forming associations between items. Serial recall is accomplished by traversing these associations, which act as the retrieval cues for sequence production. This constitutes a serial representation of order, because the information necessary for producing a sequence is not simultaneously accessible, rather it emerges dynamically as recall unfolds.

The mechanism of associative chaining appears in various theories of memory (e.g., Ebbinghaus, 1913/1964; Jones et al., 1996; Kieras, Meyer, Mueller, & Seymour, 1999; Lewandowsky & Murdock, 1989; Murdock, 1993, 1995; Wickelgren, 1965). However, the most successful application of the chaining approach to short-term serial recall is the Theory of Distributed Associative Memory (TODAM; Lewandowsky and Murdock, 1989) model. TODAM is a formal model of verbal short-term memory in which items are represented as vectors of random elements, and order information is represented by merging the vectors of pairs of contiguous items. Sequence information is encoded by adding the item and associative vector representations one by one to a common memory vector. Serial recall is initiated by probing the memory vector with a start marker that is linked to the initial item. The first item recalled is then used to cue the second item, which is used to cue the third item, and so on and so forth.

One major objection to simple chaining models of this kind in which order is encoded solely by contiguous associations between items is that if recall should fail mid sequence then the chain is broken and recall must cease. However, TODAM manages to circumvent this shortcoming using the following recall procedure. Due to its use of distributed representations, the output of TODAM in response to a recall cue is not an exact copy of an item, but rather a *blurry* approximation. To recover the item representation, the noisy output vector must first be deblurred by determining which of a pool of experimental vocabulary items it approximates best. If this process is successful then the deblurred item is retrieved and used to cue the next response. However, if this process fails then the associative chain is not necessarily broken, because the blurry output vector can still be used as a retrieval cue, often successfully retrieving the correct next item.

Lewandowsky and Murdock (1989) show that TODAM can reproduce effects of primacy and recency of the serial position curve, as well as several constraints not listed earlier. The recency effect is a consequence of retroactive interference during the encoding of item and associative information, as well as the removal of each item from the competitor set once recalled, which reduces the number of competitors towards the end of the sequence. In contrast, the primacy effect is attributable to an *ad hoc* assumption: the weighting of the encoding strength of each successive association decreases exponentially.

One major shortcoming of TODAM is that it has difficulties explaining transposition errors. Specifically, the model has no mechanism for producing positional exchange errors and the locality constraint on movement errors in general. Murdock (1995) has presented an instantiation of TODAM that incorporates remote as well as contiguous associations (cf. Ebbinghaus, 1913/1964), the strengths of which decrease as a function of the distance between items. According to Murdock (1995) this version of TODAM qualitatively meets these shortcomings[1].

Nevertheless, TODAM suffers from further problems. First, it has difficulties handling sequences containing repeated items. For example, given the sequence *A B A C,* it predicts that recall of *B* and *C* will be compromised, because they share the same recall cue. However, the Ranschburg effect shows that it is the recall of the second instance of the repeat that is impaired, not the items following the repeats. Second, a related problem occurs when participants are given

[1] Murdock (1995) does not provide the results of actual simulations of the model.

sequences containing alternating phonemically similar and dissimilar items, such as the sequence *B K P R*. The model predicts that recall of the dissimilar items *K* and *R* should be impaired, because they possess similar (confusable) recall cues. This prediction is contrary to the data (Baddeley, 1968; Henson et al., 1996), which shows that dissimilar items on mixed lists are recalled as effectively as items in corresponding positions on pure dissimilar lists, if not more so (see e.g., Farrell, 2006; Farrell & Lewandowsky, 2003). Third, TODAM predicts more infill than fill-in errors, because an item recalled too soon will cue its successor in the input sequence by virtue of its direct associative link with that item. This prediction is antithetic to the data (Henson et al., 1996; Page & Norris, 1998; Surprenant et al., 2005).

These shortcomings are not specific to TODAM; they are endemic to all associative chaining models of serial order, including a recent chaining model of serial learning (Solway, Kahana, Addis, & Murdock, in press). In this regard, the recent serial recall model of Botvinick and Plaut (2006) merits comment. These authors show that an Elman (1990) recurrent neural network, once trained to perform serial recall, is able to meet the abovementioned shortcomings of chaining models. This accomplishment is noteworthy, because network models of this kind have previously been disregarded as candidate models of serial recall on account that they operate through chaining (Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998a). This is because the output of such networks is determined by a cue that is a compound of their past contextual states. That the Botvinick and Plaut (2006) model can explain these results is a consequence of its extensive training regime during which it learns not to use chaining and instead develops some form of positional representations. Precisely what these representations are and how they emerge is not yet clear. What is apparent is that the explanatory success of the model is not attributable to a chaining-based representation of serial order.

There are also theoretical objections to chaining as a general approach to serial order. In his seminal article, Lashley (1951) integrated evidence from a variety of domains to highlight the inadequacies of chaining. He noted that the ease, with which phonemes can be combined to form new words, and words to form new sentences, is too flexible for chaining. The preponderance of anticipation errors in speech and typing, he argued, suggest that, "prior to the overt enunciation of the sentence, an aggregate of word units is partially activated or readied" (p.19), signifying that sequence information is activated in parallel, not serially as posited by chaining theories. The need for a parallel representation of serial order is evident in skilled behaviour where many actions such as the finger strokes of a musician, Lashley noted, are performed too quickly for feedback from each response to serve as the cue for the next. Lashley concluded that serial behaviour cannot be explained by a single mechanism based upon associative chaining, proposing instead a two-stage mechanism wherein the first stage all acts to be performed are simultaneously activated, whilst in the second stage a scanning mechanism selects their serial order.

Explanatory mechanisms of contemporary theories

Following the demise of chaining theory a new generation of computational theories of verbal serial memory has emerged in recent years. The wealth and complexity of contemporary theories means that a thorough treatment of each is not possible. Moreover, a focus on the detailed properties of specific models can often obscure important commonalities between them. Fortunately, there has been some theoretical convergence amongst models, and several well-specified mechanisms for the representation and generation of serial order have now been identified, which are commonly employed. Accordingly, I focus on describing the core mechanisms that have found application in different theories, and the empirical precedents upon which they are based. A list of recent computational theories of serial recall and the chief mechanisms they employ to represent serial order is given in Table 1-1.

As can be seen from inspection of this table, current models represent and generate serial order either by: (1) incorporating a *competitive queuing* sequence planning and control mechanism, (2) assuming a *primacy gradient* of activation levels, (3) associating items to temporal, absolute, or relative representations of their sequence positions using *position marking*, (4) implementing *response suppression*, (5) postulating *output interference* during recall; or through some combination of these five mechanisms. Although different combinations of these mechanisms are employed across models, I argue later that they must *all* coexist in any adequate model of verbal

Short-Term Memory Model	Explanatory Mechanisms						
	Competitive Queuing	Primacy Gradient	Position Marking			Response Suppression	Output Interference
			Temporal	Absolute	Relative		
SRN (Botvinick & Plaut, 2006)	✓	×	×	×	×	×	×
SIMPLE (Brown, Neath, & Chater, 2007)	×	×	✓	✓	×	×	✓
OSCAR (Brown, Preece, & Hulme, 2000)	✓	✓	✓	×	×	✓	✓
Burgess & Hitch (1992)	✓	×	×	✓	×	✓	×
Burgess & Hitch (1999, 2006)	✓	✓	✓	×	×	✓	×
Farrell & Lewandowsky (2004)	✓	✓	×	✓	×	✓	✓
SOB (Farrell & Lewandowsky, 2002)	×	✓	×	×	×	✓	×
C-SOB (Farrell, 2006; Lewandowsky & Farrell, 2008)	×	✓	×	✓	×	✓	✓
LIST PARSER (Grossberg & Pearson, 2008)	✓	✓	×	×	×	✓	×
SEM (Henson, 1998a)	✓	✓	×	×	✓	✓	×
Feature model (Neath, 2000)	×	×	×	✓	×	✓	×
Primacy model (Page & Norris, 1998)	✓	✓	×	×	×	✓	×

Table 1-1 *Models of verbal serial recall and the core explanatory mechanisms they instantiate. Note—Farrell and Lewandowsky (2004) compared four models of serial recall built from different combinations of explanatory mechanisms (position marking; position marking and response suppression; position marking and output interference; primacy gradient and response suppression) within a common competitive queuing response selection architecture.*

serial memory (cf. Lewandowsky & Farrell, 2008). Before doing so I begin by briefly delineating each of the mechanisms.

Competitive Queuing

Most models of verbal short-term memory generate serial order using a response selection mechanism developed by Grossberg (1978), and subsequently dubbed by Houghton (1990) as *competitive queuing* (CQ). The popularity of this mechanism is underscored by the fact that it is instantiated in eight of the twelve models presented in Table 1-1. This mechanism, which is buttressed by recent neurophysiological evidence from monkeys (Averbeck, Chafee, Crowe, & Georgopoulos, 2002), is closely related to Lashley's (1951) general theory of serial behaviour. Although popularised in models of verbal short-term memory it has found application in computational theories of serial order generation spanning a variety of serial performance domains (see Glasspool, 2005 for a review).

A schematic of a generic CQ mechanism (e.g., Bullock, 2004; Bullock & Rhodes, 2003) envisaged as a neural network model is illustrated in Figure 1-1. The model comprises two layers of localist item nodes: a *parallel planning layer* and a *competitive choice layer*. The nodes in the parallel planning layer represent the pool of items from which sequences are generated. Recalling a sequence is a two-stage process. In the first stage, an ordering mechanism activates in parallel a sub-set of the nodes in the parallel planning layer, with the relative strength of node activations coding the relative priority of items. In the second stage, the activations in the parallel planning layer are projected to corresponding nodes in the competitive choice layer. The node activations in this layer are governed by *recurrent-competitive-field dynamics*. Each item node excites itself and sends lateral inhibition to competitor nodes in the same layer. This sets-up a response competition and the item with the strongest activation level is chosen for recall, after which a feedback signal from the competitive choice layer inhibits its corresponding representation in the parallel planning layer. This maximum finding selection and suppression process continues iteratively until all sequence items have been recalled. Note that if this mechanism is disrupted by perturbing the activations in one or both of the layers (through the addition of moderate amounts of random noise to the item activations) then it predicts transposition errors akin to those observed in serial recall, including the prevalence of adjacent-neighbour movement and exchange errors.

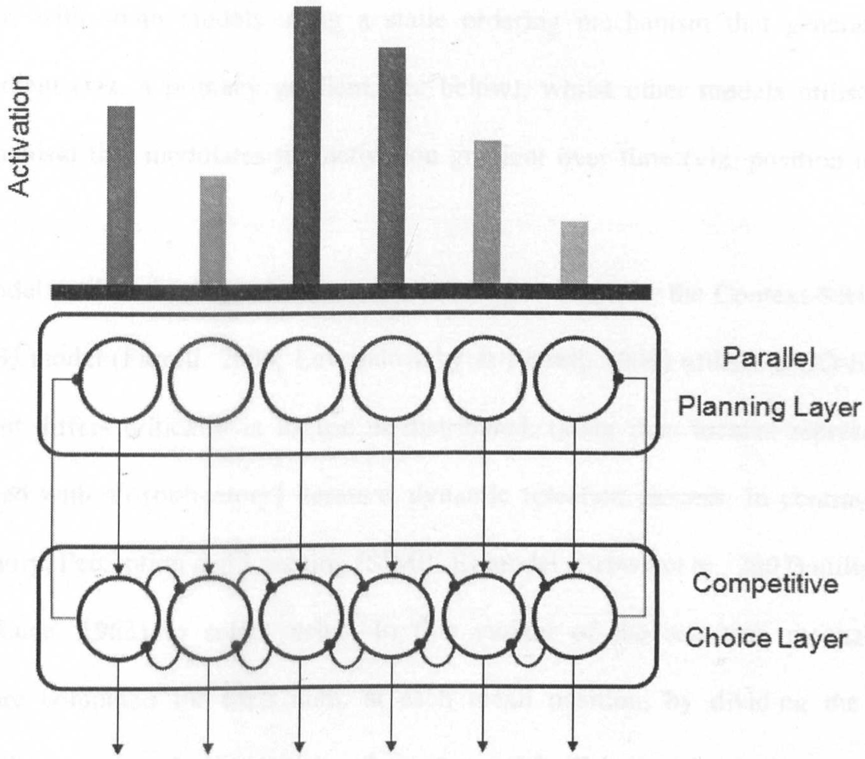

Figure 1-1 *Schematic of a two-layer competitive queuing sequence planning and control mechanism comprising a parallel planning layer (upper field of nodes) and a competitive choice layer (lower field of nodes). Lines terminating with arrows represent excitatory connections, whereas lines terminating with circles represent inhibitory connections. Note that each node in the lower competitive choice layer has an inhibitory connection to every other node in the same layer, but only adjacent-neighbour inhibitory connections are shown to prevent visual interference. Similarly, each node in the competitive choice layer has an inhibitory connection to its corresponding node in the parallel planning layer, but for clarity only feedback connections for the leftmost and rightmost nodes are illustrated. See main text for further details.*

Models that utilise this sequence planning and control mechanism are jointly known as CQ models, because the activations in the parallel planning layer are organised in a competitive queue (Davelaar, 2007). There are several variations on the basic CQ mechanism described above. For example, not all CQ models are neural network based (e.g., Henson, 1998a), and whilst some models implement the competitive choice layer as a recurrent-competitive field (e.g., Burgess & Hitch, 1999, 2006; Farrell & Lewandowsky, 2004), others simply select the strongest item based upon the activations elicited by the ordering mechanism (e.g., Henson, 1998a; Page & Norris, 1998). Models also critically differ in the mechanism that generates the activations in the parallel planning layer, with some models using a static ordering mechanism that generates a single activation gradient (viz. a primacy gradient, see below), whilst other models utilise a dynamic ordering mechanism that modulates the activation gradient over time (viz. position marking, see below).

Not all models utilise the CQ selection mechanism. For example, the Context-Serial-Order-in-a-Box (C-SOB) model (Farrell, 2006; Lewandowsky & Farrell, 2008) utilises a CQ-like selection mechanism, but differs critically in its use of distributed, rather than localist representations of items, combined with an (obligatory) iterative, dynamic selection process. In contrast, the Scale Invariant Memory, Perception and Learning (SIMPLE) model (Brown et al., 2007) utilises the *Luce choice rule* (Luce, 1963) to select items. In this variant of the selection mechanism recall probabilities are computed for each item, at each recall position, by dividing the activations generated by the ordering mechanism by their sum total. The item with the strongest recall probability is then chosen for recall. However, the Luce choice rule is arguably an abstract, non-mechanistic variant of the CQ mechanism. Page (2000) has shown that it is functionally equivalent to selecting the strongest item in a localist network governed by recurrent-competitive field dynamics – the competitive choice layer in CQ models.

Position marking

Position marking is an approach to representing serial order in which sequence items are linked with some varying contextual representation of position. The positional representations are only approximate, meaning that the representations of neighbouring positions will overlap to some degree. At recall the positional cues are reinstated in turn with each sequence item being activated to the extent that its stored position-item association overlaps with the current positional cue. Response selection proceeds by emitting the item activated most strongly in response to each positional cue. This constitutes a *dynamic* representation of serial order, because multiple activation gradients are established during the course of sequence recall.

Models of serial recall instantiating position marking differ according to whether they represent positional information using *temporal*, *absolute*, or *relative* codes (Henson, 1999). An example of a

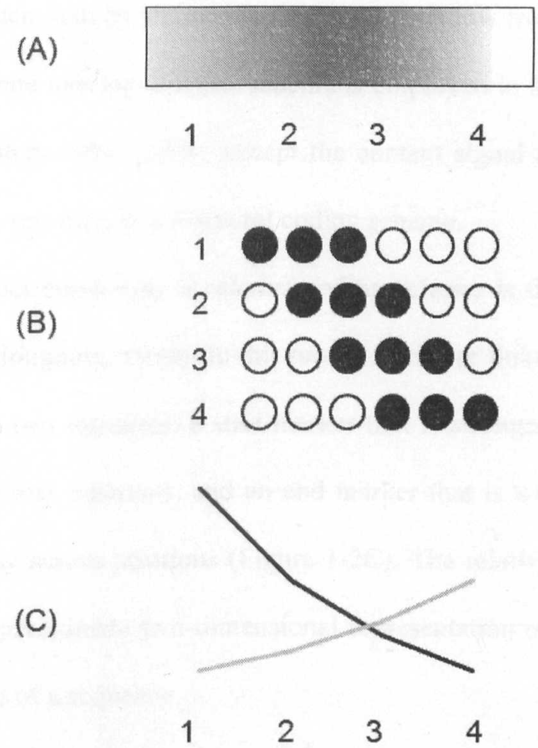

Figure 1-2 *Varieties of positional representations of serial order: (A) a temporal representation of position based on the mechanism employed in OSCAR (Brown et al., 2000). Changes in the shaded bar denote the different states of an oscillator-based timing signal, (B) an absolute representation of position based on the moving window scheme employed by Burgess and Hitch (1992), (C) a relative representation of position based upon the start and end markers in Henson's (1998a) SEM.*

model relying on a *temporal* coding scheme is the Oscillator-Based Associative Recall (OSCAR) model (Brown et al., 2000). In this model items are linked with the different states of a time-varying context signal driven by sets of oscillators operating at different frequencies (Figure 1-2A). At recall the context signal is reset to its initial state, before being replayed. The item activations elicited by the re-evolving context signal are processed by a CQ mechanism that emits the most actively cued items at different recall times. A similar but more abstract temporal coding scheme is utilised by the SIMPLE model (Brown et al., 2007).

Models employing an *absolute* coding scheme include C-SOB (Farrell, 2006; Lewandowsky & Farrell, 2008) and the original Burgess and Hitch (1992) model. For example, in the latter model, items are associated with an event-driven context signal implemented as a vector of inactive nodes containing a dynamic window of active nodes (Figure 1-2B). The context vector changes gradually with the presentation of each item by sliding the attentional window from left to right by a constant one node per item. The same moving window scheme is employed in more recent instantiations of the model (Burgess & Hitch, 1999, 2006), except the context signal is driven by the passage of time, rather than by items, resulting in a temporal coding scheme.

An example of a model employing a *relative* coding scheme is the Start-End Model (SEM; Henson, 1998a; see also Houghton, 1990). In this model items are linked to the varying states of a context signal comprising two elements: a start marker that is strongest for the first position and decreases exponentially across positions, and an end marker that is weakest for the first position and increases exponentially across positions (Figure 1-2C). The relative strengths of the start and end markers provide an approximate two-dimensional representation of the position of each item relative to the start and end of a sequence.

Primacy gradient

A simpler scheme for representing serial order is in terms of a primacy gradient of activation levels, whereby the first item is activated most strongly, and the activations of subsequent items decreases monotonically across positions (Figure 1-3A). This constitutes a *static* representation of serial order, because item priority is coded by a single activation gradient that does not change across time or recall position. When serial order is represented by a primacy gradient complemented by response suppression (see below), ordered recall of a sequence is accomplished by iteratively selecting the most active item and then removing it from the competitive queue by inhibiting its activation so the next strongest item can be emitted. This is the functional mechanism for serial recall in the Primacy model (Page & Norris, 1998), the original Serial-Order-in-a-Box (SOB) model (Farrell & Lewandowsky, 2002), and the LIST PARSE (Laminar Integrated Storage of Temporal Patterns for Associative Retrieval, Sequencing and Execution) model (Grossberg & Pearson, 2008).

Nevertheless, even models that represent serial order through position marking incorporate a primacy gradient. For example, in the OSCAR model (Brown et al., 2000) and the most recent instantiation of the SOB model (C-SOB; Farrell, 2006; Lewandowsky & Farrell, 2008) the primacy

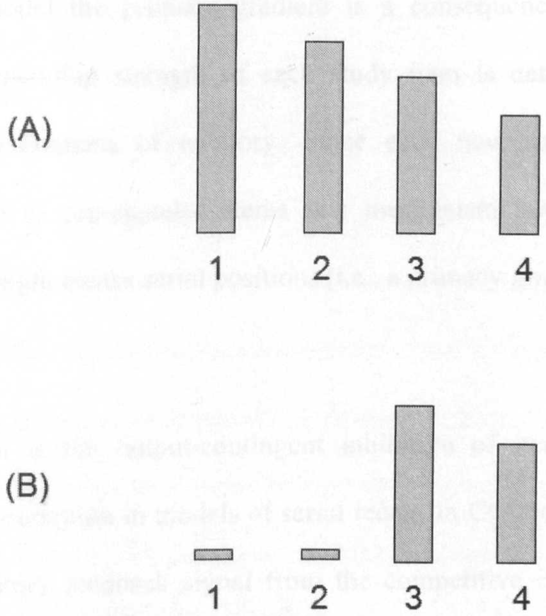

Figure 1-3 *Initial state of a primacy gradient for a four-item sequence (A), followed by partial response suppression of the first two emitted items (B).*

gradient is implemented as an exponential decrease in the strength of the associations between items and their position markers (similar comments apply to the start marker in Henson's SEM). In the Burgess and Hitch (1999, 2006) model, a primacy gradient is implemented through decaying inhibition of activated item nodes during sequence presentation. To elaborate, during presentation of a sequence each stimulus activates its corresponding item node after which it is inhibited. Critically, this inhibition wears off gradually over time, meaning that once recall is initiated earlier items will have had more time for their activations to recover from inhibition. This sets up a primacy gradient of activations over the item nodes with the outcome that the Burgess and Hitch (1999) model can accomplish forward serial recall even in the absence of its positional context signal (Page, 2005).

Thus, most models of serial recall assume the presence of a primacy gradient at some level. However, few models specify a mechanism for its generation (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Page & Norris, 1998; Henson, 1998a). Two apparent exceptions are the SOB model (Farrell and Lewandowsky, 2002; including its recent extension – C-SOB, Farrell, 2006; Lewandowsky & Farrell, 2008) and the LIST PARSE model (Grossberg and Pearson, 2008). For example, in the SOB model the primacy gradient is a consequence of a *similarity-sensitive encoding process*. The encoding strength of each study item is determined by computing its similarity to the current contents of memory. Since each new study item will bear some approximate resemblance to pre-encoded items this mechanism necessarily yields a gradual reduction in encoding strength across serial positions (i.e., a primacy gradient)[2].

Response suppression

Response suppression is the output-contingent inhibition of items, and is, as Table 1-1 indicates, a widespread assumption in models of serial recall. In CQ models, response suppression is reflected by the inhibitory feedback signal from the competitive choice layer to the parallel planning layer, following retrieval of an item. It is considered to be a defining property of those models (Davelaar, 2007). However, it is more important in CQ models (and non-CQ models e.g., Farrell & Lewandowsky, 2002) that represent serial order via a primacy gradient than models that represent serial order through position marking. In the former models, response suppression is necessary to remove an emitted item from the competitive queue so that the next strongest item can be recalled (Figure 1-3B). Without response suppression the response selection mechanism would perseverate on the initial response, which would always remain the most active. Response suppression is less important (but still employed) in models that represent serial order through position marking, because at recall the dynamically evolving positional context continuously modifies the competitive queue.

The existence of response suppression is supported by a number of direct empirical precedents. First, as noted earlier people are poor at recalling both occurrences of the same item when presented with sequences containing repeated items (the Ranschburg effect). That this difficulty is witnessed even when people can detect repetitions with a very high level of accuracy (Henson, 1998b) suggests that this response suppression is obligatory and not under volitional control. The operation of response suppression is further supported by the scarcity of erroneous repetitions in

[2] See Lewandowsky and Farrell (2008) for an explanation of conditions under which a scalloped primacy gradient is predicted by similarity-sensitive encoding.

participants recalls (the repetition constraint). Such erroneous doublings of responses when they do occur tend to be spaced several positions apart (Henson, 1996), which is consistent with the notion that suppressed items gradually decay from inhibition, as is assumed by a number of models of verbal short-term memory (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998a). However, this assumption is contradicted by evidence showing that response suppression is complete and non-decaying, with release from inhibition only being granted upon recall of the entire sequence (Duncan and Lewandowsky, 2005). Recently, Farrell and Lewandowsky (2007) have shown that the accuracy of recall of the final-item in a sequence is modulated by the number of items in the sequence that have already been recalled (and hence suppressed), indicating that response suppression contributes to the recency effect.

These various strands of evidence all point to the existence of a special purpose mechanism for preventing repetition. Indeed, Houghton and Hartley (1995) have speculated that the inhibition of previously performed acts may be a fundamental mechanism for serial order control in a variety of serial performance domains.

Output interference

Output interference refers to the interfering action of recalling an item on the representations of yet to be recalled items. An empirical contribution of output interference to serial recall would be reflected in the primacy effect. However, identifying such a contribution is rendered difficult by the fact that in serial recall the output order of items is perfectly correlated with their input order. Thus, the primacy effect may originate from input processes, such as a primacy gradient, output processes, such as output interference, or a combination of the two. When input and output position are dissociated (Cowan et al., 2002; Oberauer, 2003), by having participants start their recalls at different sequence positions, recall performance decreases across the output positions of items identifying a necessary role for output interference in the generation of the primacy effect. However, that a primacy effect is also observed over the input positions of items additionally implies a role for a primacy gradient (Oberauer, 2003).

Despite the empirical precedent, Table 1-1 identifies only four models that accord a role for output interference during recall (e.g., Brown et al., 2000, 2007; Farrell & Lewandowsky, 2004; Lewandowsky & Farrell, 2008). In the models of Brown et al. (2000) and Lewandowsky and Farrell (2008), output interference is modelled as the addition of increasing amounts of random Gaussian noise, with each successive item recalled, to the weight matrix encoding associations between items and positional markers. In the model of Farrell and Lewandowsky (2004), output interference is modelled as the addition of increasing amounts of random Gaussian noise, with each successive item recalled, to the item activations in the parallel planning layer of a CQ network. The addition of output interference in these models enables them to better accommodate effects of primacy and sequence length.

Mechanism selection

Having outlined the basic mechanisms that form the core of contemporary models of verbal serial memory an important question is as follows: given the available data and the existence of different combinations of mechanisms (as reflected by the models in Table 1-1), what combination is most likely? I have already considered direct evidence that identifies a necessary role for two of the explanatory mechanisms, namely response suppression and output interference. The former mechanism is buttressed by repetition effects amongst other ancillary results, whilst the latter mechanism is buttressed by primacy effects observed over the output positions of items in studies where input and output order was dissociated. However, a key conundrum is whether serial order information is represented by a primacy gradient of activation levels, through positional marking, or by a combination of these mechanisms. To address this question, I will examine the relative virtues and vices of the two mechanisms for representing serial order, by considering the extent to which they can accommodate the empirical benchmarks listed earlier and by determining whether or not their explanatory power can be increased through their confluence.

The models listed in Table 1-1 can be divided into two camps based upon the core mechanism they employ for sequencing. In one camp, there are models which assume that the functional mechanism for representing serial order is a primacy gradient – so called *primacy* or *ordinal* *models* (e.g., Farrell & Lewandowsky, 2002; Grossberg & Pearson, 2008; Page & Norris, 1998). In the other camp, there are models which assume that the functional mechanism for representing serial order is associations between items and position markers – so called *positional models* (Brown et al, 2000; Burgess & Hitch, 1992, 1999, 2006; Henson, 1998a; Lewandowsky & Farrell, 2008).

Primacy models show that a great deal of serial recall data can be explained using simple assumptions. When complemented with response suppression these models predict accuracy serial position curves characterised by extensive primacy and restricted recency, an inverted U shaped latency serial position curve, effects of sequence length, as well as a locality constraint on transposition errors (Farrell & Lewandowsky, 2002, 2004; Page & Norris, 1998). The primacy effect materializes because items near the start of the sequence are more distinctive, meaning these items encounter less competition than items towards the end of the sequence, whereas the recency effect manifests because towards the end of the sequence the cohort of competitor items is reduced due to the operation of response suppression. The locality constraint arises because the disparity in activation between items is smallest for those at neighbouring ordinal positions rendering near-neighbour transpositions most likely.

Primacy models are not unique in their ability to predict these effects. However, there are two constraints that necessitate a direct role for a primacy gradient. First, a primacy gradient complemented by response suppression is necessary to accommodate the finding that fill-in errors are more frequent than infill errors. Primacy models predict this outcome, because if an item i is recalled a position too soon and then suppressed, item i-1 will be a stronger competitor at the next recall position than item i+1, because the former item by virtue of being presented earlier in the sequence will have been encoded more strongly on the primacy gradient. Second, Farrell and Lewandowsky (2004) have shown that a primacy gradient accompanied by suppression of recalled items is necessary to accommodate the pattern of transposition latencies, whereby anticipation errors are slower than postponement errors.

Notwithstanding these apparent strengths, there are two constraints that are beyond the purview of primacy models. The first concerns the multiple effects of grouping, which are generally considered to be reflective of a hierarchical or multidimensional representation of serial order (see e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998a; Lewandowsky & Farrell, 2008). Such effects are inexplicable by primacy models, because they can only represent serial order information along a single dimension of activation strength. Second, the lack of any direct coding of positional information in these models means that they are unable to accommodate the interposition and protrusion constraints.

Positional models provide a remedy to these shortcomings. Like their primacy model counterparts, these models predict effects of primacy and recency of the accuracy serial position curve, an inverted U shaped latency serial position curve, effects of sequence length, as well as a locality constraint on transposition errors (Brown et al., 2000; Burgess & Hitch, 1999; Farrell & Lewandowsky 2004; Henson, 1998a; Lewandowsky & Farrell, 2008). Primacy and recency effects emerge because of the greater distinctiveness of position markers for items near the beginning and end of a sequence. Items near these boundaries have fewer neighbours than items at medial positions, meaning they encounter less competition at recall. The locality constraint arises because the positional codes for items at proximal positions are more similar than the codes for items at distant positions, rendering near-neighbour transpositions most common.

Critically, positional models can accommodate the effects of grouping that are so problematic for primacy models, because they can represent serial order information on multiple dimensions. The elaboration of these models to grouped sequences is accomplished by incorporating one dimension of ordering that represents the positions of groups in sequence and a second dimension of ordering that represents the positions of items within groups (Brown et al., 2000; Henson, 1998a; Lewandowsky & Farrell, 2008). This provides a richer and more robust multidimensional representation of positional information than the one-dimensional positional representation assumed for ungrouped sequences. This representational scheme has been shown to be sufficient to accommodate the major effects of grouping, including a reduction in order errors, effects of mini within-group primacy and recency of the accuracy serial position curve, as well as the scalloped form of the latency serial position curve (Brown et al., 2000; Henson, 1998a; Lewandowsky & Farrell, 2008).

The direct coding of positional information in these models additionally enables them to meet the interposition and protrusion constraints. Interposition errors in grouped sequences materialize because items in corresponding positions in different groups possess similar within-group positional codes, rendering them vulnerable to positional confusions. Protrusion errors are accommodated by assuming that items, as well as being coded for their position in a sequence, are additionally coded for their position within a sequence of sequences (Henson, 1998a). Such errors manifest because items occupying the same sequence position on different trials will possess similar within sequence positional codes.

This explanatory scope notwithstanding, positional models are also subject to limitations. Specifically, the position marking mechanisms in OSCAR (Brown et al., 2000), C-SOB (Farrell, 2006; Lewandowsky & Farrell, 2008), as well as the Burgess and Hitch (1992, 1999, 2006) model suffer from at least two deficiencies: (1) they produce serial position curves that are symmetrically bowed, and (2) they predict an approximately equal number of fill-in and infill errors. In brief, these obstacles arise because the position marking mechanisms in these models lack an ordinal dimension. As noted earlier, the solution has been to incorporate a primacy gradient in these models either via a gradual decrease in the strength of the associations between items and their position markers (Farrell, 2006; Lewandowsky & Farrell, 2008), or through decaying inhibition of activated item nodes (Burgess and Hitch, 1999). The incorporation of a primacy gradient in these models enables them to capture the extended primacy of the serial position curve. The implementation of the primacy gradient in the Burgess and Hitch (1999) model also enables it to meet the fill-in constraint. Note that extended primacy and fill-in are not problematic for SEM (Henson, 1998a), because the position marking mechanism in this model already incorporates a primacy gradient as one of its components.

Although positional models arguably surpass primacy models in their explanatory power it is apparent from the foregoing discussion that positional models must incorporate a primacy gradient at some level if they are to accommodate effects of primacy, as well as fill-in. Moreover, a primacy gradient complemented by response suppression is necessary to enable these models to capture the pattern of transposition latencies documented by Farrell & Lewandowsky (2004). Indeed, in addition to examining the predictions of a primacy model, these authors examined the error latency predictions of three models in which serial order was represented through positional marking (either alone, or in conjunction with output interference or response suppression) and found that none of the latter models could accommodate the empirical pattern observed across their experiments. However, recently Lewandowsky and Farrell (2008) have shown that a model previously not considered in their original work, in which the primacy gradient and response suppression mechanism was augmented with a set of positional markers, provides a more accurate account of the dynamics of transposition errors than the combination of a primacy gradient and response suppression alone.

The foregoing analysis suggests that an adequate model of verbal serial memory must incorporate both a primacy gradient and position marking. Although effects of grouping and positional errors necessitate a direct role for position marking, a primacy gradient complemented by response suppression is required to adequately explain effects of extended primacy, the fill-in constraint, and the dynamics of transposition errors. It is only through the union of these mechanisms that these more challenging data can be comprehensively explained. Combined with the independent evidence for response suppression and output interference outlined earlier, as well as direct evidence for the CQ sequence planning and performance mechanism reviewed later in this chapter, the empirical database suggests that all five explanatory mechanisms must coexist in an adequate model of verbal serial memory.

Visuospatial serial memory: Data

In contrast to the wealth of research devoted to the study of verbal serial memory, visuospatial serial memory has received considerably less attention. Indeed, until recently there existed only a trickling of studies of serial memory employing visuospatial stimuli. As noted earlier, this is perhaps because upon first inspection the problem of serial order appears more relevant to language than visuospatial processing. Fortunately, in recent years a greater appreciation of the importance of serial order processing in the nonverbal domain has been accompanied by a steady stream of studies probing serial memory employing visuospatial stimuli. The purpose of the second half of this chapter is to review the major empirical outcomes of studies in this area and consider the extent to which they can be explained by recourse to the explanatory mechanisms of verbal serial memory outlined above.

By way of introduction, it is useful to consider whether there is any need to posit a visuospatial short-term memory subsystem that is functionally distinct from a verbal short-term memory subsystem, as purported by the working memory framework (Baddeley, 1986, 2000; Baddeley & Hitch, 1974). This theoretical fractionation is buttressed by a wealth of data showing double dissociations between performance on verbal and visuospatial short-term memory tasks. This includes double dissociations obtained under dual task (Farmer, Berman, & Fletcher, 1986; Lange, 2005; Logie, Zucco, & Baddeley, 1990; Meisser & Klauer, 1999; Guerard & Tremblay, 2008) and neuroimaging (Awh, Jonides, Smith, Schumacher, Koeppe, & Katz, 1996; Smith and Jonides, 1997; Smith, Jonides, & Koeppe, 1996) conditions, as well as double dissociations between neuropsychological patients (De Renzi & Nichelli, 1975; Hanley, Young, & Pearson, 1991; Vallar & Baddeley, 1984) and different patient groups (Wang & Bellugi, 1994). Although a counterproposal, based upon the concept of a unitary memory system has been advanced (e.g., Jones, Beaman, & Macken, 1996), the data upon which this theory is founded have proven un-replicable (see e.g., Lange, 2006 and Meisser & Klauer, 1999).

An obvious starting point for exploring visuospatial serial memory is to establish whether the serial position function for visuospatial stimuli bears any resemblance to that observed with verbal stimuli. However, initial attempts to replicate the extensive primacy and restricted recency of the serial position curve were unsuccessful. For example, Philips and Christie (1977a, b) assessed item recognition in visuospatial short-term memory by presenting individuals with sequences of matrix patterns. Memory was assessed by probing each position in a sequence in reverse order, using a two-alternative forced choice test involving the correct item and a dissimilar foil. The serial position function for this task was flat, except for the final item, which exhibited a profound recency advantage. That this empirical outcome is not specific to the visual stimuli employed was confirmed by Broadbent and Broadbent (1981) who assessed item recognition using simpler visual images composed of no more than three stimulus features. Akin to the above authors, they observed a serial position function characterised by no primacy and notable one-item recency. The generality of this primacy-absent serial position function is underscored further by a study by Walker, Hitch, & Duroe (1993) in which novel visual patterns were presented sequentially in different spatial locations. Memory was assessed by presenting a single pattern and having participants determine its spatial position in the original sequence. The serial position function generated by this task was again devoid of primacy, but exhibited strong one-item recency.

The outcomes of these initial studies implied that the determinants of the serial position functions for verbal and visuospatial short-term memory are not the same. It was Avons (1998) who subsequently noted that discrepancies between the serial position functions obtained with verbal and visuospatial stimuli might be attributable to the use of different recall methodologies, as opposed to the operation of distinct mechanisms in the two domains. He noted that studies of verbal short-term memory typically employed serial recall, whereas studies of visuospatial short-term memory typically employed item recognition. The confound here is that serial recall requires, amongst other things, the explicit processing of order information, whereas item recognition requires the processing of item, but not order information. Consistent with this proposition, Ward, Avons, and Melling (2005) have shown that when item recognition is deployed with verbal stimuli a primacy-absent serial position function exhibiting one-item recency is observed.

In the analysis that follows, I review evidence demonstrating that serial position functions exhibiting effects of primacy and recency can be obtained using visuospatial stimuli under testing scenarios that emphasise the explicit processing of order information. Moreover, under such conditions it is possible to replicate many of the empirical constraints identified employing verbal material, thereby giving credibility to the hypothesis that explanatory mechanisms of verbal serial memory may be extensible to the visuospatial domain. However, before delineating this evidence I begin with a few necessary preliminaries.

First, attention must be drawn to the distinction between visual and spatial information in short-term memory. Although it was once considered that visuospatial short-term memory is a unitary system (e.g., Baddeley & Hitch, 1974), there is now a consensus that it is fractionated into distinct visual and spatial sub-components (Baddeley, 2000; Logie, 1995; Pearson, 2001). This assumption is most explicit in Logie's (1995) theory of visuospatial working memory, which consists of two components: a *visual cache* and an *inner scribe*. The visual cache is a passive perceptual input store that deals with static properties of visual images, such as colour, shape, luminance and form, whilst the inner scribe is an active rehearsal mechanism that processes dynamic information, such as the displacements of objects in space. Two strands of evidence buttress this theoretical fractionation of visuospatial short-term memory. First, dual-task studies have shown that visual primary memory tasks are vulnerable to disruption by visual, but not spatial secondary tasks, whereas the reverse pattern of disruption materializes when the primary memory task is spatial (Logie & Marchetti, 1991; Klauer & Zhao, 2004; Tresch, Sinnamon, & Seamon, 1993). Second, neuropsychological patients have been identified that exhibit impairment on visual, but not spatial short-term memory tasks, whilst other patients have been identified that exhibit the converse pattern of preservation and impairment (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999).

Partly due to this theoretical fractionation of visuospatial short-term memory, studies have examined visual and spatial serial memory independently. On the one hand, studies of visual serial memory have employed sequences composed of novel or familiar visual patterns presented from a constant spatial position, akin to the item recognition studies described above. On the other hand, studies of spatial serial memory have utilised sequences composed of novel or familiar two-dimensional spatial locations or directional movements.

The process of recalling sequence information in these tasks typically utilises the method of serial reconstruction mentioned previously. At retrieval, study items are presented simultaneously in their original spatial coordinates (in the case of spatial tasks), or at random locations (in the case of visual tasks), and recall proceeds by pointing to the items in their presentation order. Serial reconstruction is obligatory for visual tasks due to the absence of a natural response mode for visual stimuli. This is not the case for spatial tasks, because even if the locations presented at study are absent during the test phase, individuals can still point towards where they remembered seeing the locations. Thus, in spatial tasks it is possible to utilise a serial recall test procedure. Nevertheless, with few exceptions (e.g., Farrand & Jones, 1996; Farrand et al., 2001; Guerard & Tremblay, 2008), studies of spatial serial memory typically utilise serial reconstruction, in order to minimise dependence on item memory and maximise reliance on order memory.

One major obstacle when studying serial memory in the nonverbal domain concerns the problem of verbal recoding. When engaged in a visuospatial memory task, experimental participants often spontaneously generate verbal descriptors for stimuli and rehearse those descriptors as a sequence of verbal tokens, rendering the task an assessment of verbal, rather than nonverbal serial memory. Researchers have taken several procedural steps to preclude or minimise contamination by such verbal mediation. In studies of spatial serial memory sequences of unique, rather than fixed spatial coordinates are often employed, to prevent participants assigning labels such as digits to locations. In studies of visual serial memory novel or complex familiar stimuli are used for which simple verbal descriptors are difficult to generate. Another approach has been to incorporate an *articulatory suppression* manipulation (Murray, 1967), whereby participants are required to utter a verbal token aloud repeatedly (e.g., "the", "the", "the"...) during sequence presentation. This prevents the use of the speech apparatus for the sub-vocal rehearsal of verbal labels, permitting an examination of visual and spatial serial memory that is uncontaminated by verbal recoding.

A review of recent studies of serial memory employing visual and spatial stimuli now follows.

Visual serial memory

The first study of visual serial memory was conducted by Avons (1998) who examined serial reconstruction of sequences of visual matrices ranging from four to six-items in length. Contrary to the item recognition studies of Philips and Christie (1977a, b), Avons observed serial position functions characterised by effects of primacy and recency, as well as deleterious effects of increases in sequence length on reconstruction accuracy akin to those observed in the serial recall of verbal material. That these effects were not the consequence of contamination by verbal recoding was confirmed by their survival under conditions of articulatory suppression. In a subsequent study, Avons and Mason (1999) extended these observations by analysing the pattern of transposition and item errors underpinning this task. Their analyses revealed that transposition errors were more common than item errors, and that the former errors most frequently involved the displacement of items to positions immediately proximal to their correct positions, with transpositions spanning longer displacements becoming increasingly less likely. Thus, the pattern of transposition errors was found to obey the locality constraint. The authors additionally observed detrimental effects of the visual similarity of items reminiscent of the phonological similarity effect observed with verbal materials.

More recent studies in this domain have employed sequences of unfamiliar faces as stimuli. One of the virtues of using such stimuli is that because adult face recognition is a skilled process (Smyth et al., 2005), faces can be processed quickly and effortlessly (unlike the complex visual matrix stimuli employed by Avons), thereby reducing the demands on stimulus encoding processes, which when overly taxed might obscure serial memory characteristics. Studies by Smyth et al. (2005) and Ward et al. (2005) have shown that serial reconstruction of sequences of unfamiliar faces exhibits effects of primacy and recency of the serial position function, as well as detrimental effects of sequence length on reconstruction accuracy. The former authors also observed more transposition than item errors, a locality constraint on transpositions, detrimental effects of the visual similarity of items, and the survival of all these effects under conditions of articulatory suppression.

Despite the paucity of available studies it is apparent that visual serial memory exhibits numerous characteristics reminiscent of verbal serial memory. These include: (1) effects of primacy and recency of the serial position curve, (2) detrimental effects of increases in sequence length on accuracy, (3) a greater incidence of transposition errors than item errors, (4) a locality constraint on transpositions, and (5) deleterious effects of item similarity. The resistance of these effects to articulatory suppression suggests they are not the consequence of contamination by verbal recoding.

Spatial serial memory

Studies of spatial serial memory have made use of the Corsi-Blocks Task (Corsi, 1972), or basic variants of it. The original task comprises a wooden board containing nine randomly positioned blocks that are tapped by the experimenter according to a particular sequence in the presence of a participant who must subsequently imitate it. However, more recent studies have typically utilised computer versions of the task, which offer among others the advantage of being able to record response latencies. A typical computerised instantiation of the task consists of a visual display containing nine haphazardly organised icons. A sequence is generated by having a sub-set of the icons flash off and then back on again or alternatively by changing the shading of icons temporarily. At recall memory for the sequence is assessed via serial reconstruction.

Although this task has a long history in the neuropsychology literature (see e.g., Corsi, 1972; De Renzi, Faglioni, & Previdi, 1977; De Renzi & Nichelli, 1975; Milner, 1971; Smirni, Villard, & Zappala, 1983), it was not until a study by Smyth and Scholey (1996) that serial position data were first documented. In that study the authors reported effects of primacy and recency of the accuracy serial position curves, a locality constraint on transposition errors, as well as detrimental effects of increases in sequence length on reconstruction accuracy (the latter being documented also in Smyth, 1996; Smyth, Pearson, & Pendleton, 1989; Smyth & Scholey, 1994). Concerning effects of primacy and recency, although these have been documented in further studies using the Corsi-Task (e.g., Avons, 2007; Meisser & Klauer, 1999) many authors continue to fail to report serial position data, relying instead on aggregate measures of recall performance.

Other studies have focused on the temporal dynamics of recall and the role of organisational factors in the Corsi-Task. For example, Fisher (2000) demonstrated that the latency to initiate the first response in the sequence increases as an approximately linear function of sequence length, as has been documented with verbal sequences (Anderson & Matessa, 1997; Maybery et al., 2002). In another set of studies, De Lillo (De Lillo, 2004; De Lillo & Lesk, 2009) has shown that grouping effects akin to those observed with verbal stimuli can be obtained by organising locations into groups based upon their spatial proximity to one another. Most notably, the latencies to group initial items are considerably longer than to group non-initial items, consistent with a hierarchical representation of the sequence. These studies point to functional similarities in the temporal dynamics of recall in spatial and verbal serial memory.

Another task that has been employed to assess spatial serial memory is the Dots-Task of Jones and colleagues (Jones, Farrand, Stuart, & Morris, 1995). In this task, seven locations represented by dots are presented sequentially within an invisible square matrix encased by a border. Unlike the Corsi-Task, which uses a fixed set of spatial coordinates, in the Dots-Task the spatial locations vary randomly across trials. In a further departure from the Corsi-Task, the Dots-Task uses a sequential rather than a simultaneous spatial presentation array. Instead of presenting the locations simultaneously at encoding and then indicating their serial order, each location is presented singly and then cleared from the screen before presentation of the next location. At test the locations re-appear simultaneously prompting serial reconstruction.

A wealth of studies employing this task have shown that it exhibits a plethora of characteristics which are functionally similar to those observed in the serial recall of verbal material. These include effects of primacy and recency of the accuracy serial position curve (Farrand & Jones, 1996; Farrand, Parmentier, & Jones, 2001; Parmentier, Andres, Elford, & Jones, 2006; Parmentier, Elford, & Maybery, 2005; Tremblay, Guerard, Parmentier, Nichols, & Jones, 2006; Tremblay, Macken, & Jones, 2001; Tremblay, Nichols, Parmentier, & Jones, 2005); a latency serial position function characterised by a long initial latency, followed by an inverted U shaped inter-response time profile; deleterious effects of increases in sequence length on reconstruction accuracy (Jones et al., 1995); more transposition errors than item errors (Guerard & Tremblay, 2008); a locality constraint on transposition errors (Jalbert, Saint-Auben, & Tremblay, 2008; Parmentier et al., 2006); and effects of temporal grouping, including a recall advantage for grouped relative to ungrouped sequences, effects of within group primacy and recency of the accuracy serial position curve, a reduction in transposition errors between groups, and a scalloping of the response latency serial position curve (Parmentier et al., 2006).

Another group of studies have explored spatial serial memory using a task that is different in character to those described above. Agam and colleagues (Agam et al., 2005, 2007; Sekuler et al., 2003) have examined serial memory for sequences of connected linear movements. In the spatial task employed by these authors a dot presented in the central screen position moves along a trajectory that is divided into five or six linear connected directional movement segments. At retrieval, the participant is required to draw the sequence of movements of the disc with a stylus on a graphics tablet or imitate the sequence via hand and arm gestures. Agam and colleagues found that this task yields effects of primacy and recency of the serial position curve, as well as a locality constraint on transposition errors.

Before concluding this section, it merits comment that work by Parmentier and colleagues (Parmentier & Jones, 2000; Parmentier, Maybery, & Jones, 2004; Tremblay et al., 2006) has investigated spatial serial memory for heard locations, observing effects of primacy and recency of the accuracy serial position curve, a latency serial position curve exhibiting a long initial latency followed by an inverted U shaped trend, a locality constraint on transposition errors, and effects of temporal (and pitch) grouping similar to those observed by Parmentier et al. (2006) in spatial serial memory for seen locations.

In summary, studies of spatial serial memory have identified a number of characteristics shared with verbal serial memory. These include: (1) serial position functions exhibiting effects of primacy and recency, (2) deleterious effects of increases in sequence length on reconstruction accuracy, (3) latency serial position functions characterised by a long initial latency followed by an inverted U shaped trend, (4) more transposition than item errors, (5) a locality constraint on transpositions, and (6) multifarious effects of temporal grouping. That these effects have been documented using the Dots-Task, which employs unique items across trials, and the Corsi-Blocks task, which has been shown to be insensitive to articulatory suppression (Smyth et al., 1989), suggests that they are not the consequence of contamination by verbal mediation.

Visuospatial serial memory: Theory

The foregoing analysis has identified a number of correspondences between verbal and visuospatial serial memory. It would appear that initial discrepancies in the serial position functions observed for verbal and visuospatial tasks reflected a lack of parity in recall methodologies rather than necessarily the operation of distinct mechanisms in the two domains. What then do these functional similarities indicate?

The only extant theory that has anything to say on this matter is the Object-Oriented Episodic Record (O-OER) model of Jones and colleagues (Jones et al., 1995, 1996), which assumes that verbal, visual, and spatial information is stored in a unitary memory system in which serial order information is represented via associative links between contiguous items. Thus, common effects across serial memory domains are attributed to their reliance on a common associative chaining mechanism. However, it is apparent from the discussion on chaining earlier that such an approach, whether based on entirely contiguous associations or contiguous in conjunction with remote associations, is inadequate as a solution to the problem of serial order in verbal short-term memory. Furthermore, the data just reviewed are equally forceful in ruling out a chaining account of serial order in visuospatial short-term memory.

Given the non-availability of adequate theories of serial order in visuospatial short-term memory a useful starting point for theorising is to hypothesise that the explanatory mechanisms of verbal short-term memory delineated previously, for which there is considerable empirical precedent, are extensible to the visuospatial domain. This might be because some of those mechanisms are shared between the verbal and visuospatial short-term memory subsystems. For example, positional information might be encoded in both domains by associating items to position markers through a common context or timing signal, like those underpinning the operation of the OSCAR (Brown et al., 2000) and Burgess and Hitch (1999, 2006) models. From the perspective of the working memory model this shared episodic contextual control signal might be envisaged to rely on the episodic buffer – a component of the working memory architecture proposed by Baddeley (2000), which facilitates cross talk between the working memory slave-systems. Other mechanisms, such as competitive queuing, a primacy gradient of activations, and response suppression, might be implemented separately in the verbal and visuospatial short-term memory subsystems.

Given the success of explanatory mechanisms of verbal serial memory in explaining data patterns common to both the verbal and visuospatial domains, there is no doubt that these mechanisms are viable candidates for representing the serial order of visuospatial sequences. Furthermore, it is tempting to speculate that because the verbal data implicate a role for all five of the explanatory mechanisms, the functional similarities necessitate that these mechanisms also play a similar role in the visuospatial domain. However, the data do not necessarily warrant this conclusion. For example, the functional similarities might be driven by only a small number of shared or comparable mechanisms that happen to be exerting a large impact on performance. The question, which arises therefore, is what combination of the explanatory mechanisms delineated previously is most likely to underpin the representation and generation of serial order in the visuospatial domain?

I would like to begin by proposing that the functional similarities are most compatible with the notion that both verbal and visuospatial serial memory utilise a competitive queuing sequence planning and control mechanism. Such a mechanism, equipped with an appropriate activation gradient, can accommodate a host of serial memory effects characteristic of the two domains, most notably the preponderance of near-neighbour transpositions, as well as other ancillary outcomes, such as effects of serial position and sequence length. All CQ models predict that the most common errors will be movements and exchanges between items. This prediction is a natural consequence of the parallel sequence dynamics assumed by these models, which when disrupted through the addition of noise can alter the relative priority of items. Near-neighbour transpositions predominate, because the representation of serial order via an activation gradient necessarily implies that the strongest competitors to the target item at each recall position will be items from adjacent serial positions.

These predictions, including the prevalence of near-neighbour transpositions, are not unique to CQ models. However, recent electrophysiological recording studies have provided direct evidence for the dynamics assumed by the CQ mechanism (Averbeck, Chafee, Crowe, & Georgopoulos, 2002; Averbeck, Chafee, Crowe, & Georgopoulos, 2003; Averbeck, Crowe, Chafee, & Georgopoulos, 2003). For example, in a study by Averbeck et al. (2002), macaque monkeys were trained to imitate geometric shapes (triangle, square, trapezoid, inverted triangle) made up of serial movement segments. Prior to sequence imitation a parallel representation of the sequence existed in prefrontal cortex (area 46). Each movement segment was represented by a distinct pattern of neural activity and their relative strengths of activation predicted their relative performance order. During copying neuronal ensembles coding the forthcoming movement segment increased in firing rate until the movement was executed, after which the neural activity associated with the performed segment decreased sharply. Analyses of transposition errors revealed that near-neighbor transpositions were most frequent, and materialized when a neural ensemble coding a movement segment from an incorrect serial position was activated more strongly than the neural ensemble coding the movement segment for the current position. These neural response profiles were strikingly similar to those predicted by Houghton's CQ model (Houghton, 1990). These results with monkeys confirm that the brain utilises the parallel response activation and sequential selection process postulated by the CQ mechanism. I take the preponderance of near-neighbour transposition errors in verbal and visuospatial serial memory to be indirect behavioural signatures of their reliance on such a mechanism.

There are further grounds for supposing that a CQ selection process underpins verbal and visuospatial sequence planning and recall. Specifically, there is mounting evidence to suggest that CQ is a general basis for all sequence planning and control (Bullock, 2004; Bullock & Rhodes, 2003). For example, computational models that utilise the CQ mechanism have been developed and successfully applied to data from a range of serial performance domains, including typing

(Rumelhart & Norman, 1982), speech production (Dell, 1986; Dell, Burger, & Svec, 1997; Hartley & Houghton, 1996; Houghton, 1990; Bohland, Bullock, Guenther, 2009), sequence learning (Rhodes & Bullock, 2002), spelling (Glasspool & Houghton, 2005; Glasspool, Houghton, & Shallice, 1995; Glasspool, Shallice, & Cipolotti, 2006; Houghton, Glasspool, & Shallice, 1994), saccade generation (Brown, Bullock, & Grossberg, 2004), action planning (Cooper & Shallice, 2000), and of course short-term memory (Burgess & Hitch, 1992, 1999, 2006; Henson, 1998a; Page & Norris, 1998, 2009). The success of CQ models in these various domains is attributable to their ability to capture error patterns, such as transpositions, that appear to be characteristic of all serial behaviours.

Having identified CQ as a preferred mechanism for sequence planning and control, the next question, is what combination of the remaining explanatory mechanisms of a primacy gradient, position marking, response suppression, and output interference underpins the representation of serial order in the visuospatial domain? To this question there is currently no answer, because the existing database of empirical findings does not provide direct evidence for either of these mechanisms. Appreciation of why this is the case can be gleaned by consulting Table 1-2, which lists the major empirical constraints of verbal serial memory delineated at the outset of this chapter and indicates which, if any, of the abovementioned explanatory mechanisms they are uniquely attributable to, based upon the evidence reviewed earlier. The table also summarizes which of those constraints have been replicated in the visual and spatial domains and by extension the explanatory mechanisms they implicate.

As can be seen from inspection of this table, the core effects replicated in these domains include effects of primacy and recency of the accuracy serial position curve; effects of serial position on recall latency; effects of sequence length; a greater incidence of transposition than item errors; the locality constraint on transpositions; as well as effects of item (visual) similarity. As impressive as this list may be, neither of these effects are unique signatures of any of the four explanatory mechanisms. On the contrary, they can be explained with reference to various different combinations of those mechanisms. Note that although the primacy effect is often attributed to a

Major empirical effects	Explanatory mechanisms: Primacy gradient	Explanatory mechanisms: Position marking	Explanatory mechanisms: Response suppression	Explanatory mechanisms: Output interference	Effect demonstrated: Visual	Effect demonstrated: Spatial
1. SPC						
Accuracy						
Primacy						
Standard	?	✗	✗	?	✓	✓
Input position	✓	✗	✗	✗	?	?
Output position	✗	✗	✗	✓	?	?
Recency	✗	?	?	✗	✓	✓
Latency						
Preparatory latency	?	?	?	?	?	✓
Inverted U shaped trend	?	?	?	?	?	✓
2. Sequence length	?	?	?	?	✓	✓
3. Basic error patterns						
More transposition than item errors	?	?	?	?	✓	✓
Locality constraint	?	?	?	?	✓	✓
Transposition latencies	✓	✓	✓	✗	?	?
Fill-in constraint	✓	?	✓	?	?	?
4. Grouping effects						
Grouping advantage	✗	✓	✗	✗	?	✓
Acc SPC	✗	✓	✗	✗	?	✓
Latency SPC	✗	✓	✗	✗	?	✓
5. Positional errors						
Interposition constraint	✗	✓	✗	✗	?	✗
Protrusion constraint	✗	✓	✗	✗	?	?
6. Repetition effects						
Repetition constraint	✗	✗	✓	✗	?	?
Ranschburg effect	✗	✗	✓	✗	?	?

Table 1-2 *Major empirical effects of verbal serial memory, the explanatory mechanisms they are attributable to, and whether those effects have been replicated in visual and spatial serial memory. Note—that under the primacy effect 'standard' indicates the primacy effect observed in studies in which input and output position are empirically confounded. 'Input position' and 'output position' represent the primacy effects observed across input and output positions, respectively, when input and output processes are empirically dissociated, as in the study of Oberauer (2003) described earlier. See main text for further details.*

primacy gradient, it can also be explained by output interference. The contribution (or lack thereof) of these mechanisms to the primacy effect can only be disentangled when input and output order is dissociated, as in the study of Oberauer (2003), but no studies of visual and spatial serial memory have yet done so. The more diagnostic constraints listed in Table 1-2, such as the pattern of recall latencies accompanying transposition errors, the fill-in constraint, repetition effects, and positional errors, have not yet been investigated in the visuospatial domain.

One apparent exception is the effects of temporal grouping reported by Parmentier et al. (2006) in spatial serial memory, which point to a role for position marking. However, it must be emphasised that there are aspects of this study that raise doubts about whether it provides evidence for this mechanism. Specifically, Parmentier et al. (2006) used temporal pauses between groups that are considerably longer than those employed in studies of temporal grouping using verbal stimuli. It is therefore possible that the effects of grouping reported by these authors materialized due to the extra opportunities afforded for rehearsal during the long pauses between groups, as opposed to the recruitment of a hierarchical positional representation of the sequence. Moreover, the authors failed to detect an elevation in interposition errors for grouped spatial sequences. This absence is noteworthy given that interpositions are a major hallmark of grouping in verbal serial memory, and the main basis for the claim that grouping effects are attributable to positional marking.

Thus, despite the successful replication of a host of empirical effects observed with verbal stimuli, the existing empirical database does not permit identification of a preferred combination of explanatory mechanisms in visual and spatial serial memory. This evidential gap constitutes the key conundrum motivating the current thesis: what combination of a primacy gradient, position marking, response suppression, and output interference is most likely to contribute to the representation of serial order in the visuospatial domain?

This thesis attempts an answer to this question by exploring whether three empirical constraints listed in Table 1-1, which are direct signatures of specific explanatory mechanisms, generalise to visual and spatial serial memory. These constraints include: (1) the pattern of transposition error latencies documented by Farrell and Lewandowsky (2004), which have been attributed to a representational mechanism combining a primacy gradient, position marking, and response suppression (Lewandowsky & Farrell, 2008), (2) effects of temporal grouping (including interposition errors), which have been attributed to position marking, and (3) the fill-in constraint, which has been attributed to a primacy gradient complemented by response suppression.

Note that although Lewandowsky and Farrell (2008) argued on the basis of their computational modelling work that the pattern of transposition latencies documented in the experiments of Farrell and Lewandowsky (2004) are attributable to a representational mechanism composed of a primacy gradient, position marking, and response suppression, these authors also examined the error latency predictions of four alternative representational mechanisms. These included: (1) position marking, (2) position marking with output interference, (3) position marking with response suppression, and (4) a primacy gradient with response suppression. Importantly, they found considerable heterogeneity in the error latency predictions of the different representational mechanisms. Thus, by evaluating the empirical error latency patterns in visual and spatial serial memory with reference to these alternative theoretical predictions, it will be possible to identify a preferred combination of explanatory mechanisms. Given their apparent diagnosticity, a substantial component of this thesis is devoted to the examination of transposition error latency patterns in visual and spatial serial memory.

Before closing this section it is important to emphasise that I am not arguing that there is no role for domain specific mechanisms and processes in verbal and visuospatial serial memory. However, given the existence of the common behavioural features documented in Table 1-2 and the availability of a set of explanatory mechanisms that can accommodate those shared attributes, it is more parsimonious to work from the assumption that (at least some of) those mechanisms apply across both domains than to speculate from the outset entirely novel alternatives. The extent to which those mechanisms do generalise will in turn determine the degree to which domain specific mechanisms and processes must be invoked.

Behavioural experiments

This thesis contains twelve experiments, which examined serial memory employing a variety of stimuli, manipulations, and performance measures. Some of the experiments used visual stimuli (unfamiliar faces; Experiments 1-6), whilst others used spatial stimuli (two dimensional locations;

Experiments 7-10, & 12), or verbal stimuli (words or consonants; Experiments 6, 10, & 11). All of the experiments employed the recall method of serial reconstruction of order. The experimental manipulations included: sequence length (Experiments 1, 2, 11, & 12), articulatory suppression (Experiments 2 & 5), visual similarity (Experiment 5), temporal grouping (Experiments 4, 7, 9, & 10), and post-sequence interference (Experiment 7). The basic measures of performance for all experiments were accuracy serial position curves and transposition gradients. Other experiments additionally report latency serial position curves and transposition error latencies (Experiments 1-9), distributions of errors within and between groups (Experiments 7, 9, & 10), and relative incidences of fill-in and infill errors (Experiments 1, 11, 12).

Computational modelling

As well as behavioural experiments, this thesis makes extensive use of computational models in order to explore the consequences of different design choices for representing serial order. Specifically, in Chapter 2 a generic modelling framework is developed based upon a dynamic CQ architecture within which serial order information can be represented via different combinations of a primacy gradient, position marking, response suppression, and output interference. This single dynamic architecture, which is based upon that employed by Farrell and Lewandowsky (2004), is used in Chapter 3 to examine the transposition latency predictions of five models built from different combinations of the explanatory mechanisms. These predictions are tested empirically in visual and spatial serial memory in Chapters, 4, 5, and 6. One might question the need to implement these models given that their error latency predictions have already been documented in Farrell and Lewandowsky (2004). However, it is important to show that the predictions documented by these authors replicate before using them to draw important theoretical inferences from new data. Furthermore, as will become apparent in Chapter 3, there are aspects of the modelling reported by these authors that require further detailed computational investigation.

This generic modelling architecture is employed again in Chapter 8 to reinterpret the explanatory mechanisms necessary to accommodate the fill-in constraint. Although the fill-in constraint is generally compatible with a primacy gradient complemented by response suppression, an acknowledged shortcoming of this mechanism is that it over-predicts the extent of fill-in errors, suggesting it is incomplete. To determine what further assumptions are necessary to accommodate the fill-in constraint several models are constructed that supplement the basic primacy gradient and response suppression mechanism with additional assumptions. Of particular interest is whether the shortcomings of this mechanism can be overcome by augmenting it with a set of positional markers. The patterns of fill-in and infill errors in verbal, visual, and spatial serial memory are then explored and the different models fit to the data.

Computational modelling is further employed in Chapter 7 to explain differences in temporal grouping effects in verbal and spatial serial memory observed in Experiment 10. Although this experiment revealed common effects of grouping on recall accuracy and latency in the two domains, it also failed to detect an increase in interposition errors for grouped spatial sequences (cf. Parmentier et al., 2006). It is speculated that this discrepancy is due to different representations of the positions of items in grouped verbal and spatial sequences, with verbal items being coded for their positions within groups, and spatial items being coded for their positions in sequence. These hypotheses are tested using a version of Henson's (1998a) Start-End Model.

Overview of thesis

Chapter 2 presents an introduction to computational and mathematical modelling. The generic modelling architecture employed to compare the transposition latency and fill-in and infill error predictions of different combinations of explanatory mechanisms in Chapters 3, 4, 5, 6, and 8 is subsequently introduced. This is followed by a description of model evaluation and selection techniques adopted in this thesis.

Chapter 3 reports transposition latency predictions for five alternative models and associated mechanisms for representing serial order. Fits of two of the models to verbal serial recall data taken from Farrell and Lewandowsky (2004) support the notion that the serial order of a verbal sequence is represented through the combination of a primacy gradient, position marking, and response suppression.

Chapter 4 examines transposition latencies in visual serial memory across a range of experimental manipulations (Experiments 1-6). The data consistently support the error latency predictions of a model combining a primacy gradient, position marking, and response suppression. Quantitative fits of models to data taken from Experiment 2 confirm that when model parameters are estimated from the data this model is the only one that predicts the observed empirical pattern.

Chapter 5 examines transposition latencies in spatial serial memory across manipulations of temporal grouping (Experiments 7 & 9) and post-sequence interference (Experiment 8). The results of these experiments, combined with quantitative fits of the models to data from Experiment 9, are once again in accordance with the error latency predictions of a model combining a primacy gradient, position marking, and response suppression. Effects of temporal grouping are also reported providing further support for the role of position marking. However, contrary to studies of grouping using verbal stimuli no increase in interposition errors is observed for grouped relative to ungrouped spatial sequences.

Chapter 6 presents parameter space sensitivity analyses of the models and mechanisms for representing serial order presented in Chapter 3. The purpose of these analyses is to examine the transposition latency predictions of the models across a wide range of their parameter settings, in order to establish the robustness of their predictions presented in Chapters 3, 4, and 5, which are based upon single parameter values. The outcomes of these analyses indicate that the predictions of the models witnessed in those chapters are representative of their wider behaviour and can therefore be attributed to their core principles. Critically, the modelling outcomes indicate that the better account of the empirical pattern provided by the model combining a primacy gradient, position marking, and response suppression is not attributable to it being overly complex.

Chapter 7 directly compares effects of temporal grouping in verbal and spatial serial memory (Experiment 10). Common effects of grouping on recall accuracy and latency are observed in the two tasks, but consistent with Experiments 7 and 9, no increase in interposition errors is observed for grouped relative to ungrouped spatial sequences. It is speculated that this discrepancy is due to different positional representations for grouped sequences in the two domains. Simulations of the

Start-End Model (Henson, 1998a) which varied the nature of grouped positional representations are conducted to test this hypothesis.

Chapter 8 presents three experiments that examined the relative distribution of fill-in and infill errors in verbal (Experiment 11), visual (re analysis of Experiment 1), and spatial (Experiment 12) serial memory. Applications of four models of serial order to the data from each of these experiments provided further evidence for the combination of a primacy gradient, positional marking, and response suppression in the three domains, whilst additionally providing tentative support for a fourth mechanism: namely a restricted end positional marker that codes the position of the final item in a sequence relative to the end of that sequence.

Chapter 9 summarises the evidence for the operation of competitive queuing, a primacy gradient of activations, positional marking, and response suppression in visual, spatial, and verbal serial memory obtained in this thesis. It also considers the relationship between the mechanisms of serial order in verbal and visuospatial short-term memory, and their possible locus within the multi-component model of working memory introduced at the outset of this chapter. The chapter concludes by considering some future empirical and modelling directions for the research.

2

Introduction to computational modelling

Abstract

This chapter provides a brief introduction to computational and mathematical modelling of cognitive processes, before describing the generic modelling architecture employed to simulate serial recall in subsequent chapters. The final section delineates model evaluation and selection methods relevant to the comparisons of models throughout this thesis.

Introduction

The goal of cognitive psychology is to explain the mechanisms underpinning such diverse mental phenomena as attention, learning, and memory. Cognitive researchers accomplish this by running behavioural experiments that target the cognitive processes of interest and by drawing inferences about those processes from resulting data. From these inferences, functional accounts of the cognitive processes under investigation are derived known as *cognitive models*.

Cognitive models appear in many different guises. The most common, by far, are *verbal theories*, which provide functional accounts couched in terms of verbal statements. Often they take the form of box and arrow diagrams accompanied by descriptions of the different components and stages of information processing. The much celebrated theory of working memory developed by Baddeley and Hitch (1974) is a classic example of this verbal style of theorizing. Verbal models play a fundamental role in cognitive psychology: they provide a broad initial foundation upon which to understand a problem domain, and they pose new questions and generate new data, which can then be used to refine the theory. The working memory framework (Baddeley & Hitch, 1974) which has fuelled more than three decades of research again serves as an illustrative example, and constitutes a parade case for the need for broad verbal theories of cognition. Notwithstanding the import of such theories, it will be argued below that a desirable goal for all models of cognition is an unambiguous, formal instantiation of their core principles.

Another class of models that offer this theoretical precision are *mathematical models*. Instead of verbal statements, the assumptions of these models are embedded in equations and algorithms that formally specify the cognitive processes they seek to explain. Although mathematical models can be classified in several ways, a useful taxonomy is between *analytical* and *computational* (or simulation) models (Lamberts, 2005). Analytical models are composed of equations (such as closed-form expressions) that describe the different components and stages of information processing. The operations performed by these models are typically linear, enabling predictions to be obtained directly by assigning values to variables and by solving the model equations. Thus, model predictions can be generated via one-shot calculations.

Computational models, on the other hand, are simulation models implemented as computer programs that mimic the operations hypothesised to underpin the cognitive processes being modelled. Such models are specified in terms of algorithms, which are assumed to reflect the kinds of processing performed by some component of the cognitive system. Connectionist and neural network models fall into this category (see Anderson, 1995; McLeod, Plunkett, & Rolls, 1998; Rumelhart & McClelland, 1986). Typically the computations performed by these models are nonlinear and complex, meaning that an analytical solution cannot be derived. Instead, the predictions of such models must be obtained by running many iterations of the computer programme in which the model is instantiated. All the models employed in this thesis fall into this category.

All mathematical models of cognition include parameters. They correspond to psychological constructs that are hypothesised to be necessary to explain the cognitive process being modelled. For example, a model of short-term memory might include a parameter that reflects the phonological similarity between items, while another parameter might correspond to a response threshold that an item must exceed in strength in order to be retrieved. The parameters of a model assume numerical values that are typically free to vary, within some bounded constraints. This freedom is granted because human behaviour is not regular and constant, but instead varies depending upon internal and situational factors. The parameters combine together in one or more equations that determine the behaviour of the model. The number of free parameters a model employs is an important property that determines its complexity and the range of data to which it can be applied. The issue of model complexity in model evaluations and comparisons is considered later in this chapter.

The goal of mathematical psychology is the development of theories of cognition couched in terms of mathematical principles, as opposed to verbal statements. It entails a commitment to the view that human cognition operates in a mechanistic fashion. The onus on researchers in this domain is to formally specify the processes that characterise the different facets of the cognitive system. The field of mathematical psychology has developed in part due to dissatisfaction with the lack of specificity associated with purely verbal theorising. The reason for this dissatisfaction is summed up by Roger Ratcliff, quoting a comment made by Jonathan Cohen of Carnegie Mellon University:

> "Boeing would not design a jet liner based on verbal hypotheses, and certainly the US FAA would not certify such an aeroplane. If the brain/mind is of at least the complexity of a jet aeroplane, how could it be possible that verbal hypotheses alone would allow us to understand mind and behaviour?" Ratcliff (1998, p.129).

The objection is that unaided verbal hypotheses provide restricted accounts of cognitive phenomenon and must therefore be supplemented by formal approaches. Mathematical modelling provides the analytical tools to help bridge this theoretical gap. A fundamental goal of mathematical psychology in this endeavour is the development and comparison of competing models of a cognitive process, in a bid to identify the one that best explains the data (Estes, 1975).

In the past two decades interest in the mathematical modelling of cognition has soared and the field of human memory has been one of the greatest beneficiaries of this burgeoning area (see e.g., Lewandowsky & Heit, 2006). Mathematical models have been developed to explain a diverse range of memory phenomenon, including: recognition memory (McClelland & Chappell, 1998; Norman & O'Reilly, 2003; Shiffrin & Steyvers, 1997), working memory (Oberauer & Kliegl, 2006), serial recall memory (Botvinick & Plaut, 2006; Brown et al., 2000; Burgess & Hitch, 1992, 1999, 2006; Farrell & Lewandowsky, 2002; Henson, 1998, Lewandowsky & Farrell, 2008, Page & Norris, 1998), and free recall memory (Davelaar, 2007; Davelaar, Goshen-Gottstein, Askenazi, Haarman, & Usher, 2005; Davelaar, Haarman, Goshen-Gottstein, & Usher, 2006; Howard & Kahana, 2002; Polyn, Norman, & Kahana, 2009; Raaijmakers & Shiffrin, 1981). Models have even been developed that attempt to unify data obtained from a variety of tasks and paradigms under a common theoretical umbrella (Brown et al., 2007; Grossberg & Pearson, 2008).

What then are the reasons for the growing popularity of mathematical models? The key benefit of mathematical modelling is that it enables a theory to be specified in a precise manner. Unlike a verbal theory, there is no ambiguity in a mathematical model; its operations are clearly defined by the equations/algorithms that govern its behaviour. This enables the theory to be communicated more clearly, which reduces the likelihood of misinterpretation. Indeed, mathematical models often exhibit predictions that are at odds with intuition (see Lewandowsky, 1993, for illustrative examples), and therefore serve as important tools for preventing theoretical misconceptions (Hintzman, 1991). Another merit is that they enable researchers to move beyond predictions based on ordinal relationships between experimental conditions (Myung, Pitt, & Kim, 2003). Mathematical models can be employed to explore complex aspects of data, such as the shape of distributions and the magnitude and generality of effects. Furthermore, a mathematical model can be evaluated, amongst other ways, by directly testing its ability to fit data. Consequently, there can be no doubt whether or not a model can reproduce an empirical outcome under scrutiny.

Mathematical models are particularly useful for testing novel and complex ideas that might otherwise prevent the formulation of clear predictions (Lewandowsky, 1993). A formal model can thus serve as an aid to help overcome limitations of human reasoning. In advocating the need for formal models of cognition, Hintzman (1991) identified more than ten such limitations; they include, but are not restricted to: constraints on working memory capacity, confirmation bias, and failure to recognise evidence that discounts a theory. These limitations can be circumvented through the use of formal models, because they are implemented and tested on a step-by-step basis (offsetting working memory demands) and a model that generates predictions that are at odds with an empirical outcome can quickly nullify a researchers false assumptions about that model (overcoming confirmation bias and failure to recognise discounting evidence).

Another virtue of formal models is that it is reasonably easy to trace a model's behaviour back to its core principles. This can be accomplished, for example, by varying the parameters controlling a mechanism of interest, while holding the parameters for all other mechanisms constant. Any changes in the behaviour of the model can then be traced to the mechanism being manipulated. These kinds of explorations of a model's components can yield revealing insights into its behaviour, which can often generate novel, testable predictions. A similar approach is useful when one is interested in which combination of a set of hypothetical mechanisms best explains a cognitive process. One can derive a number of competitor models that instantiate different combinations of those mechanisms, preferably employing a nested structure. By testing the predictions of the models against appropriate data, it is possible to isolate a model and an associated combination of mechanisms that provides the best account. This approach is used frequently in the context of this thesis.

Despite its many benefits, mathematical modelling is not without its pitfalls. First, it is easy to develop a model that is too flexible because it contains too many free parameters. Such a model might fit relevant data better than its rivals, but it may accomplish this not because it is a better approximation of the cognitive process under study, but rather because it has more degrees of freedom to capture the data (Pitt & Myung, 2002). Second, although a fundamental goal of mathematical modelling is the comparison of competing models, the means by which this is accomplished are far from trivial and constitute perhaps the greatest hazard of the modelling enterprise. Fortunately, recent years have seen considerable advancements in the development of model comparison techniques that provide a firm foundation upon which to evaluate and adjudicate between rival theories. Some of these techniques are introduced later in the chapter.

Having provided a general introduction to mathematical modelling, the next section describes the generic modelling architecture employed in this thesis. However, before moving on, it is important to mention some changes in terminology relative to Chapter 1. Specifically, in this and subsequent chapters, I frequently refer to the mechanisms for representing serial order outlined in the previous chapter (i.e., position marking, primacy gradient, response suppression, and output interference) as *representational principles*. The central reason for this is that the modelling framework delineated in this chapter only specifies the nature of the representations resulting from a particular mechanism of serial order, without however specifying the mechanism that generates those representations. To provide a clear example of this distinction, the implementation of the primacy gradient described later specifies the nature of the primacy gradient representation (the activations of items decreases across input positions), but without specifying the process that gives rise to the gradient. An example of a process that gives rise to a primacy gradient would be the similarity-sensitive encoding process of the SOB model of short-term memory (Farrell & Lewandowsky, 2002; Lewandowsky & Farrell, 2008).

Modelling approach

This thesis uses computational modelling to explore the predictions of different combinations of principles for representing serial order in short-term memory. One way to accomplish this would be to implement the computational models of serial recall listed in Table 1-1, which differ in terms of the core principles to which they are committed. However, it would be complex and time consuming to implement but a handful of those models. Moreover, they contain ancillary assumptions that can often obscure the predictions of their core principles. A better approach employed by Farrell and colleagues (Farrell & Lewandowsky, 2004; Farrell & Lelievre, 2008) – and the one adopted here – is to use a common modelling architecture within which the predictions of different combinations of representational principles can be examined in a controlled environment where the only differences are the representations under manipulation.

As noted in Chapter 1, most contemporary theories of short-term memory represent serial order by: (1) incorporating a *primacy gradient* of activations, by (2) associating items to ordinal or temporal representations of their sequence positions using *position marking*, by (3) inhibiting recalled items using *response suppression*, and by (4) implementing *output interference* during recall, or through some combination of these four principles. This thesis contrasts the predictions of models built from different combinations of these representational principles within a common competitive queuing response selection architecture that permitted the derivation of response probability as well as latency predictions. The predictions of the models are then explored empirically, in order to identify a preferred combination of principles for representing serial order in visual and spatial serial memory.

The common response selection architecture, which is based upon that employed by Farrell and Lewandowsky (2004) to model transposition latencies in serial recall, is outlined next, after which the implementation of the representational principles is described.

Common competitive queuing response selection architecture

Most contemporary theories of serial recall posit a two-stage mechanism for recall (Botnivik & Plaut, 2006; Brown et al., 2000, 2007; Burgess & Hitch, 1992, 1999, 2006, Henson, 1998; Page & Norris, 1998). In the first stage, sequence items are simultaneously activated by the ordering mechanism, with their relative activation levels coding their relative priority. The activations elicited by the ordering mechanism are then projected to a second response selection stage, which is used to select an item from amongst the response alternatives. Although the mechanism for response selection varies across models, most employ the competitive queuing mechanism delineated in Chapter 1.

In the current modelling, competitive queuing is used as a generic response selection architecture within which the predictions of different combinations of representational principles can be contrasted. In departure from the competitive queuing mechanism delineated in Chapter 1 (see Figure 1-1), in the implementation reported here the parallel planning layer in which the activations elicited by the ordering mechanism are maintained, was omitted. Instead, activation gradients corresponding to a particular representation of serial order were specified using appropriate parameters (see below) and used directly as input to a competitive choice layer. Thus, no attempt was made to model the encoding of serial order information, because the selection mechanism is insensitive to the processes that generate the activations representing the strength with which items compete at retrieval. These changes were made, because interest centered not on providing a process implementation of the different mechanisms, but rather on examining their predictions based upon the retrieval conditions they elicit.

A schematic of the architecture employed is shown in Figure 2-1. It comprises a single competitive layer of localist item nodes corresponding to the pool of response elements from which sequences can be generated. Each node has a recurrent self-excitatory connection, plus lateral inhibitory connections to all other nodes. The excitatory and inhibitory weights are a hardwired property of the network and were set to constant values of 1.1 and -0.1, respectively. Serial order is represented in this network by setting starting activation values for the item nodes for each output position, the derivation of which is determined by the representational principles being modelled (see below). The activation of each node is determined by this initial external input, plus recurrent self-excitation, and lateral inhibition received from all other item nodes, which are jointly determined by the following equation (Houghton, 2005):

$$Int_{j,t} = a_{j,t-1}w^{+} + w^{-}\sum_{i \neq j} a_{i,t-1} \tag{2-1}$$

Where Int_j constitutes the internal input a unit receives from within the layer, a_j represents the initial activation of unit j determined by its external input, a_i constitutes the activation of all other nodes in the layer, w^+ and w^- represent the excitatory and inhibitory weight values, respectively, and t corresponds to time[1]. The first term on the right hand side of equation 2-1 represents the recurrent self excitation, whereas the second term represents the lateral inhibition received from all other nodes in the layer. This sets up a "winner-takes-all" response competition over the item nodes, and the initially most active node (the node receiving the highest external activation) has the

[1] Note that negative activations are not allowed to spread in equation 2-1 (negative activations are set to zero) otherwise a node with a negative activation would send excitation, rather than inhibition, to other nodes within the layer.

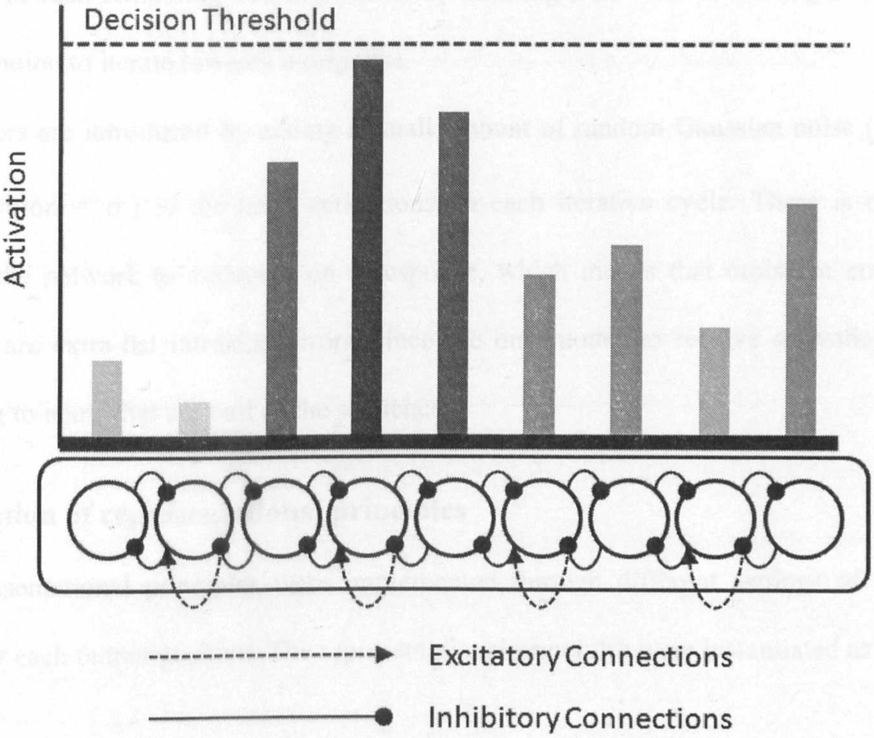

Figure 2-1 *A schematic of the lateral inhibition network. This single layer, winner-takes-all network, corresponds to the competitive choice layer in competitive queuing models of serial behaviour (Bullock, 2004; Grossberg, 1978; Houghton, 1990). Each localist item node possesses a single recurrent excitatory connection, in addition to lateral inhibitory connections to all other item nodes. Nodes are fully interconnected, but only excitatory connections to nodes 2, 4, 6 and 8 and near-neighbour inhibitory connections are shown to prevent visual cluttering. At any time the activation of a node within the layer is determined by the external input it receives from outside the layer, plus positive recurrent self-excitation, and within-layer inhibition. Note—the number of nodes in the network is dependent on the sequence-length being modelled.*

advantage that it will send more activation to itself than any other node and will also receive the least lateral inhibition. The node activations are iteratively updated over time by using the pattern of activation over the nodes at time *t*-1 as the input to the network at time *t* before re-computing the node activations using equation 2-1. Over time the activation of the strongest node will gradually increase, whereas the activations of the less active nodes will gradually decrease as they become more inhibited by lateral inhibition. This iterative updating process continues until the strongest node exceeds a response threshold *T* (set to 1.0 for all simulations) and the number of iterative cycles required to determine the response is taken as the networks recall latency. This process is then repeated at each remaining output position by defining a new set of starting activations and allowing activation to iterate towards a response.

Order errors are introduced by adding a small amount of random Gaussian noise (mean = μ; standard deviation = σ) to the node activations on each iterative cycle. There is no temporal deadline for the network to converge on a response, which means that omission errors are not possible, nor are extra-list intrusion errors, since the only nodes to receive activation are those corresponding to items that are part of the sequence.

Implementation of representational principles

The representational principles were implemented through different settings of the starting activations for each output position. The representational principles were instantiated as follows:

Position marking

Position marking was implemented by specifying activations for item nodes that reflected the confusability of item positions. Specifically, the activation a of the item node j for the current output position p was strongest, whilst the activations of neighbouring item nodes decreased as an exponential function of their positional distance from the target item:

$$a_j = \lambda\phi^{|j-p|} \tag{2-2}$$

Where λ is a parameter controlling the strength of the item node corresponding to the current recall position ($0 < \lambda < 1$), and ϕ is a parameter controlling the distinctiveness of the activations for all other item nodes ($0 < \phi < 1$). The node activations for each output position were then normalized by expressing each node's activity as a proportion of the summed activity of all nodes:

$$a_j = \frac{a_j}{\sum_i a_i} \tag{2-3}$$

This parameterization produces activation gradients akin to those generated by the positional context signals in the Burgess and Hitch (1992, 1999) and OSCAR (Brown et al., 2000) models and corresponds to an absolute representation of position. Retrieval proceeded by imposing the activation gradient for each output position over the item nodes, one at a time, and allowing activation to iterate towards a response. Example starting activations for position marking for the third output position in a six-item sequence, when $\lambda = 1$ and $\phi = .65$, are illustrated in Figure 2-2A.

Primacy gradient

The primacy gradient was implemented as a decrease in activations across input positions. Two versions of the gradient were explored: an exponential and a linear version. For the exponential primacy gradient the activation of a node j was given by:

$$a_j = a_1 \gamma^{j-1} \tag{2-4}$$

Where a_1 is the activation of the item node corresponding to the first input position ($0 < a_1 < 1$), and γ is a parameter controlling the steepness of the primacy gradient ($0 < \gamma < 1$). For the linear primacy gradient the activation of a node was defined by:

$$a_j = \begin{bmatrix} a_1 & \text{if } j = 1 \\ a_{j-1} - \gamma & \text{otherwise} \end{bmatrix} \tag{2-5}$$

Retrieval commenced by imposing the entire primacy gradient over the item nodes for the first output position and allowing activation to iterate towards a response. This process was then repeated for each subsequent output position by imposing the same primacy gradient over the item nodes, but with suppression (see below) of those nodes corresponding to recalled items. Example starting activations for an exponential primacy gradient for the first output position in a six-item sequence, when $a_1 = 1$ and $\gamma = .8$, can be inspected in Figure 2-2B.

Primacy gradient + position marking

In some simulations serial order was represented through the combination of a primacy gradient and position marking. Under these circumstances the activation of the item node for the current output position was calculated under the primacy gradient and position marking schemes described above, before adding their activations as follows:

$$a_j = (1-\omega)\cdot a_{j,pm} + \omega \cdot a_{j,pg} \tag{2-6}$$

Where $a_{j,pm}$ represents a node's position marking activation, $a_{j,pg}$ represents its primacy gradient activation, and ω is a weighting parameter that determines the attentional weight given to the two dimensions of ordering ($0 < \omega < 1$). Unless specified otherwise, ω was set to a constant value of .5, thereby giving equal weight to the two dimensions. Retrieval proceeded by imposing the activation gradient for each output position over the item nodes, one at a time, and allowing activation to iterate until a response was made. Example starting activations for an exponential primacy gradient combined with position marking for the third output position in a six-item sequence, when $a_1 = .6$; $\gamma = .9$; $\lambda = 1$; $\phi = .65$, can be seen in Figure 2-2C.

Response suppression

Response suppression was implemented as the proportional reduction of an item's activation once it had been selected, expressed as:

$$a_{j,rs} = a_j(1-\alpha) \tag{2-7}$$

Where $a_{j,rs}$ represents the activation of a node once suppressed and α is a parameter controlling the extent of response suppression ($0 < \alpha < 1$). The resulting value was then used as the node's starting activation for all subsequent output positions. However, if the same node was selected again at a later output position then the nodes starting activity at any remaining positions was based upon the new suppressed value. Example starting activations for response suppression for the fourth output position in a six-item sequence, when $\alpha = .95$, are shown in Figure 2-2D.

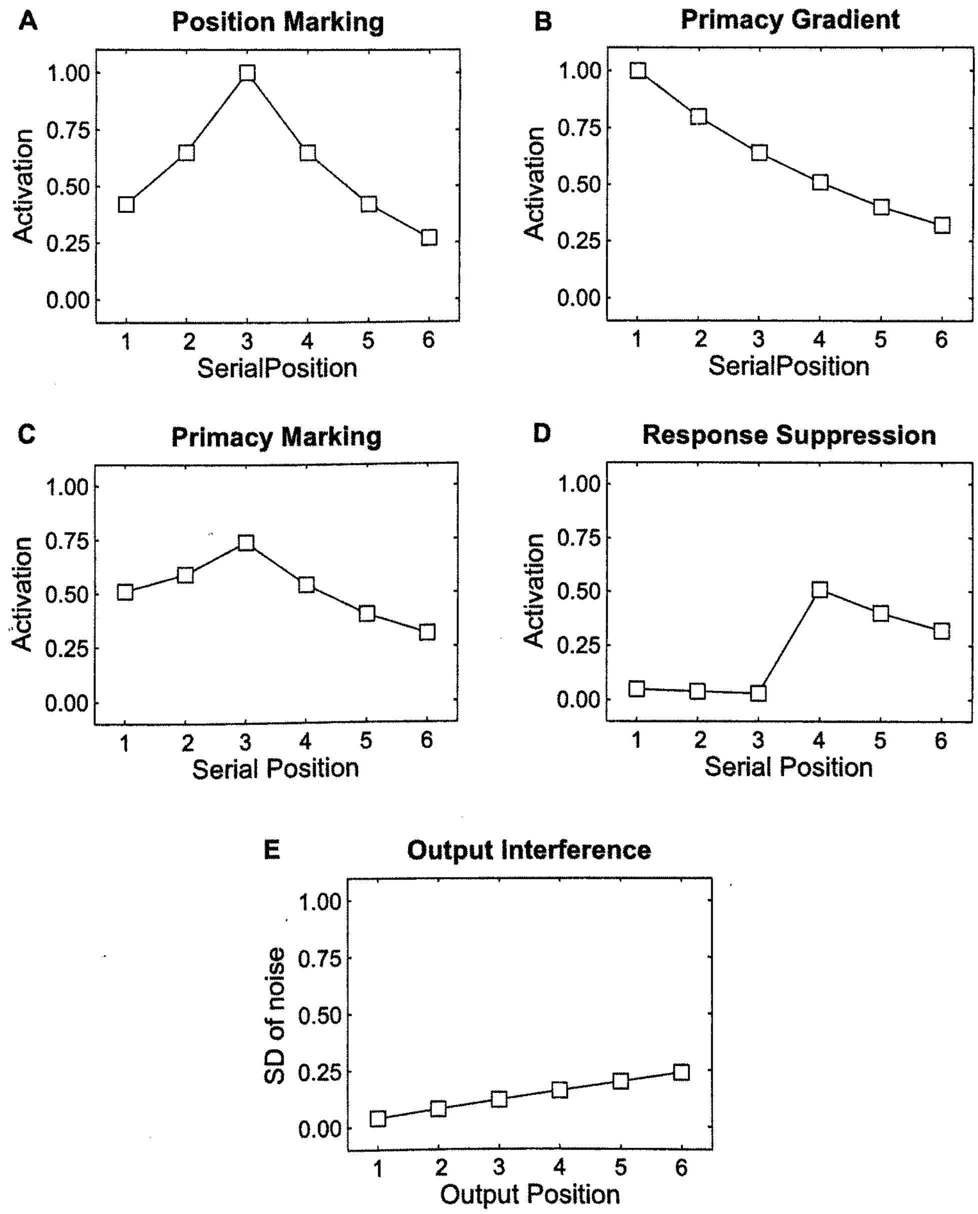

Figure 2-2 *Example starting activations for five representational principles: (A) position marking: activations shown are for the third output position, (B) primacy gradient: activations shown are for the first output position (C) Primacy gradient + position marking: activations shown are for the third output position, (D) response suppression: activations show a primacy gradient with suppression of the first three recalled items, (E) output interference: activations show the increase in noise across output positions.*

Output interference

Output interference was modelled by assuming that recall of an item added noise to the representations of yet to be recalled items. Accordingly, random Gaussian noise with a standard deviation ($0 < \delta < 1$) that increased as a function of output position was applied to the starting activations generated by the serial ordering principle(s) being modelled (e.g., position marking) and was determined by $\delta \times p$. Figure 2-1E shows the increase in the standard deviation of random Gaussian noise applied to the starting activations across output positions for a sequence of six-items when $\delta = .04$.

Model comparisons

The generic modelling architecture described above is used to test the transposition error latency predictions of models built from different combinations of the representational principles in Chapters 3, 4, and 5. The models examined in these chapters include: (1) position marking; (2) position marking + response suppression; (3) position marking + output interference + response suppression; (4) primacy gradient + response suppression; (5) primacy gradient + position marking + response suppression.

This framework is also employed to test the fill-in and infill error predictions of models built from different combinations of the representational principles in Chapter 8. The models examined in that chapter include: (1) position marking + response suppression, (2) primacy gradient + response suppression, (3) primacy gradient + position marking + response suppression. The implications of different formalisations of the primacy gradient (i.e., exponential versus linear) are also examined.

The intellectual basis for the specific combinations of representational principles examined across models is delineated in the chapters in which they appear. However, in general, combinations of principles were chosen based upon those commonly utilized in models of verbal short-term serial recall (see e.g., Table 1-1 of Chapter 1).

Forgetting in the models

Before describing how the above models will be evaluated and compared, in this section I consider the nature of forgetting in those models and how this relates to recent data exploring the mechanisms of forgetting from short-term memory. Contemporary theories of serial recall explain forgetting by recourse to a range of different explanatory principles. For example, the Burgess and Hitch (1999) model and primacy model (Page & Norris, 1998) appeal to the notion of time-based decay, according to which information is forgotten from short-term memory, because memory traces fade over time, thereby reducing the likelihood of correct remembering. The SIMPLE (Brown et al., 2007) model appeals to the notion of temporal distinctiveness, according to which the passage of time causes the memory traces of items to become less distinctive, and therefore less easily remembered. Other theories, such as C-SOB (Lewandowsky & Farrell, 2008), ascribe forgetting to interference from subsequent events.

In the modelling architecture described above forgetting is essentially attributable to interference. The random Gaussian noise added to the item activations during recall is tacitly taken to be the result of interference between the representations of list items stored in short-term memory. Importantly, no role is accorded to decay or temporal distinctiveness. The rejection of decay and temporal distinctiveness as explanatory principles is buttressed by the results of recent studies exploring the nature of forgetting from verbal short-term memory.

For example, Lewandowsky, Duncan, and Brown (2004) have provided evidence against decay. They trained participants to recall verbal lists at three different speeds (.4, .8, and 1.6 sec per item) in order to manipulate the amount of time preceding recall of the entire list (which varied from 6 to 10 sec). Participants were also required to engage in articulatory suppression during list recall to prevent rehearsal from being deployed to revivify decaying memory traces. Decay theories predict that memory performance should be best in the fast recall condition (.4 sec per item), because the memory traces of items will have had the least amount of time to undergo decay; performance should be worse in the medium recall condition (.8 sec per item), because the memory traces of items will have had more time to undergo decay; performance should be worse still, in the slow recall condition (1.6 sec per item), where the memory traces of items will have had the most time to undergo decay. However, the author's found that performance did not vary across recall conditions, at odds with the predictions of decay theories.

In a subsequent study, Oberauer and Lewandowsky (2008) directly pitted decay theory against interference theory. They manipulated total recall duration by varying the number of distracters that participants had to speak between recalling each list-item. Sometimes participants were required to utter a distracter word once (e.g., "*super*") between recall of each item, whereas sometimes they were required to utter the same distracter word three times (e.g., "*super, super, super*"). Performance in these two conditions was compared with a baseline condition without distracters. Decay theories predict that performance should be worse in the single distracter condition than the baseline condition, because the requirement to utter a single distracter between recall of each item will delay total recall time, thereby increasing the opportunity for forgetting due to decay. Decay theories predict performance should be worse still in the three distracter condition, because total recall time will be delayed even further by the requirement to utter three distracters, as opposed to one. Interference-based models, such as C-SOB (Lewandowsky & Farrell, 2008) also predict performance to be better in the baseline than the single distracter condition. This is because the requirement to utter a single distracter between recall of each item will interfere with the representations of the to-be-recalled items. Crucially however, C-SOB predicts that performance should not differ between the single distracter, and three distracter conditions, because repeating the same word three times generates the same representation as established by the first utterance, thereby adding no further interference. The results were consistent with the interference-based account offered by C-SOB.

That forgetting in short-term memory is the result of interference rather than decay is further underscored by a study by Lewandowsky, Geiger, and Oberauer (2008), which took the basic design of the above study, but incorporated a condition in which the words in the three distracter condition varied (e.g., "*January, February, March*"). Performance was worse in this condition, relative to when the distracters were repetitions of the same word (e.g., "*January, January, January*"). Interference-based models like C-SOB predict this pattern, because uttering three different words will lay down a unique representation of each distracter word, which will be more detrimental to recall than the single distracter representation lay down when the same distracter word is repeated.

Whilst the above studies have failed to provide support for decay, other studies have additionally failed to provide support for temporal distinctiveness. One major prediction of temporal distinctiveness models like SIMPLE is that items that are spaced widely apart in time are more distinctive, and hence should be better remembered than items that are spaced close together in time. Several studies have examined whether an item's degree of temporal isolation from its list neighbours is a determinant of its recall accuracy. For example, Lewandowsky and Brown (2005) presented participants with lists of items in which the length of the pause preceding, and following an item, systematically varied in an unpredictable fashion. They examined the recall accuracy of items as a function of the length of the pre- and post-item intervals and failed to observe an effect of temporal isolation. This finding has subsequently been replicated across several serial recall studies (e.g., Lewandowsky, Brown, Wright, & Nimmo, 2006; Nimmo & Lewandowsky, 2005, 2006; Parmentier, King, & Dennis, 2006)[2]. That people are insensitive to temporal isolation in serial recall, suggests that temporal distinctiveness is not the basis for forgetting from short-term memory.

In summary, although the source of forgetting from short-term memory is controversial, recent evidence supports the view that forgetting is attributable primarily to interference, with little or no role for decay or temporal distinctiveness, thereby justifying the exclusion of the latter explanatory principles from the modelling framework described above. Of course, these conclusions are based

[2] There is an exception to this general pattern. Specifically, temporal isolation effects have been observed in tasks in which the output order of items is unconstrained (e.g., Brown, Morin, & Lewandowsky, 2006; Lewandowsky, Nimmo, & Brown, 2008). However, I set aside these studies, since the default paradigm for examining short-term memory is forward serial recall. Moreover, Oberauer and Lewandowsky (2008) have shown that the SIMPLE temporal distinctiveness model is unable to accommodate the qualitative pattern of the data from their forgetting study (described above) providing further independent evidence for the rejection of temporal distinctiveness as an explanatory principle of short-term memory.

on studies of verbal short-term memory, and so it remains possible that decay or temporal distinctiveness may be viable explanatory principles in the nonverbal domain. However, concerning the latter principle at least, Parmentier et al. (2006) have failed to observe temporal isolation effects in a spatial serial recall task using unpredictable inter-item intervals, suggesting that temporal distinctiveness is also an unviable explanatory principle of nonverbal short-term memory[3].

Model evaluation and selection

Having outlined the common modelling architecture and the different models whose predictions are to be examined, the next section focuses on the pragmatics of how they are evaluated and compared, a process known as *model selection.* This process constitutes one of the greatest difficulties of the modelling enterprise. Accordingly, a number of important methods and criteria for model evaluation are introduced that are employed in subsequent chapters to distinguish between the models.

Parameter estimation and goodness-of-fit

One traditionally important criterion in cognitive psychology for evaluating a model is to determine its ability to fit observed data. The process of fitting models to data is known as *parameter estimation* and involves identifying the parameter values of a model that maximise the correspondence between its predictions and observed data. There are two methods of parameter estimation (Myung, 2003): least squares estimation (LSE) and maximum likelihood estimation (MLE). Each approach has its own tools for evaluating a model's data fitting ability. These are

[3] In contrast to the results of Parmentier et al. (2006) a recent study by Guerard, Neath, Surprenant, and Tremblay (2010) has shown that spatial serial memory is sensitive to effects of temporal isolation. However, this finding was restricted primarily to conditions in which the length of the inter-item intervals was predictable. Lewandowsky, Wright, and Brown (2007) have shown that under such conditions temporal isolation effects are likely attributable to a top-down selective encoding strategy, rather than temporal distinctiveness.

known collectively as measures of *goodness-of-fit* (GOF). In LSE a popular measure of GOF is the *root mean square deviation* (RMSD):

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{j=1} (obs_j - pred_j)^2}{N}} \tag{2-8}$$

Where obs_j is the value j observed, $pred_j$ is the corresponding value predicted by the model, and N is the number of data points. Large values of RMSD indicate that a model's fit to data is poor. Other least squares measures of GOF include the residual sums of squares and the coefficient of variation (R^2). In estimating parameters using LSE one seeks to minimise the value of the GOF measure.

In MLE the GOF measure is the likelihood of the data given the model parameters. In this approach, a model is construed as a parametric family of probability distributions (Myung, 2003), where each distribution is indexed by a different combination of model parameter values. The distributions represent the model's predictions about the probability of certain observations occurring, or not occurring. If one's aim is to test a model's prediction then a probability distribution is generated by choosing a specific combination of parameter values and data are collected to evaluate the prediction. Maximum likelihood estimation is the inverse of this process; the data have already been collected and the goal is to identify the parameter values most likely to have produced the data, if the model is correct.

Obtaining likelihood estimates depends upon specifying an appropriate likelihood function, the choice of which is based upon distributional assumptions underpinning the data. In this thesis, the data to be fitted typically consist of a confusion matrix containing the frequency of responses at each output position as a function of input position. Because the response frequencies for each output position have a multinomial distribution, the multinomial likelihood function is used:

$$L(\mathrm{X},\mathrm{P}) = N! \prod_{j=1}^{J} P_j^{\,Fj} / F_j! \tag{2-9}$$

Where L(X, P) is the likelihood of the data given the model parameters, F is a data vector composed of the frequency of responses for each of j categories, and P is a vector of the model's predicted probabilities that responses will be assigned to each category of F. When estimating parameters using maximum likelihood we seek model parameters that maximise L(X, P). However, because the values of L(X, P) will tend to be very small, it is computationally more convenient to use log-likelihoods. As there is a monotonic relationship between numbers and their logs, the parameters that maximise log-likelihood will be the same as the parameters that maximise likelihood. The multinomial log-likelihood function is defined as:

$$\ln L(\mathrm{X},\mathrm{P}) = \ln N! - \sum_{j=1}^{J} \ln F_j! + \sum_{j=1}^{J} F_j \cdot \ln P_j \qquad (2\text{-}10)$$

For the confusion matrix, it is necessary to calculate the value of ln$L$(X, P) for the data vector representing each output position, since each will have its own multinomial distribution. The resulting values are then summed to produce a joint multinomial log-likelihood estimate and it is this value that is maximised during MLE.

Given the availability of two methods for parameter estimation, which is preferable? The benefit of MLE is that it can be used for statistical hypothesis testing, since maximum likelihood is the cornerstone of many inferential statistical tests, including the chi-square and g-square tests, and numerous model selection tests, including the Akaike Information Criterion (Akaike, 1973) and the Bayesian Information Criterion (Schwarz, 1978), amongst others. The limitation of this approach is that sometimes it is not possible to specify an appropriate likelihood function for the data being fit. The benefit of LSE is that GOF measures such as RMSD are easy to compute and require minimal distributional assumptions. The latter means that they can be applied flexibly to a variety of data fitting scenarios, including conditions under which specification of an appropriate likelihood function is not possible. The drawback of LSE is that, as a general rule, it cannot be used for statistical hypothesis testing[4]. Thus, MLE is a preferred method for parameter estimation, but depending on the data being fit it is not always possible to use this approach. When an appropriate likelihood function is unavailable the only option is to use a least squares approach. Accordingly, in this thesis parameter estimation is based upon MLE whenever possible and LSE at all other times.

Optimization

Optimization refers to the systematic process by which parameter estimates are obtained. Usually it is carried out using an effective algorithm that searches through a model's parameter space to find the optimum solution. Optimization algorithms tend to work by minimizing a GOF function, which is the desired outcome when fitting models using LSE. However, when using MLE we seek to maximise the GOF function. The solution when using optimization procedures to obtain MLE estimates is to minimize the negative log-likelihood. For models composed of a single parameter a variety of optimization algorithms exist (see Lamberts, 2005, for a review). Because all the models considered in this thesis possess multiple parameters, these algorithms are not described here.

For multiple parameter estimation, the simplest method for optimization is to perform a grid search of a model's parameter space. This involves the setting up of grids, in which each point corresponds to a vector of model parameter values. The GOF function is evaluated at each of these points and the parameter vector with the best value represents the solution to the optimization problem. A major drawback with this approach is that it is computationally costly as the number of points to be evaluated increases exponentially with the number of model parameters. Accordingly,

---

[4] This claim comes with one important qualification. Namely, when the observations in the data are independent and are normally distributed with a common variance the minimised residual sums of squares corresponds to a maximum likelihood estimate (Myung, 2003). Under these circumstances a minimised residual sums of squares score can be converted to a maximum likelihood score, and used for statistical hypothesis testing in the same way as a maximum likelihood score obtained directly using MLE.

the approach is not adopted here for optimization purposes. However, it is used to evaluate model flexibility, as will be described later.

A better method for multiple parameter estimation, and the one adopted here, is the *simplex algorithm* (Nelder & Mead, 1965; see also Lagarias, Reeds, Wright & Wright, 1998). In this method, the solution to the optimization problem is located at the vertex of a geometric figure known as a simplex. The simplex is composed of as many points as there are model parameters, plus an extra point. Thus, for a two parameter model the simplex would correspond to a triangle. Each point of the simplex represents a vector of parameter values.

The method requires that the modeller initially specify a set of well-informed starting parameter values assumed to lie close to the global minimum of the GOF function. The algorithm proceeds by defining an initial simplex, based on these starting values, and the GOF function is evaluated at each point. The algorithm must then decide which of three possible measures will minimize the GOF function. One measure is to replace the worst point with a point reflected through the centre of the remaining points. If this point is better than the best current point then this is followed by a further expansion away from the worst point. A second measure is to shrink the simplex by contracting the value of the worst point. The third measure is to shrink the simplex by contracting multiple points towards the best point.

The vertices of the new simplex are then evaluated as before and the decision process is repeated. This continues iteratively as the simplex moves towards the minimum of the GOF function and away from the maximum. When this function can be minimized no further the parameter values located at the vertex of the final simplex with the smallest GOF value represent the algorithms solution.

Like all optimization algorithms, simplex often gets stuck in local minima, a problem that can be minimized by repeating the search process with many different starting points.

Limitations of goodness-of-fit

Although the data fitting competencies of a group of competitor models are necessary benchmarks for model selection, they are not sufficient (Pitt & Myung, 2002). This is because a model can provide a good fit to data even if it is a poor approximation of the cognitive process being modelled. The reason for this is that human data are inherently noisy, containing information about the cognitive process under study, but also noise arising due to experimental error (Pitt & Myung, 2002). When models are fit to data they invariably capture some of this experimental error, as well as the cognitive process of interest. A real danger when fitting models is the possibility of *overfitting*. This is the term used to refer to the fitting of a model that is too complex, for instance, because it has too many free parameters. This may enable the model to provide a superior fit than its rivals, but it does so by capturing noise that is irrelevant to the cognitive process under study. The cost of the model's superior fit is that the final parameter values will be arbitrary, resulting in the model generalizing poorly to circumstances other than the current data. Consequently, contemporary model selection techniques emphasise the role of additional sources of information when choosing between models, including information about their relative complexity, flexibility and generalizability. I take up discussion of these criteria next.

Flexibility and complexity

Two important and related criteria for model selection are flexibility and complexity (Myung, 2000; Myung, Balasubramaian, & Pitt, 2000). Flexibility refers to a model's ability to fit many different data patterns, some of which may relate to the cognitive process of interest and some of which may not. Flexibility is directly related to a model's complexity (Myung, 2000; Pitt & Myung, 2002), which refers to those properties of a model that govern its flexibility. One of these properties is the number of parameters a model possesses; a model with many parameters is more complex than a model with fewer parameters. A second property is the functional form of the model equation(s) governing how parameters are combined (Myung, 2000). Two models can possess the same number of parameters, but the functional form of one may be more complex than the other.

The complexity of a model is an important determinant of its data fitting ability. All things being equal, a more complex model will tend to provide a better fit than a simpler rival model, because it will have more degrees of freedom to capture noise in the data. Consequently, when comparing models that differ in their relative complexity it is important to examine their associated flexibility. This involves searching through a large portion of a model's parameter space and recording the different predictions generated. The simplest approach, and the one adopted in this thesis, is to explore a model's prediction via a grid search. This involves setting upper and lower boundaries and an increment value for each model parameter. The boundaries represent the range of values that a parameter can assume and the increment value specifies the interval values within this range to be examined. All factorial combinations of the parameter values are then explored by simulation, and the associated predictions for each parameter vector are recorded (see Farrell & Lewandowsky, 2004 and Oberauer & Lewandowsky, 2008, for illustrative examples).

The variability in a model's prediction can then be evaluated, by for example, plotting its predictions as a distribution by sorting and placing them into a number of bins, one of which might reflect the empirical outcome of interest. An overly flexible model might be expected to generate predictions falling into a number of different bins, with only a small proportion falling into the bin corresponding to the data pattern of interest. This would indicate that the model's predictions are not robust and that its ability to approximate the data pattern of interest is the result of too much flexibility. By contrast, a desirable model would be expected to predict the outcome of interest as its main data pattern, with less variability in its predictions.

A more sophisticated approach to parameter space analysis is known as parameter space partitioning (PSP) (Pitt, Kim, Navarro, & Myung, 2006; Pitt, Myung, Montenegro, & Pooley, 2008). In PSP, the modeller first specifies *a priori*, a number of alternative qualitative data patterns. For example, these might represent the different possible ordinal relationships between a set of experimental conditions, one of which corresponds to the data pattern seen empirically. An efficient search algorithm then explores the entire model parameter space, partitioning it into sub-regions corresponding to the different data patterns. This approach, although highly promising, is difficult to implement (requiring a sophisticated Markov chain Monte Carlo search algorithm) and is computationally costly, as the entire parameter space (rather than some sub-set of that space) must be mapped. Consequently it is not employed in the current thesis.

Generalizability

Generalizability is defined as a model's ability to fit all data relating to a cognitive process, as opposed to just the current data being modelled (Myung et al., 2005). Measures of generalizability incorporate information about a model's GOF to current data, in addition to information about its complexity. The dimension(s) of complexity considered depend upon the specific generalizability measure being employed, but GOF is always based upon a model's maximum likelihood value, obtained via MLE. The aims of such measures are to neutralize differences between the GOF values of rival models that emanate from the greater complexity of some models over others. Thus, by handicapping more complex models, model selection can commence on an even playing field.

Attention is drawn here to two specific measures of generalizability employed throughout this thesis; they are the Akaike Information Criterion (AIC; Akaike, 1973) and the Bayesian Information Criterion (BIC; Schwarz, 1978). Beginning with the former, the AIC is defined as:

$$\mathrm{AIC}_i = -2\ln L_i + 2v_i \qquad (2\text{-}11)$$

Where $\ln L$ is the maximum log-likelihood, v represents the number of free model parameters and i indexes the model among k competitors for which AIC is being computed. Equation 2-11 shows that the AIC rewards GOF and imposes a penalty that increases as a function of the number of free model parameters. When comparing models with AIC the one with the lowest value is preferred; this is the model with the minimum of free parameters that best explains the data.

A problem with comparing raw AIC values is that it can be difficult to determine whether differences in those values are meaningful, because they are not organised along a continuous measure. Fortunately, Wagenmakers and Farrell (2004) have prescribed a simple approach to transform these raw values into so called Akaike weights, which represent the conditional probabilities that each model should be selected. To obtain Akaike weights, one first subtracts the AIC value of the model with the smallest AIC from the AIC values of all other models:

$$\Delta_i(\mathrm{AIC}) = \mathrm{AIC}_i - \min \mathrm{AIC} \qquad (2\text{-}12)$$

In the next step, an estimate of the relative likelihood L of model i is obtained by the following transform:

$$L(M_i \mid \text{data}) \propto \exp\left\{-\frac{1}{2}\Delta_i(\text{AIC})\right\}, \tag{2-13}$$

Akaike weights are then calculated by normalising the relative model likelihoods (dividing each likelihood by the sum of all likelihoods):

$$w_i(\text{AIC}) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(\text{AIC})\right\}}{\sum_{k=1}^{K}\exp\left\{-\frac{1}{2}\Delta_k(\text{AIC})\right\}}, \tag{2-14}$$

The resulting AIC weights represent the conditional probabilities that each model should be chosen given the data and the set of competing models. The use of AIC weights can greatly facilitate model comparison, because the strength of evidence in favour of each model can be expressed on a continuous measure.

One limitation of the AIC is that its complexity measure fails to take into account the sampling variability of estimated parameters, which increases with the number of free model parameters. This can sometimes lead to the selection of an overly complex model. An alternative selection criterion that does take this variability into account is the BIC (Schwarz, 1978), which is defined as:

$$\text{BIC}_i = -2\ln L_i + v_i \log(n) \tag{2-15}$$

Where n is number of observations in the calculation of ln$L$. Like the AIC, the BIC rewards GOF, but penalizes a model not only on the basis of its number of free parameters, but also the sample size. Because the BIC uses two penalties in its complexity term, it punishes complexity more so than the AIC. As for AIC, raw BIC values can be converted into BIC weights by replacing the AIC values in equation 2-14 with BIC values instead.

Chapter summary

This chapter began by providing an introduction to mathematical modelling and outlining the virtues of this approach. The generic modelling framework employed in this thesis was then outlined and a number of models were prescribed for comparison based upon the combinations of representational principles commonly instantiated in models of verbal short-term serial recall. Finally, the methods and criteria used for model evaluation and selection were described. Subsequent chapters compare the models on the basis of these criteria in a bid to identify a preferred combination of principles for representing serial order in visual and spatial serial memory.

3

Modelling transposition latencies

Abstract

This chapter explores the transposition error latency predictions of five models and associated mechanisms for the representation of serial order using the dynamic competitive queuing response selection architecture delineated in Chapter 2. The results of initial simulations of the models demonstrate that the different mechanisms for representing serial order can each explain the core data patterns of serial memory, but make qualitatively different predictions concerning the dynamics of transposition errors. Quantitative fits of a sub-set of the models to verbal serial recall data taken from Farrell and Lewandowsky (2004) buttress the notion that the serial order of a sequence of verbal items is represented by a mechanism combining a primacy gradient, positional marking, and response suppression.

Introduction

As noted in Chapter 1, recent studies have shown that visual and spatial serial memory exhibit a number of constraints that were previously thought to be unique features of verbal serial memory. These constraints include the characteristic form of the accuracy and response latency serial position curves, effects of sequence length, and the locality constraint on transposition errors. However, many different combinations of the representational principles delineated in that chapter can accommodate these constraints. The clearest example of this comes from the computational modelling of Farrell and Lewandowsky (2004) who showed that models of serial recall built from different combinations of a primacy gradient, position marking, response suppression, and output interference can each explain these core serial memory data. Consequently, a preferred combination of principles for representing serial order in the visuospatial domain is currently lacking.

Although models built from different combinations of the representational principles cannot be differentiated on the basis of the abovementioned constraints, Farrell and Lewandowsky (2004) have shown that adjudication between the models becomes possible by considering their predicted *latency-displacement functions*. To explain, a latency-displacement function (LDF) plots the mean response latencies of transposition errors as a function of *transposition displacement* (note that response latency refers to the time elapsed between the current and the previous response, except in the case of the first response for which the latency is the time elapsed from the onset of the recall phase). Recall from Chapter 1 that transposition displacement refers to the numerical difference between an item's presentation (input) and recall (output) positions. Transpositions with negative displacement values are known as *anticipation errors* and correspond to items recalled ahead of their correct positions. For example, a –4 displacement refers to an item recalled four positions before its correct position. Transpositions with positive displacement values are known as *postponement errors* and correspond to items recalled following their correct positions. For example, a +2 displacement corresponds to an item recalled two positions after its correct position. A displacement value of zero corresponds to items recalled in their correct positions. Thus, a LDF plots the same information as a transposition gradient, but with mean latency rather than proportion of responses as the dependent measure of interest.

Farrell and Lewandowsky (2004) used the dynamic modelling architecture delineated in Chapter 2 to compare the predicted LDFs of four models built from different combinations of principles for representing serial order: (1) position marking (PM); (2) position marking in conjunction with response suppression (PM+RS); (3) position marking in conjunction with output interference (PM+OI); and (4) a primacy gradient in conjunction with response suppression (PG+RS). Although the models generated qualitatively similar accuracy and response latency serial position curves and transposition gradients (see Figure 3-1), there was considerable heterogeneity in their corresponding LDFs (see Figure 3-2). The PM model predicted a symmetrical function in which responses for correct recalls were fastest and responses to anticipations and postponements became longer as the absolute transposition displacement value increased. The PM+RS and

PM+OI[1] models both produced partially symmetric functions in which the slopes of the functions for postponements were shallower than the slopes of the functions for anticipations. Finally, the PG+RS model produced an asymmetric function in which responses for anticipations became longer as the absolute transposition displacement value increased, whereas responses for postponements became accelerated as the absolute transposition displacement value increased.

Across three experiments involving keyboard timed serial recall of verbal items, Farrell and Lewandowsky (2004) consistently found that latency is a negative function of transposition displacement, as predicted by the PG+RS model. However, one noteworthy discrepancy between the observed LDFs and those predicted by the model was that the slopes of the functions for postponements were either flat or weakly positive in the former case, whereas they were consistently negative in the latter case. Thus, although the PG+RS model provided the best account of the LDFs of the competitor models considered, that account was nevertheless incomplete. In a subsequent article, Lewandowsky and Farrell (2008) reported simulations of a further model that augmented the PG+RS model with a set of positional markers (PG+PM+RS). That model was successfully able to accommodate the overall negative relationship between transposition latency and transposition displacement, in addition to the flattening of the slope of the function for postponements (see Figure 3-2), thereby providing support for a mechanism combining three representational principles.

The objectives of this chapter are two-fold. The first objective is to replicate the modelling of transposition latencies reported by Farrell and Lewandowsky (Farrell and Lewandowsky, 2004; Lewandowsky & Farrell, 2008) by generating qualitative predictions for the models described above using the same published model parameter settings employed by these authors. The purpose of these simulations is to verify that the implementation of the dynamic modelling architecture and

[1] Figures 3-1 and 3-2 do not show the predictions of the PM+OI model, but instead show the predictions of a generalised version of this model in which position marking and output interference has been augmented with response suppression (PM+OI+RS). Nevertheless, the qualitative predictions of the PM+OI+RS model shown in these figures are the same as those described for the PM+OI model in the main text.

representational principles of Chapter 2 behaves as expected, based upon the modelling of Farrell and Lewandowsky. The transposition error latency predictions of the models obtained from these initial simulations, which are indeed consistent with those obtained by Farrell and Lewandowsky, are tested in Chapters 4 and 5 using serial reconstruction of visual and spatial sequences, respectively. The second objective is to provide quantitative fits of the PG+RS and PG+PM+RS models to a sub-set of the Farrell and Lewandowsky (2004) data. This fitting exercise is important, because although Farrell and Lewandowsky (2004) reported quantitative fits of the PG+RS model to their data, Lewandowsky and Farrell (2008) only reported qualitative predictions of the PG+PM+RS model. Thus, it remains to be seen whether the latter model provides a better account of the empirical pattern than its more parsimonious rival model when it is directly fit to behavioural data. Notably, in order to facilitate across domain comparisons it is important to identify the preferred combination of principles for explaining LDFs in verbal serial memory, before examining LDFs in visual and spatial serial memory.

Qualitative modelling of transposition latencies

Simulation details

The computational modelling of transposition latencies was conducted using the dynamic competitive queuing response selection network architecture outlined in Chapter 2. The generic model parameters, which includes the response threshold (T), excitatory (w^+) and inhibitory (w^-) weights and the mean (μ) and standard deviation (σ) of Gaussian noise, were set to constant values of 1.0, 1.1, –0.1, 0, and .04, respectively, and were based on values used in the simulations of Farrell and Lewandowsky (2004). The error latency predictions for five models were contrasted: (1) position marking (PM); (2) position marking and response suppression (PM+RS); (3) position marking and output interference and response suppression (PM+OI+RS)[2]; (4) primacy gradient and response suppression (PG+RS); (5) primacy gradient and position marking and response

[2] This model was not examined by Farrell and Lewandowsky (2004), who instead focused on a model combining position marking and output interference.

suppression (PG+PM+RS). These models capture the range of representational mechanisms employed across most contemporary theories of verbal short-term memory (see Figure 1-1 of Chapter 1).

The parameter values chosen for each of the representational principles to generate the initial predictions were the same as those reported in Farrell and Lewandowsky (2004). Specifically, for position marking, the weighting parameter for the position markers (λ) was set to 1, whilst the parameter controlling the distinctiveness of the position markers (ϕ) was set to .65. For the primacy gradient, the weighting parameter for the node reflecting the first input position (a_1) was set to .6, whilst the parameter controlling the steepness of the gradient (γ) was set to .85 (note that the version of the primacy gradient employed here was the exponential). For the combination of a primacy gradient and position marking, the parameter controlling the attentional weight given to the two dimensions of ordering (ω) was set to .5. The response suppression (α) and output interference (δ) parameters were set to values of .95 and .04, respectively. Predictions were generated for each model using 50,000 simulation trials of six-item sequences.

Simulation results

The predictions of the models for accuracy and response latency serial position curves and transposition gradients can be observed in Figure 3-1. These predictions can be compared with those presented in Figure 3 of Farrell and Lewandowsky (2004) and Figure 1 (leftmost and central panels) of Lewandowsky and Farrell (2008). Considering first the accuracy serial position curves (Figure 3-1A), all models predict a primacy and a recency effect, with poorer performance at medial serial positions, consistent with the data from studies of verbal serial recall (e.g., Crowder, 1972; Jahnke, 1965, 1969, Murray, 1967), as well as visual (e.g., Avons, 1998; Smyth et al., 2005) and spatial (e.g., Jones et al., 1995; Smyth & Scholey, 1996) serial reconstruction. However, the PM model predicts a symmetrical serial position curve that is at odds with the empirical data. This behaviour materialises because of the symmetrical manner in which positional uncertainty is

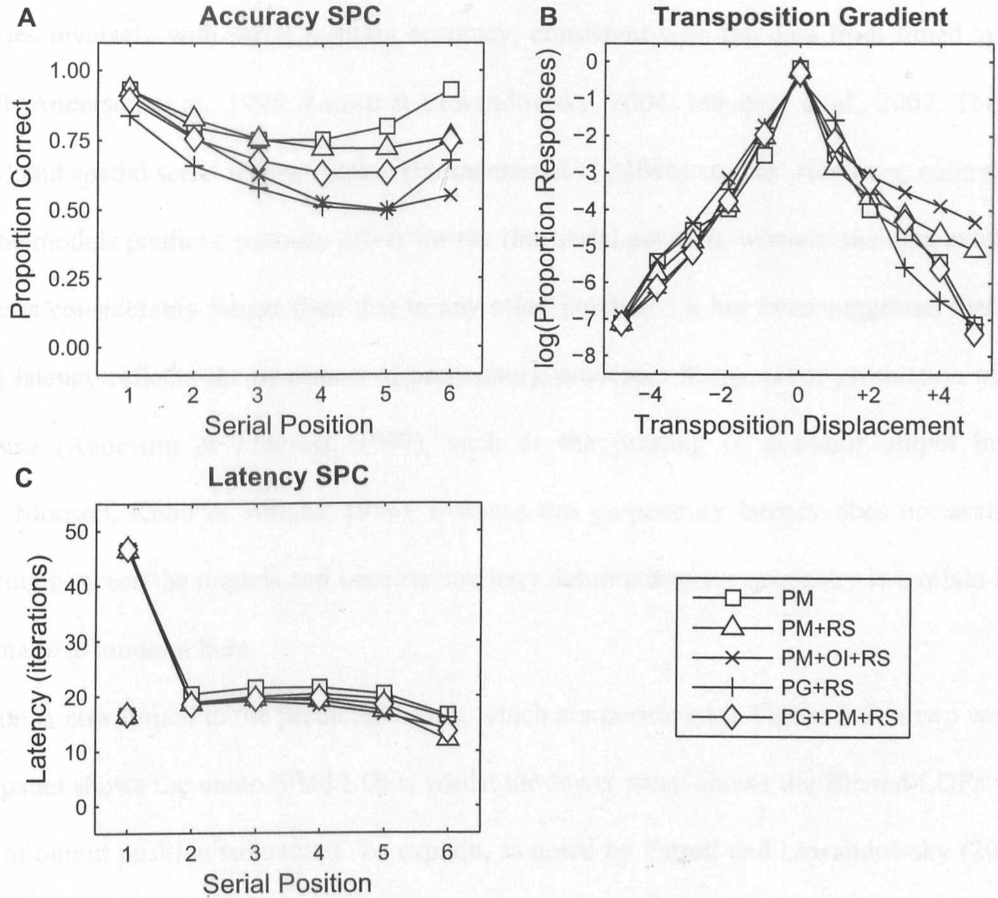

Figure 3-1 *Predicted accuracy serial position curves (A), transposition gradients, (B), and latency serial position curves (C), for five models of serial order. Note—PM = position marking; PM+RS = position marking and response suppression; PM+OI+RS = position marking, output interference, and response suppression; PG+RS = primacy gradient and response suppression; PG+PM+RS = primacy gradient, position marking, and response suppression.*

distributed in the position markers. The remaining models all exhibit more realistic asymmetric serial position curves characterised by greater primacy than recency.

Considering now the predictions for transpositions (Figure 3-1B), all models predict the three commonly observed features of transposition gradients (Farrell and Lewandowsky, 2004). First, the proportion of responses peaks at displacement value zero, reflecting that correct responses are most prevalent. Second, the proportion of transpositions is greatest for displacements with an absolute value of one, and decreases gradually thereafter – the *locality constraint* (Henson, 1996). Third, the transposition gradient is approximately symmetrical; that is, the proportion of transpositions for each absolute displacement is approximately the same for anticipations and postponements.

Turning to the response latency serial position curves (Figure 3-1C), all models predict that latency varies inversely with serial position accuracy, consistent with the data from timed verbal serial recall (Anderson et al., 1998; Farrell & Lewandowsky, 2004; Maybery et al., 2002; Thomas et al., 2003) and spatial serial reconstruction (Parmentier et al., 2006) studies. However, contrary to that data, the models predict a primacy effect for the first serial position, whereas the data exhibit a latency that is considerably longer than that to any other position[3]. It has been suggested that this long initial latency reflects the operation of preparatory processes that precede production of the first response (Anderson & Matessa, 1997), such as the priming of a motor output buffer (Sternberg, Monsell, Knoll & Wright, 1978). Because this preparatory latency does not assist in discriminating between the models and because ancillary assumptions are necessary to explain it no attempt is made to model it here.

Attention is now turned to the predicted LDFs, which are portrayed in Figure 3-2 in two ways[4]. The upper panel shows the unmodified LDFs, whilst the lower panel shows the filtered LDFs with the effects of output position subtracted. To explain, as noted by Farrell and Lewandowsky (2004), one factor that is confounded with transposition displacement is the output position of transposed items. Transpositions that are anticipation errors tend to occur at early output positions, whereas transpositions that are postponement errors tend to occur at late output positions. This is problematic, because latencies for early output positions will tend to be longer than latencies for late output positions, thereby artificially inflating latencies for anticipations and artificially accelerating latencies for postponements. Following Farrell and Lewandowsky (2004), to deal with this confound the modified LDFs have had these effects removed by subtracting the mean latency

[3] To facilitate graphical comparison of the predicted latency serial position curves with the data presented in subsequent chapters Figure 3-1B includes the predicted latency of each of the models for the first output position, plus additional points reflecting those latencies with a 30-iteration constant added.

[4] Repetition errors (both occurrences of the repeat) were excluded from the LDFs shown in Figure 3-2 because the models incorporating response suppression predict that the error latencies for repeated and non-repeated items behave differently.

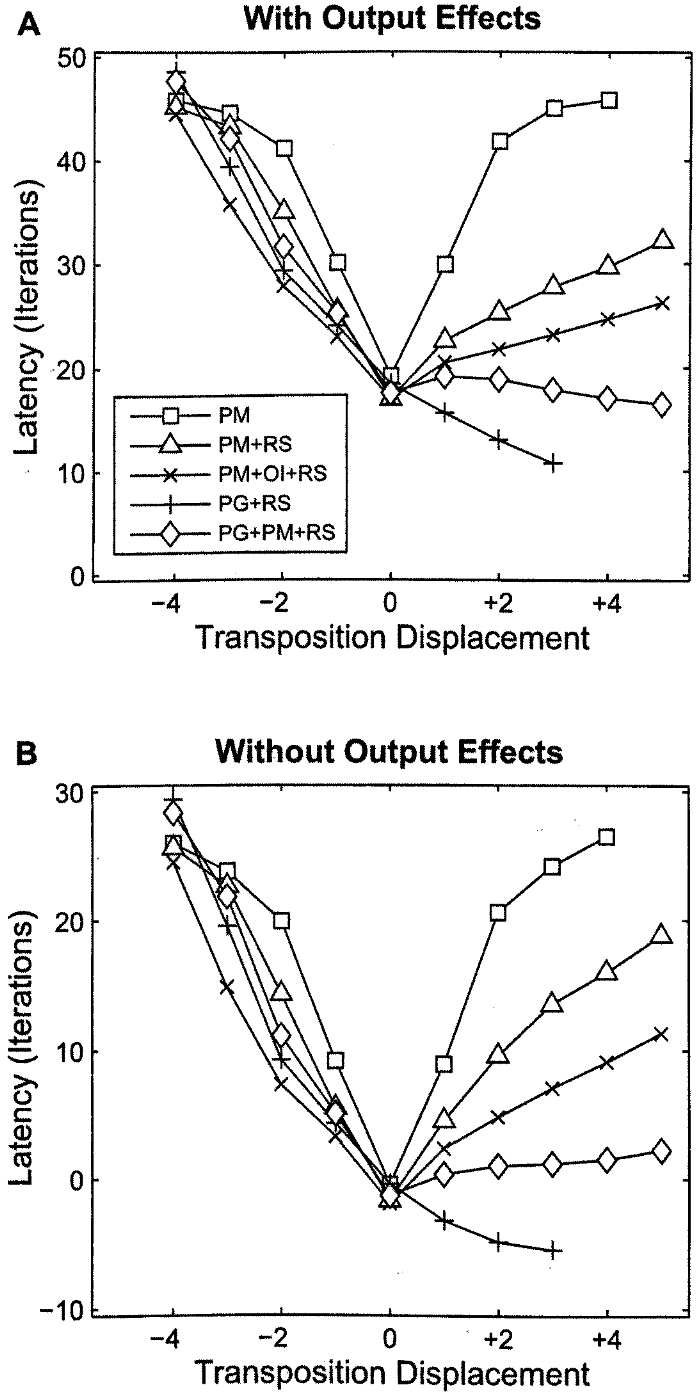

Figure 3-2 *Predicted latency-displacement functions for five models of serial order with the effects of output position on latency included (A) and excluded (B).*

for each output position from the individual latencies at corresponding output positions. Inspection of Figure 3-2 reveals that the predictions of the models for the unmodified and modified LDFs are qualitatively very similar.

The first thing to note about these LDFs is that all the models predict that the slope of the function for anticipations is negative, indicating that latencies for anticipations increase as a function of transposition displacement. This is because whether serial order is represented by a primacy gradient, position marking, or a combination of the two, an anticipation error involves a weakly activated item being recalled from amongst a set of stronger competitors from earlier input positions. The further an item is anticipated, the greater the number of competitors it must overcome, and consequently the longer it will take for that item to win the output competition. It follows from these similarities that the models cannot be differentiated on the basis of their predicted anticipation slopes.

Adjudication becomes possible, however, when one considers the model's predicted postponement slopes. Considering first the position marking model; this model predicts that latencies for postponements increase as a function of displacement in the same way as they do for anticipations. This is a natural consequence of the symmetrical manner in which positional uncertainty is distributed in the positional markers. The effect of augmenting position marking with response suppression is to reduce the slope for postponements, because the suppression of recalled items reduces the number of competitors at late output positions. Since postponement errors will tend to occur towards the end of the sequence, the reduced competitor set results in shorter latencies for these errors. Combining output interference with the combination of position marking and response suppression engenders a further reduction in slope for postponements, because the impact of accumulating output interference is to raise the activations of item nodes taking them closer to the response threshold. In departure from the above, the model combining a primacy gradient with response suppression predicts a negative postponement slope in which latencies for postponements become accelerated with increasing positive displacements. This is because a postponement error involves a strongly activated item being recalled from amongst a set of weaker competitors from later input positions. The longer an item is postponed the greater the disparity will be between its activation and that of its weaker rivals enabling it to quickly suppress those items through lateral inhibition and win the output competition. Finally, the model combining a primacy gradient with position marking and response suppression predicts a flat postponement slope when the effects of output position are included in the LDF (Figure 3-2A), and a shallow positive postponement slope (Figure 3-2B) when those effects are excluded from the LDF (the latter prediction is most relevant, as all LDF analyses in this thesis involve removal of the contaminating effects of output position). The flat/shallow-positive postponement slope predicted by this model arises because the impact of adding the primacy gradient and position marking activations is to flatten the component of the position marking activations representing positional uncertainty with respect to the beginning of the sequence.

Quantitative modelling of transposition latencies

The qualitative predictions of the models obtained from the initial simulations reported above are consistent with those reported by Farrell and Lewandowsky (Farrell and Lewandowsky, 2004; Lewandowsky & Farrell, 2008) in every way. Having verified that the implementation of the dynamic modelling architecture and representational principles of Chapter 2 behave as expected, based upon the initial modelling of these authors, this section now reports quantitative fits of the PG+RS and the PG+PM+RS models to data taken from Farrell and Lewandowsky (2004)[5]. These authors reported quantitative fits of four models to data from their Experiment 3, in which participants engaged in keyboard timed serial recall of ungrouped and temporally grouped nine-item verbal sequences composed of digits. Those data have been re-plotted in Figures 3-3 and 3-4, which show the accuracy and response latency serial position curves and transposition gradients (Figure 3-3), and the LDFs with the effects of output position subtracted (Figure 3-4). Farrell and Lewandowsky (2004) fitted their models to data from the ungrouped condition and concluded that the PG+RS model was the only one to predict the qualitative pattern observed in the empirical LDFs. However, as noted earlier, the predictions of the PG+RS model were not entirely consistent with the empirical data. Specifically, the PG+RS model predicted a negative postponement slope, whereas the data shown in Figure 3-4 exhibit shallow positive postponement slopes. Subsequently,

[5] Thanks go to Simon Farrell for making these data available and for his permission to use them for the simulations reported here.

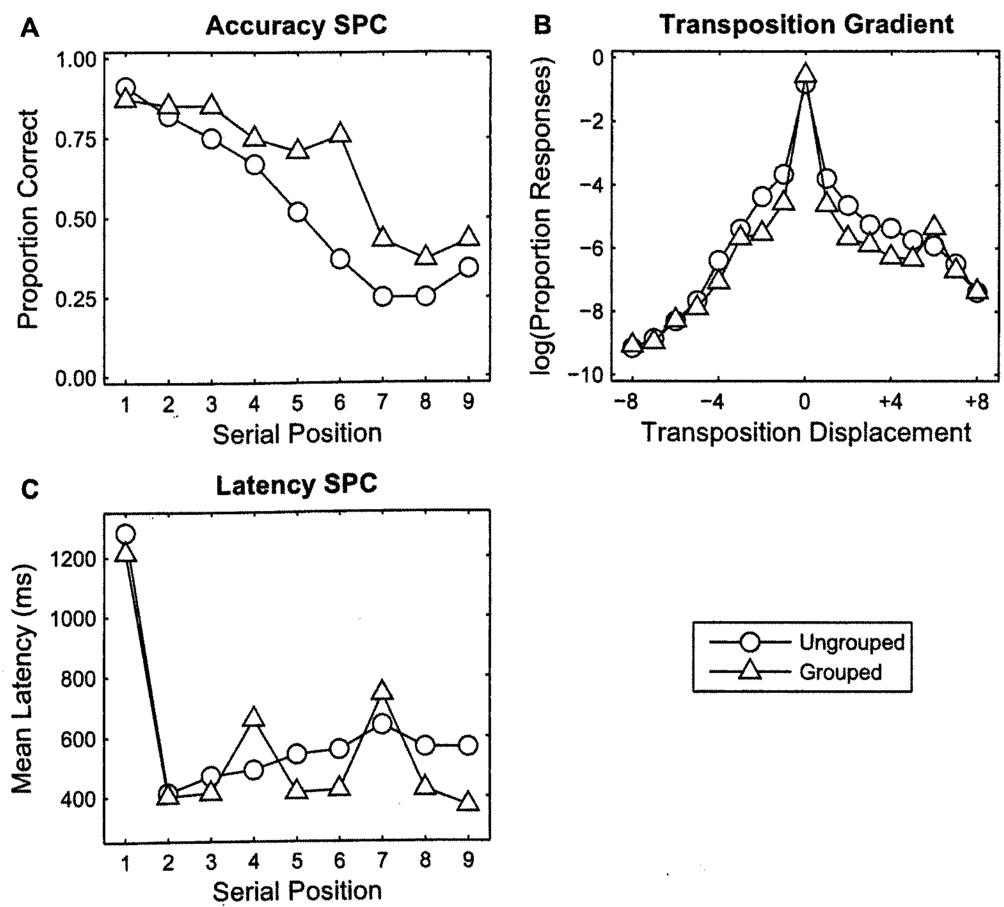

Figure 3-3 *Serial recall performance measures for ungrouped and temporally grouped verbal sequences for Experiment 3 of Farrell and Lewandowsky (2004). Panels show accuracy serial position curves (A), transposition gradients (B), and latency serial position curves, (C).*

Lewandowsky and Farrell (2008) showed that a model in which the basic primacy gradient and response suppression mechanism was complemented with a set of positional markers (PG+PM+RS) was able to capture this empirical outcome. Nevertheless, the authors only presented qualitative predictions of the model so it remains to be seen whether it maintains the same predictions when its parameters are estimated directly from the behavioural data.

Simulation details

To address this issue, the PG+RS and PG+PM+RS models were fit to the same target data using the same procedure as that employed by Farrell and Lewandowsky (2004). Specifically, the models were fit to the data of individual participants (26 in total) and the model predictions for

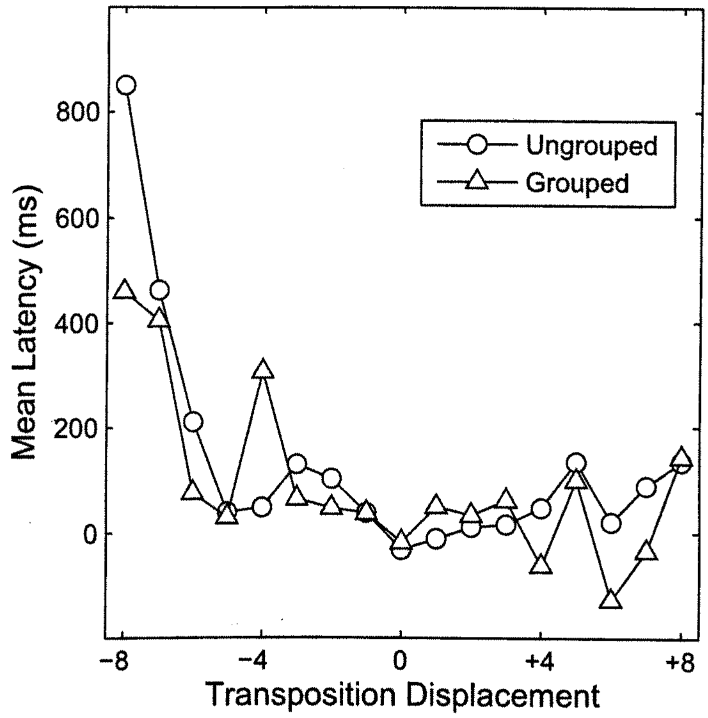

Figure 3-4 *Latency-displacement functions for ungrouped and temporally grouped verbal sequences for Experiment 3 of Farrell and Lewandowsky (2004).*

individual participants were then averaged to give aggregate predictions. The modelling procedure was the same as that used for the first set of simulations, except that model parameter values were varied using the simplex algorithm (Nelder & Mead, 1965), in order to minimize the root mean square deviation (RMSD) between the data and the prediction of each model (summed across accuracy serial position curve and transposition gradient), for each individual participant.

The parameters that were free to vary during the fitting process for the PG+RS model were the steepness of the primacy gradient (γ), and the standard deviation of noise (σ). For the PG+PM+RS model, these parameters were the distinctiveness of the position markers (ϕ), the steepness of the primacy gradient (γ), and the standard deviation of noise (σ). The parameters that were held constant for both models included the amount of response suppression ($\alpha = .95$), and the weighting parameter for the primacy gradient ($a_1 = 1$). The weighting parameter for the position markers was also held constant for the PG+PM+RS model ($\lambda = 1$), as was the attentional weight given to the primacy gradient and position markers ($\omega = .5$).

In summary, the number of free model parameters was two for the PG+RS model and three for the PG+PM+RS model. To increase the chances of finding the global minimum of the goodness-of-fit functions parameter estimates were obtained using multiple starting points for the search algorithm. These points were chosen by selecting two values for each free parameter, and then factorially crossing these to create a grid of starting values. Each parameter vector explored by the search algorithm involved 10,000 model simulation trials.

Simulation results

The minimized RMSD values for the fits of the models to the data for each participant can be scrutinized in Table A1-1 of Appendix 1. The average RMSD was 0.09 for the PG+RS model compared to a value of 0.08 for the PG+PM+RS model. Thus, the PG+PM+RS model provided only a slight improvement in fit to the accuracy serial position curves and transposition gradients, and at the expense of an additional free parameter. Of course, this outcome is not surprising given that the initial simulations reported above have already shown that the two models are capable of generating qualitatively similar serial position curves and transposition gradients. Adjudication between the models can only be accomplished by comparison of their predicted LDFs. Before considering these, I first describe the predictions of the models for the core serial memory performance measures.

The predictions of the models for accuracy serial position curves, transposition gradients, and latency serial position curves, averaged across fits to individual participants, are shown in Figure 3-5. Starting with the serial position curves for accuracy (Figure 3-5A), both models predict an extensive effect of primacy coupled with a restricted effect of recency, as seen empirically (Figure 3-3A). Both models also predict an asymmetric transposition gradient characterised by more postponement than anticipation errors (Figure 3-5B), in accordance with the data (Figure 3-3B).

Turning to the response latency serial position curves (Figure 3-5C), these have had a constant 40 iterations added to the mean response latencies for the first output position to facilitate graphical comparison with the data. As can be seen both models predict that latencies increase across output positions, consistent with the data (Figure 3-3C). In summary, when the models are contrasted on

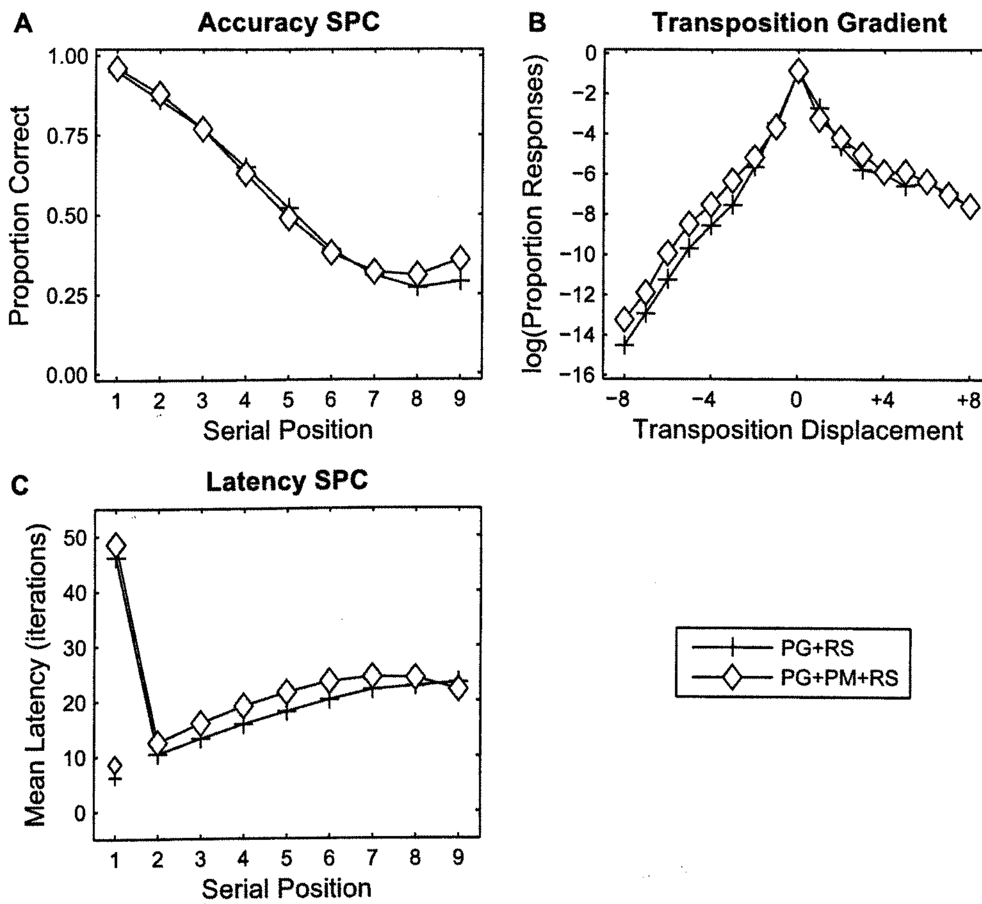

Figure 3-5 *Fits of the PG+RS and PG+PM+RS models to the ungrouped condition of Experiment 3 of Farrell and Lewandowsky (2004). Panels show accuracy serial position curves (A), transposition gradients (B), and latency serial position curves (C).*

the basis of their predictions for the core serial memory performance measures the more parsimonious PG+RS model fares equally well as the PG+PM+RS model.

Notwithstanding this correspondence between the models, their predicted LDFs shown in Figure 3-6 tell a different story. These functions, which have been averaged across fits to individual participants, were generated by the same best fitting parameter values that generated the predictions shown in Figure 3-5. They have had the effects of output position removed by calculating the predicted mean latency for each output position, for the fits to each participant's data, before subtracting these from the individual predicted latencies at corresponding output positions. It is apparent that although both models predict a negative slope for anticipations they generate qualitatively different predictions concerning the slope for postponements. Specifically, the PG+RS

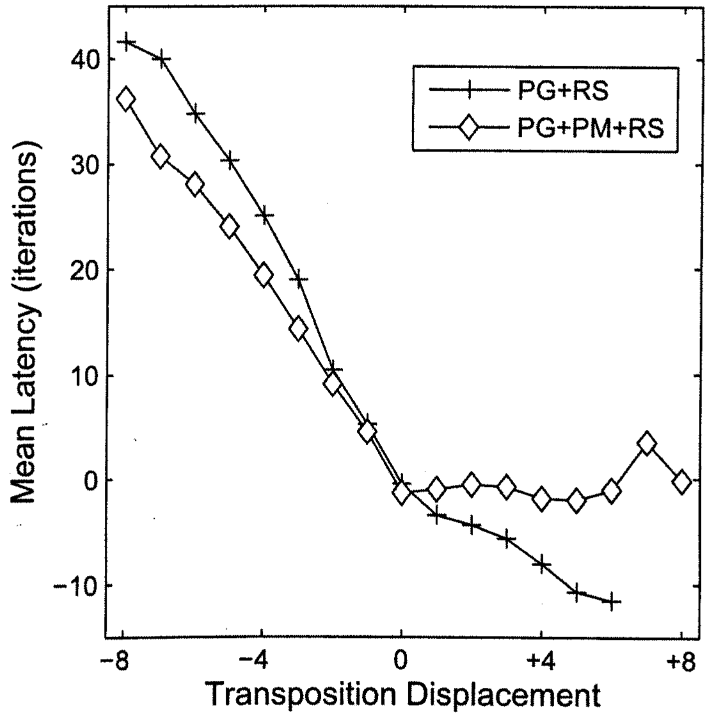

Figure 3-6 *Latency-displacement functions predicted by the PG+RS and PG+PM+RS models after fitting to the accuracy serial position curves and transposition gradients of the ungrouped condition of Experiment 3 of Farrell and Lewandowsky (2004)*

model predicts a steeply negative postponement slope, whereas the PG+PM+RS model predicts a flat postponement slope. It is the prediction of the latter model that is most consistent with the empirical pattern portrayed in Figure 3-4. Thus, the outcomes of the quantitative model fitting exercise harmonize well with the qualitative predictions presented initially, and suggest that the basic primacy gradient and response suppression mechanism must be complemented with a set of positional markers to provide a complete account of LDFs in verbal serial memory.

General discussion

This chapter examined the transposition error latency predictions of five alternative models and associated mechanisms for the representation of serial order, before providing quantitative fits of two of those models to verbal serial recall data taken from Farrell and Lewandowsky (2004). Below I consider the core outcomes of the two sets of simulations in turn.

The predictions of the models generated from the initial simulations are consistent with those reported by Farrell and Lewandowsky (Farrell and Lewandowsky, 2004; Lewandowsky & Farrell, 2008). They confirm that the different combinations of principles for representing serial order cannot be differentiated on the basis of their predicted accuracy and latency serial position curves and transposition gradients – arguably the core data patterns explored in visual and spatial serial memory to date (latency serial position curves notwithstanding). The results of these simulations therefore provide qualified support for the claim made in Chapter 1 that the empirical constraints so far examined in visual and spatial serial memory cannot be attributed to a unique combination of principles for representing serial order. Nevertheless, the same simulations showed that the different representational mechanisms could be theoretically distinguished by considering their predicted LDFs.

The outcomes of the quantitative fitting of the PG+RS and PG+PM+RS models to the data of Farrell and Lewandowsky (2004) revealed that both models provided equally good accounts of the accuracy and latency serial position curves and transposition gradients (notwithstanding the slightly smaller mean RMSD value for the PG+PM+RS model). Nevertheless, the models predicted qualitatively different LDFs. Specifically, although both models predicted negative anticipation slopes, the PG+RS model predicted a negative postponement slope, whereas the PG+PM+RS model predicted a flat postponement slope. It is the prediction of the latter model that is most compatible with the empirical data. The results of the quantitative fitting of models therefore identify a primacy gradient of activation, position marking, and response suppression as a preferred combination of principles for representing serial order in verbal serial memory.

Chapter summary

This chapter examined the transposition error latency predictions of five alternative models and associated mechanisms for the representation of serial order. The results of quantitative fits of a sub-set of these models to data taken from Farrell and Lewandowsky (2004) suggest that the serial order of a sequence of verbal items is represented by a competitive queuing system equipped with a primacy gradient of activation, associations between items and positional markers, and suppression of recalled items. The next two chapters test the error latency predictions of the five models using serial reconstruction of visual (Chapter 4) and spatial (Chapter 5) sequences, in order to identify a preferred combination of principles for representing serial order in visual and spatial serial memory.

4

Transposition latencies in visual serial memory

Abstract

The aim of this chapter was to identify a preferred combination of principles for representing serial order in visual short-term memory. In service of this goal, the error latency predictions of the five mechanisms for representing serial order presented in Chapter 3 were tested across six experiments, using a task requiring serial reconstruction of sequences of unfamiliar faces. The experiments consistently revealed that transposition latency is an overall negative function of transposition displacement, but with a shallow positive slope for postponement errors. This empirical outcome is consistent with the error latency prediction of a model in which serial order is represented by a primacy gradient of activation, associations between items and positional markers, and suppression of recalled items.

Introduction

The preceding chapter examined the error latency predictions of five alternative models and mechanisms for representing serial order. Fits of two of the models to verbal serial recall data taken from Farrell and Lewandowsky (2004) suggest that the serial order of a sequence of verbal items is represented by a primacy gradient of activation, associations between items and positional markers, and suppression of recalled items. The purpose of the present series of experiments was to identify a preferred combination of principles for representing serial order in visual short-term memory, via an empirical examination of the latency-displacement functions (LDFs) underpinning a visual serial reconstruction of order task.

As noted in Chapter 1, initial studies of serial position memory in the visual domain showed that it is characterised by a serial position curve that is different in appearance to that documented using verbal stimuli. Recall that Philips and Christie (Philips, 1983; Philips & Christie, 1977a, 1977b) examined serial position curves for sequences of matrix patterns composed of black and white squares presented from a constant spatial location, using a series of item recognition probe tests. In departure from the serial position curves documented using verbal items they observed a pronounced recency effect coupled with flat pre-recency performance. This empirical outcome, which was replicated independently by subsequent authors (Broadbent & Broadbent, 1981; Walker et al., 1991), was used to argue that serial position curves in verbal and visual short-term memory are mediated by modality specific mechanisms that operate according to fundamentally different principles.

Avons (Avons, 1998; Avons & Mason, 1999) subsequently argued that different serial position curves observed employing verbal and visual stimuli might actually be the consequence of the use of different recall methodologies in the two domains. He noted that studies showing primacy-dominant serial position curves with verbal stimuli used order-sensitive recall tasks, such as serial recall and serial reconstruction, whereas studies of visual memory showing primacy-absent serial position curves used order-insensitive tasks, such as item recognition. Avons tested this hypothesis by examining visual memory using serial reconstruction of sequences of randomly filled matrix patterns. In contrast to the item recognition studies of Philips and Christie, Avons observed effects of extended primacy and restricted recency of the accuracy serial position curves, as well as other canonical effects of serial memory, including detrimental effects of increases in sequence length, a greater incidence of order errors than item errors, a locality constraint on order errors, in addition to deleterious effects of the visual similarity of items reminiscent of the phonological similarity effect documented in the verbal domain (e.g., Baddeley, 1966; Conrad & Hull, 1964). That these constraints are not specific to the visual stimuli employed was confirmed more recently in studies by Ward et al. (2005) and Smyth et al. (2005), who replicated these findings using serial reconstruction of sequences of unfamiliar human faces as stimuli.

These studies demonstrate that when judgements of serial order are required, visual short-term memory exhibits a number of constraints previously thought to be unique characteristics of verbal short-term memory. Nevertheless, as noted in Chapter 1, these constraints can be accommodated by a variety of different mechanisms for representing serial order. The intellectual basis for this claim comes from the computational modelling work presented in Chapter 3 demonstrating that different combinations of a primacy gradient, position marking, response suppression, and output interference can each accommodate the core effects of memory for serial order. Consequently, a preferred combination of principles for representing serial order in visual memory is currently indefinable. The current chapter attempts to plug this evidential gap by evaluating the LDFs underpinning a task involving serial reconstruction of sequences of unfamiliar faces, with reference to the LDFs predicted by the models presented in Chapter 3.

As noted by Smyth et al. (2005), a key benefit of employing unfamiliar faces as stimuli for exploring visual serial memory is that adult face recognition is a skilled process, which means that unfamiliar faces can be encoded quickly and easily. Indeed, Smyth et al. (2005) have shown that serial position effects using sequences of unfamiliar faces can be detected using presentation rates as little as 300ms per-item. In contrast, the visual matrix stimuli employed by Avons do not correspond to a familiar cognitive category and require longer presentation rates to produce stable serial memory performance. Accordingly, one concern is that the novelty and complexity of such stimuli may place undue demands on their encoding, thereby potentially obscuring the estimation of serial memory effects in the visual domain. Importantly, Smyth et al. (2005) have shown that serial memory phenomenon observed with unfamiliar faces are resistant to verbal encoding, confirming that such stimuli can be used to explore visual serial memory devoid of contamination by verbal mediation.

In what follows, I present the results of six experiments, which examined the LDFs underpinning serial reconstruction of sequences of unfamiliar faces across a range of experimental manipulations, including sequence length (Experiments 1 & 2), articulatory suppression (Experiment 2 & 5), the dynamics of the reconstruction array (Experiment 3), temporal grouping (Experiment 4), visual similarity (Experiment 5), as well as a direct comparison of serial reconstruction for verbal and visual stimuli (Experiment 6). To foreshadow the main results, the experiments consistently revealed LDFs characterised by negative anticipation slopes and shallow positive postponement slopes. This outcome is in accordance with the empirical pattern reported by Farrell and Lewandowsky (2004) for the serial recall of verbal items and is consistent with the error latency prediction of a model in which serial order is represented via a mechanism combining a primacy gradient, positional marking, and response suppression. Quantitative fits of a sub-set of the models from Chapter 3 to representative data taken from Experiment 2 confirm that when model parameters are estimated from behavioural data the abovementioned model is the only one to predict the observed pattern of the empirical LDFs.

Experiment 1

The task employed to probe visual serial memory in the experiments of this chapter involved serial reconstruction of sequences of unfamiliar faces. Participants were presented with pictures of unfamiliar faces, one item at a time, in the centre of a computer screen. Once the sequence had been presented, the faces reappeared in random positions within a circular array and participants were asked to reconstruct the order of the sequence by clicking on the faces using a mouse-driven pointer.

The first experiment examined the LDFs underpinning this task for sequences containing four, five, and six unfamiliar faces. The rationale for the sequence length manipulation was manifold. First, it permitted an analysis of potential performance related variability in the LDFs, since previous research has shown that serial memory for visual stimuli (Avons, 1998; Smyth et al., 2005; Ward et al., 2005) deteriorates as the length of the to-be-remembered sequence increases, in the same way as it does for verbal stimuli (Anderson et al., 1998; Crannell & Parrish, 1957; Maybery et al., 2002). Second, it enabled an examination of those functions under conditions that were expected to engender changes in response times. For example, in studies of typing Sternberg and colleagues (Sternberg et al., 1978) have shown that the mean latency to perform the first item in a verbal sequence recalled from short-term memory increases as an approximately linear function of sequence length. Additionally, the average response time for the subsequent inter-response intervals, known as the rate, also increases approximately linearly with sequence length. Rhodes, Bullock, Verwey, Averbeck, and Page (2004) have dubbed these chronometric effects as the sequence length effect on latency and the sequence length effect on rate, respectively. That these chronometric effects are a characteristic of verbal serial memory was confirmed by Maybery et al. (2002). Since visual serial memory is susceptible to sequence length effects on accuracy like verbal serial memory, it is anticipated that it too will exhibit the sequence length effect on latency and the sequence length effect on rate. Third, the sequence length manipulation enabled an examination of the sensitivity of the LDFs to changes in the range of possible transposition displacements, which naturally increases with the number of sequence-items.

Method

Participants

Eighteen participants recruited from the student population at the University of York took part in the experiment in exchange for course credit or an honorarium of £10.

Apparatus & Stimuli

The stimuli were sequences of four, five, and six unfamiliar faces of the same gender drawn randomly without replacement from an ensemble of 814 front profile images of unfamiliar faces subject to the constraint that no face was presented on more than two occasions. They were presented in greyscale on a white background at a height of 1.5 inches. Stimulus presentation and data collection were controlled using software developed in-house running on a Dell Optiplex (Intel Core 2 Duo, 2.13 GHz processor) PC equipped with a 19" monitor and a Razer Copperhead high precision mouse. The same apparatus were used for all subsequent experiments.

Design & Procedure

The experiment manipulated a single independent variable: Sequence Length (four-, five-, and six-items), which was a within-subjects factor. The six permutations of the three conditions were fully counterbalanced across participants.

Participants were tested individually in a quiet room in the presence of the experimenter. They initiated each trial by selecting a 'begin trial' icon in the centre of the computer display using the mouse-driven pointer. A central fixation cross then appeared for 1500ms and was replaced by a sequence of faces presented singly for 500ms each and separated by a 500ms blank interval. The final face was followed by a 1000ms delay, after which the faces reappeared simultaneously on-screen in a noisy circular array. Beneath the array were a series of initially empty response windows, the number of which was the same as items in the sequence. Participants were required to reconstruct the sequence in forward serial order by selecting the faces using the mouse-driven pointer. Once an item was selected it was deleted from the reconstruction array, before reappearing in the response window representing the current output position. Participants could choose to omit items by selecting an icon containing a question mark located in the centre of the reconstruction array. This resulted in a question mark being assigned to the appropriate response window. Once participants had reconstructed the sequence they selected a 'validate input' icon, following which the contents of the screen cleared and the reconstruction time for the sequence was presented in the central screen position for 3000ms, after which the 'begin trial' icon for the next trial appeared.

Participants took part in two testing sessions (spaced at least 24-hours apart) each of which contained three blocks of 50 trials of each sequence length. There were two practice trials at the start of each block and enforced 1-minute rest periods were imposed after every twenty five experimental trials. Each testing session lasted approximately 70 minutes.

Results

The data were analysed using a strict serial reconstruction scoring procedure: an item was scored correct only if it was reported in the same serial position that it was presented. The results are organised in to four sections: (1) accuracy serial position curves, (2) transposition gradients, (3) latency serial position curves, and (4) LDFs.

Accuracy serial position curves

The accuracy serial position curves are depicted in Figure 4-1A. It is evident from inspection of this figure that the serial position curves exhibit effects of both primacy and recency. Also apparent is that serial reconstruction accuracy deteriorates as a function of increasing sequence length. Statistical confirmation of the former claim was obtained by entering the data for each sequence

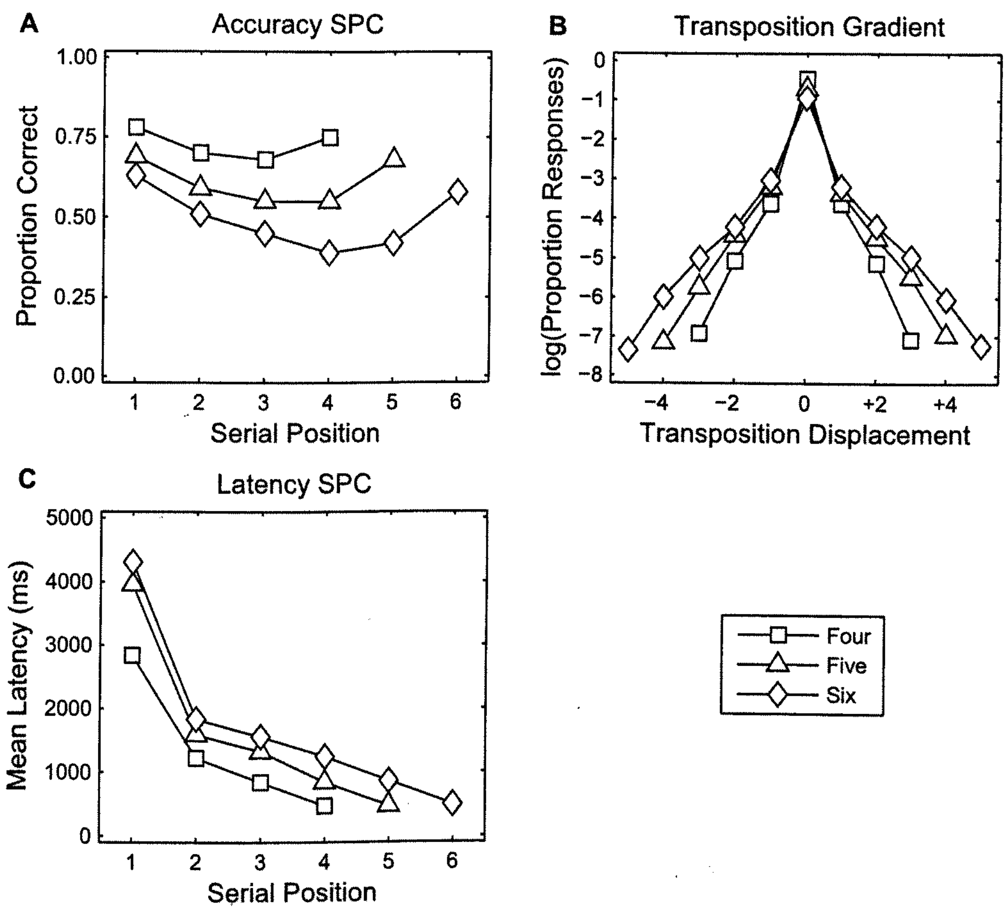

Figure 4-1 *Serial memory performance measures for Experiment 1. Panels show accuracy serial position curves (A), transposition gradients, (B) and latency serial position curves (C).*

length into one-way serial position ANOVAs[1]. Reliable effects of serial position were obtained for four-item, $F(3, 51) = 17.366$, $MSE = .045$, $p < .001$, five-item, $F(4, 68) = 16.969$, $MSE = .005$, $p < .001$, and for six-item sequences, $F(5, 85) = 36.409$, $MSE = .231$, $p < .001$, and subsequent trend analyses revealed significant quadratic trends for each sequence length: $F(1, 17) = 53.730$, $MSE = .123$, $p < .001$, for four-item, $F(1, 17) = 45.833$, $MSE = .302$, $p < .001$, for five-item, and, $F(1, 17) = 104.957$, $MSE = .652$, $p < .001$, for six-item sequences. The effect of sequence length was statistically confirmed by a one-way ANOVA considering the mean proportion of correct responses for each length. This revealed a reliable effect of sequence length, $F(2, 34) = 95.615$, $MSE = .235$, $p < .001$, with four-item sequences being recalled better than five-item sequences, $t(17) = 6.260$, p

[1] For all within-subjects ANOVAs reported in this thesis violations of the assumption of sphericity were accommodated by use of the Greenhouse-Geisser correction.

< .001, and five-item sequences in turn being recalled better than six-item sequences, $t(17) = 9.125$, $p < .001$.

Transposition gradients

The transposition gradients are portrayed in Figure 4-1B and exhibit three major characteristics. First, the gradients peak at displacement value zero, reflecting the majority of responses are correct responses. Second, the proportion of transpositions is greatest for displacements with an absolute value of one, and decreases gradually as the absolute value of the displacement increases. Third, the transposition gradients are symmetrical: that is the proportion of transpositions for each absolute displacement value is approximately the same for anticipations and postponements. Also apparent is that the proportion of anticipations and postponements increases as a function of sequence length, consistent with the accuracy serial position analysis.

Latency serial position curves

Figure 4-1C shows the mean response latencies associated with correct responses as a function of serial position. The functions are characterised by a long initial preparatory latency and an overall monotonically negative relationship between latency and serial position. It is also apparent that latencies become longer at each serial position as the sequence length increases.

The latency serial position curves were analysed by entering the data for each sequence length into separate one-way ANOVAs. There were significant effects of serial position for all sequence lengths: $F(3, 51) = 125.334$, $MSE = 4.655E7$, $p < .001$, for four-item, $F(4, 68) = 104.972$, $MSE = 1.144E8$, $p < .001$, for five-item, and, $F(5, 85) = 140.914$, $MSE = 9.926E7$, $p < .001$, for six-item sequences. The negative relationship between latency and serial position was confirmed by trend analyses, which revealed significant linear components for four-item, $F(1, 17) = 180.524$, $MSE = 5.007E7$, $p < .001$, five-item, $F(1, 17) = 176.473$, $MSE = 1.073E8$, $p < .001$, and for six-item sequences, $F(1, 17) = 221.175$, $MSE = 1.279E8$, $p < .001$.

The presence of a sequence length effect on latency was evaluated via a one-way ANOVA on the mean latencies for the first serial position for each sequence length. This revealed a significant effect of sequence length, $F(2, 34) = 44.186$, $MSE = 1.067E7$, $p < .001$, which was determined to contain a significant linear component by a subsequent trend analysis, $F(1, 17) = 99.785$, $MSE = 1.95E7$, $p < .001$, indicating that latencies for the first position increased as an approximately linear function of sequence length. The presence of a sequence length effect on rate was evaluated in a similar fashion, but using the mean of the inter-response latencies (excluding the first response), as the dependent measure. This revealed a significant effect of sequence length, $F(2, 34) = 28.671$, $MSE = 791054.858$, $p < .001$, which was again characterized by a significant linear trend, $F(1, 17) = 35.852$, $MSE = 1095219.459$, $p < .001$, indicating that the mean inter-response rate also increased as an approximately linear function of sequence length.

Latency-displacement functions

Turning now to the data of chief interest, Figure 4-2 shows the LDFs, which plot the average latencies of transpositions as a function of transposition displacement[2]. The effects of output position have been removed from these data in a manner akin to that described for the model predictions in Chapter 3, by subtracting each participant's mean latency for each output position from their individual latencies at corresponding output positions. Note that the negative latencies associated with some transposition displacements are a consequence of this filtering process. As can be seen from inspection of this figure, the LDFs are characterised by steep negative slopes for anticipations and shallow positive slopes for postponements.

The LDFs were analysed using the following two-stage procedure. In the first stage, regression analyses were performed examining the relationship between transposition latency and

[2] Exploratory analyses of the latency-displacement functions for all experiments in this chapter were performed initially to examine their sensitivity to potential outliers. These analyses revealed that the qualitative form of the latency-displacement functions (the sign and steepness of the anticipation and postponement slopes) was generally unaffected by whether all latencies were included, or only responses within the range of 2.5, or 3 standard deviations from the mean. The empirical pattern was also similar when the mean of the median latencies was used as the dependent measure instead of the mean latency. Given the insensitivity of the qualitative form of the latency-displacement function to these different approaches to dealing with response time outliers all responses were retained for the transposition latency analysis and the mean latency was employed as the dependent measure.

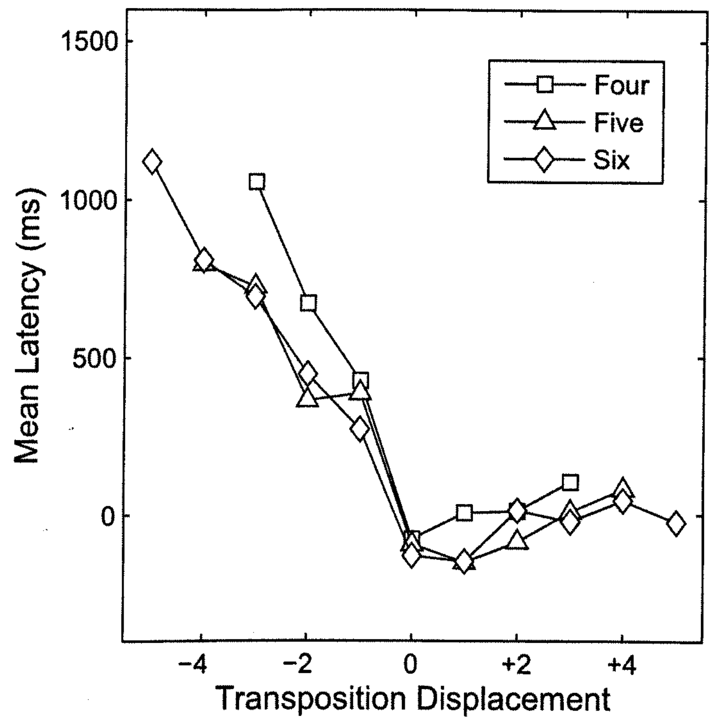

Figure 4-2 *Latency-displacement functions for Experiment 1.*

transposition displacement for each individual participant. One analysis examined the relationship between transposition latency and transposition displacements that were anticipations (negative displacement values), whilst a second examined the relationship between transposition latency and transposition displacements that were postponements (positive displacement values). Thus, regression equations were computed for each participant by regressing transposition latency on transposition displacements that were anticipations and postponements separately. Each regression equation represents the best description (in a least squares sense) of the relationship between transposition latency and the predictor variable (anticipations or postponements) for a specific participant. The transposition displacement values included in the regression analyses for anticipations ranged from –3 to 0 for four-item sequences, -4 to 0 for five-item sequences, and –5 to 0 for six-item sequences. The transposition displacement values included in the regression analyses for postponements ranged from 0 to +3 for three item sequences, 0 to +4 for five-item sequences, and 0 to +5 for six-item sequences.

In the second stage of the analysis, the regression slope parameter estimates for anticipations and postponements were subjected to one-sample t-tests to determine whether they deviated

Parameter	Estimate	*SE*	*t*	*p*
Four				
Anticipation	-386.61	87.44	-4.422	.00
Postponement	58.66	18.44	3.180	.01
Five				
Anticipation	-223.09	107.83	-2.069	.05
Postponement	50.90	14.38	3.539	.00
Six				
Anticipation	-249.40	42.14	-5.918	.00
Postponement	38.95	14.71	2.647	.02

Table 4-1 *Mean regression slope parameter estimates for the latency-displacement functions of Experiment 1.*

reliably from zero. The mean regression slope parameter estimates are shown in Table 4-1. As can be seen from inspection of this table, the mean slope estimates for anticipations were steeply negative and deviated significantly from zero: $t(17) = -4.422$, $p < .001$, for four-item sequences, $t(17) = -2.069$, $p = .05$, for five-item sequences, and, $t(17) = -5.918$, $p < .001$, for six-item sequences. The mean slope estimates for postponements were shallowly positive in comparison, but also deviated significantly from zero: $t(17) = 3.180$, $p < .01$, for four-item sequences, $t(17) = 3.539$, $p < .01$, for five-item sequences, and, $t(17) = 2.647$, $p < .05$, for six-item sequences.

To give some indication of the variability in the LDFs, the anticipation and postponement slope estimates for individual participants are portrayed graphically in Figure 4-3. As can be seen from inspection of this figure, the majority of participants contributed steep negative slopes for anticipations, whereas the majority of participants contributed shallow positive slopes for postponements. Inspection of the individual regression slope estimates therefore confirms that the empirical pattern of the averaged LDFs portrayed in Figure 4-2 is robust and not the consequence of a small number of participants exerting undue influence on the data.

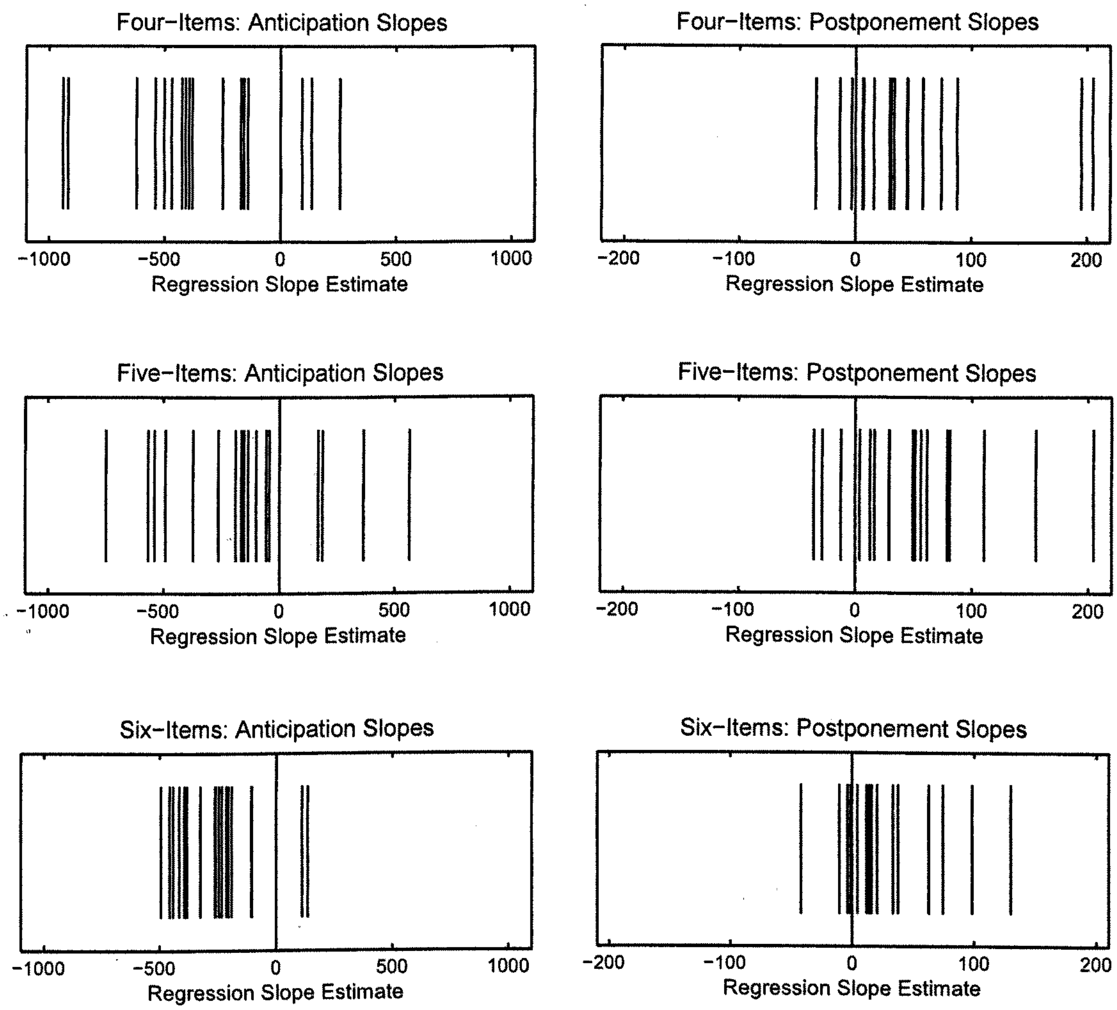

Figure 4-3 *Individual regression slope parameter estimates for the latency-displacement functions of Experiment 1. Each vertical line in a panel represents a regression slope estimate for a participant for a particular experimental condition. The left hand panels show slope estimates for anticipations, whilst the right hand panels show slope estimates for postponements. The top panels show the slope estimates for four-item sequences, the middle panels show the slope estimates for five-item sequences and the bottom panels show the slope estimates for six-item sequences. Note— the scales of the x-axes for the slope estimates for anticipations and postponements are different.*

Discussion

Before interpreting the results of the current experiment, a brief recap of the predictions of the models presented in Chapter 3 is in order. Recall from that chapter that although the five models and associated mechanisms for representing serial order predicted qualitatively similar serial position curves and transposition gradients (see Figure 3-1) they also predicted qualitatively different LDFs (see Figure 3-2B, which shows the LDFs predicted by the models with the effects of output position subtracted, as in the data shown in Figure 4-2). While all the models predicted that the slope of the LDF for anticipations is steeply negative there was considerable heterogeneity in their predictions for the slope of the function for postponements. Specifically, the models involving positional marking, either alone, or in conjunction with response suppression or output interference, predicted steep positive postponement slopes; the model combining a primacy gradient with response suppression predicted a negative postponement slope; whilst the model combining a primacy gradient, position marking, and response suppression predicted a shallow positive postponement slope.

The results of the current experiment are unambiguous. The LDFs for serial reconstruction of visual sequences varying in length were consistently characterised by steep negative slopes for anticipations and shallow positive slopes for postponements. This empirical outcome is in accordance with the pattern observed in the verbal serial recall studies of Farrell and Lewandowsky (2004) and is most compatible with the error latency prediction of a representational mechanism combining a primacy gradient of activation, positional marking, and response suppression (i.e., the PG+PM+RS model). The same results are incompatible with the error latency predictions of the four alternative mechanisms for representing serial order. This conclusion is subject to one important qualification: because items were automatically deleted from the reconstruction array after being selected the data do not identify a role for response suppression. That is, the current data only mandate a role for a primacy gradient and positional marking.

Attention is now given to the empirical patterns obtained for the three additional serial memory performance measures. Considering first the accuracy serial position curves, these exhibited effects of both primacy and recency as predicted by all the models (see Figure 3-1A), with a more marked primacy effect, similar to that which is typically seen in verbal serial recall and serial reconstruction. Serial memory performance was also shown to deteriorate as a function of increasing sequence length, consistent with previous studies employing visual (Avons, 1998; Smyth et al., 2005; Ward et al., 2005), verbal (Anderson et al., 1998; Anderson & Matessa, 1997; Crannell & Parrish, 1957), and spatial (Jones et al., 1995; Smyth & Scholey, 1996) stimuli. These findings add to a wealth of studies indicating that verbal, visual, and spatial short-term memory exhibit functionally similar serial position curves when judgements of serial order are required.

Turning to the latency serial position curves, these were also similar to those obtained in verbal studies, in two respects. First, the latency for selecting the first item in the sequence was considerably longer than that for any other item (e.g., Anderson et al., 1998; Anderson & Matessa, 1997; Farrell & Lewandowsky, 2004; Maybery et al., 2002; Thomas et al., 2003). Second, there were sequence length effects on the first response latency and the inter-response rate, with latencies for both measures increasing approximately linearly as a function of sequence length (e.g., Maybery et al., 2002). It should be borne in mind, however, that these sequence length effects on latency may in part reflect differences in visual search times to find items in the noisy reconstruction array. Notwithstanding these similarities, the functions diverge from those reported in verbal studies in one important respect: the relationship between latency and serial position was monotonically negative in the current study, whereas in verbal studies latency typically varies inversely with serial position accuracy (see e.g., Farrell, 2008; Farrell & Lewandowsky, 2004; Maybery et al., 2002; Parmentier & Maybery, 2008), as is predicted by the models (see Figure 3-1C). Potential reasons for this discrepancy are considered in the general discussion. However, for present purposes the reader is reminded that the effects of output position on latency were removed from the LDFs, both for models and data. The departure of the empirical latency serial position curves from the predictions of the models does not therefore compromise the main theoretical conclusions arising from the current experiment.

Finally, the transposition gradients exhibited the three major characteristics of the transposition gradients observed in verbal serial recall (e.g., Farrell & Lewandowsky, 2004) and predicted by the models (see Figure 3-1B). First, the transposition gradients peaked at displacement value zero, reflecting that most responses were correct responses. Second, the probability of a transposition decreased as the absolute transposition displacement value increased, consistent with the locality constraint identified in verbal serial recall (Henson, 1996). Third, the transposition gradients were approximately symmetrical: that is the error gradients for anticipations and postponements were mirror images of each another.

In summary, the main finding of the current experiment was the observation of LDFs characterised by steep negative anticipation slopes and shallow positive postponement slopes, consistent with the error latency prediction of a representational mechanism combining a primacy gradient with positional marking. Subsequent experiments in this chapter examine the generality of this empirical pattern, and by extension the representational mechanism it supports.

Experiment 2

One conjecture that might be levelled at the findings of Experiment 1 is that the similarities between the observed LDFs and those obtained in the verbal serial recall studies of Farrell and Lewandowsky (2004) might be attributable to participants developing verbal descriptors of the face stimuli and rehearsing these as sequences of verbal tokens. If this were so, then the assertion would be that the task is verbal rather than nonverbal in nature, and it is for this reason that the data resemble those for verbal serial recall. Given that Smyth et al. (2005) have shown using a very similar experimental design and procedure that serial memory effects obtained with sequences of unfamiliar faces are not due to verbal encoding this account seems unlikely. Nevertheless, the aim of Experiment 2 was to examine whether the relationship between transposition displacement and latency observed across the different sequence length conditions of Experiment 1 generalized to conditions in which the opportunity to verbally encode faces was eliminated by having participants engage in articulatory suppression during encoding. Because of the unrealistically large number of trials that would have been required to have individuals complete both a quiet condition and an articulatory suppression condition, only the latter condition was incorporated and an estimate of the impact of articulatory suppression was obtained by means of a between-experiment comparison with Experiment 1.

Method

Participants

Eighteen participants recruited from the student population at the University of York took part in the experiment in exchange for course credit or an honorarium of £10.

Design & Procedure

The design and procedure were the same as for Experiment 1, with one exception. At the beginning of each trial, once the participant had selected the 'begin trial' icon they were required to repeat the sequence of digits "1", "2", "3", "4" out loud at the rate of three digits per second until the reconstruction array appeared. The articulation rate was demonstrated to the participant prior to the first practice trial using the built-in digital metronome on an Apple MacBook laptop. The experimenter remained present during testing to ensure compliance with the suppression protocols. If the participant failed to keep to the rate of three utterances per second the rate was demonstrated again using the digital metronome and the individual was warned by the experimenter to try harder.

Results

Accuracy serial position curves

The accuracy serial position curves are shown in Figure 4-4A. They exhibit effects of primacy and recency, as well as a decline in serial reconstruction accuracy with increasing sequence length. Statistical confirmation of the effects of serial position was obtained by entering the data for each sequence length into separate one-way ANOVAs. Reliable effects of serial position were obtained for all sequence lengths: $F(3, 51) = 24.183$, $MSE = .110$, $p < .001$, for four-item, $F(4, 68) = 26.746$, $MSE = .281$, $p < .001$, for five-item, and, $F(5, 85) = 57.095$, $MSE = .244$, $p < .001$, for six-item sequences. Subsequent trend analyses identified significant quadratic components for each sequence length: $F(1, 17) = 61.130$, $MSE = .192$, $p < .001$, for four-item, $F(1, 17) = 123.206$, $MSE = .497$, $p < .001$, for five-item, and, $F(1, 17) = 223.523$, $MSE = 1.179$, $p < .001$, for six-item sequences. The impact of the sequence length manipulation was verified by a one-way ANOVA on the mean proportion of correct responses for each sequence length. This effect was significant, $F(2, 34) = 79.459$, $MSE = .166$, $p < .001$, with four-item sequences being recalled more accurately than five-item sequences, $t(17) = 7.270$, $p < .001$, and five-item sequences in turn being recalled more accurately than six-item sequences, $t(17) = 6.463$, $p < .001$.

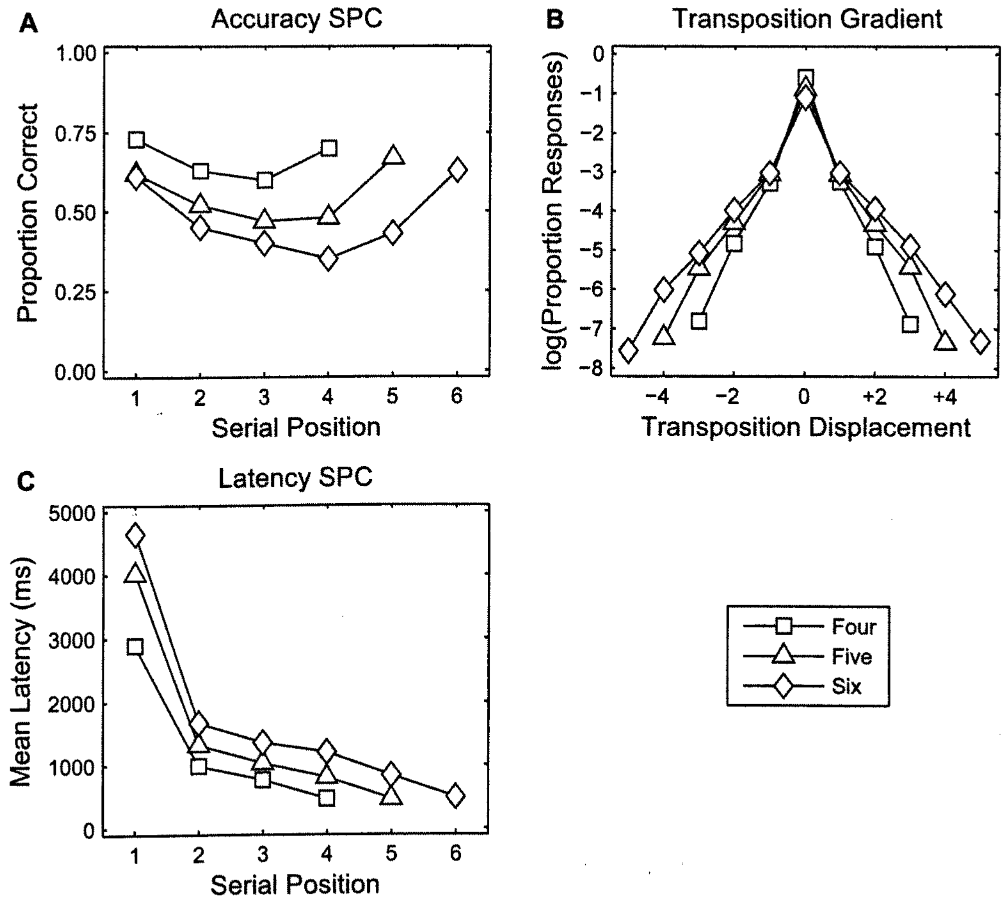

Figure 4-4 *Serial memory performance measures for Experiment 2. Panels show accuracy serial position curves (A), transposition gradients, (B) and latency serial position curves (C).*

To ascertain the impact of the articulatory suppression manipulation, a 2 (Experiment) X 3 (Sequence length) ANOVA was conducted on the mean proportion of correct responses for each sequence length for Experiments 1 and 2. This revealed a reliable main effect of sequence length, $F(2, 68) = 173.698$, $MSE = .464$, $p < .001$, but neither the main effect of Experiment, $F(1, 34) = 1.680$, $MSE = .060$, $p = .204$, nor the Experiment X Sequence length interaction, $F(2, 68) = 1.690$, $MSE = .005$, $p = .197$, reached significance.

Transposition gradients

The transposition gradients depicted in Figure 4-4B exhibit steep peaks at displacement value zero, a locality constraint on the distribution of transpositions, and symmetrical error gradients for anticipation and postponement errors. Consistent with the serial position analysis, the proportions of anticipations and postponements increases as a function of sequence length.

Latency serial position curves

Figure 4-4C shows the response latency serial position curves for correct responses. It is apparent that the latency for the first item is considerably longer than that for any other item in the sequence, and that latency is an overall monotonically negative function of serial position. Also apparent is that latencies become longer at each serial position as sequence length increases. The latency serial position curves were evaluated by conducting one-way serial position ANOVAs on the data for each sequence length. These revealed reliable effects of serial position in all instances: $F(3, 51) = 92.278$, $MSE = 6.026E7$, $p < .001$, for four-item, $F(4, 68) = 128.334$, $MSE = 1.238E8$, $p < .001$, for five-item, and, $F(5, 85) = 91.137$, $MSE = 1.343E8$, $p < .001$, for six-item sequences. Subsequent trend analyses confirmed that the serial position curves possessed a significant linear component: $F(1, 17) = 120.871$, $MSE = 5.053E7$, $p < .001$, for four-item, $F(1, 17) = 160.616$, $MSE = 1.037E8$, $p < .001$, for five-item, and, $F(1, 17) = 124.866$, $MSE = 1.406E8$, $p < .001$, for six-item sequences.

The presence of a sequence length effect on latency was examined via a one-way ANOVA on the mean response latencies for the first serial position as a function of sequence length. The effect was significant, $F(2, 34) = 55.188$, $MSE = 2.019E7$, $p < .001$, and was characterised by a reliable linear trend, $F(1, 17) = 66.826$, $MSE = 2.764E7$, $p < .001$, indicating that latencies for the first serial position increased as an approximately linear function of sequence length. The presence of a sequence length effect on rate was similarly evaluated by a one-way ANOVA performed this time on the mean of the inter-response latencies (excluding the first response latency) as a function of sequence length. The effect was again significant, $F(2, 34) = 32.291$, $MSE = 1181860.880$, $p < .001$, and characterised by a reliable linear trend, $F(1, 17) = 40.118$, $MSE = 1606228.477$, $p < .001$, indicating that the inter-response rate also increased as an approximately linear function of sequence length.

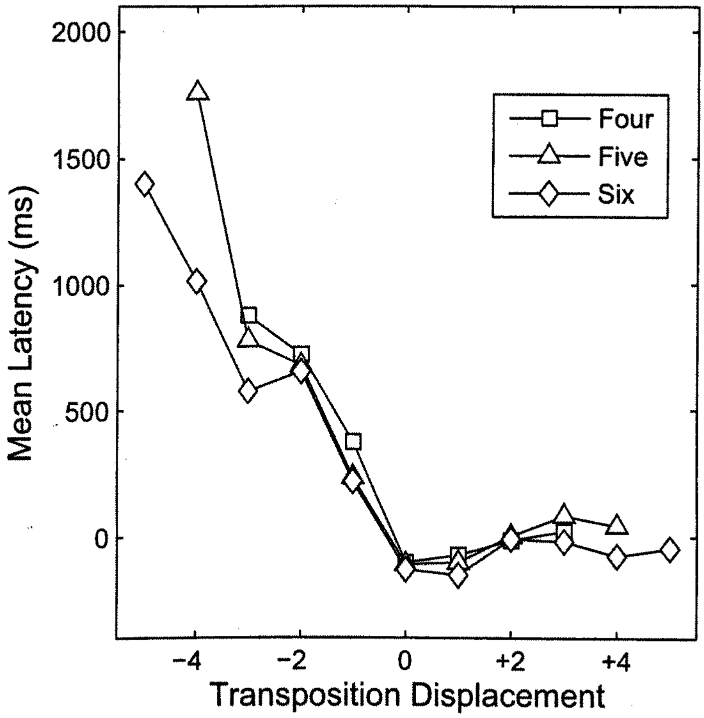

Figure 4-5 *Latency-displacement functions for Experiment 2.*

Latency-displacement functions

The LDFs with the effects of output position subtracted can be inspected in Figure 4-5. Consistent with the previous experiment, the slopes of the functions for anticipations are steeply negative, whereas the slopes of the functions for postponements are shallowly positive. The data were analysed as per Experiment 1 by conducting regression analyses on the LDFs for each participant examining the relationship between transposition latency and transposition displacements that were anticipations and postponements separately. The regression slope parameter estimates were subsequently evaluated using one-sample t-tests to determine whether they deviated reliably from zero.

The mean regression slope parameter estimates for anticipations and postponements for the different sequence length conditions are summarised in Table 4-2. The mean regression slope parameter estimates for anticipations were steeply negative and deviated reliably from zero: $t(17) = -3.874$, $p = .001$, for four-item, $t(17) = -2.081$, $p = .05$, for five-item, and, $t(17) = -2.407$, $p < .01$, for six-item sequences. The mean regression slope parameter estimates for postponements were

Parameter	Estimate	*SE*	*t*	*p*
Four				
Anticipation	-328.23	84.73	-3.874	.00
Postponement	43.37	10.72	4.046	.00
Five				
Anticipation	-441.46	212.13	-2.081	.05
Postponement	51.29	16.14	3.177	.00
Six				
Anticipation	-334.81	139.11	-2.407	.03
Postponement	23.75	8.11	2.928	.01

Table 4-2 *Mean regression slope parameter estimates for the latency-displacement functions of Experiment 2.*

shallowly positive by comparison, but also deviated reliably from zero: $t(17) = 4.046, p = .001$, for four-item, $t(17) = 3.177, p < .01$, for five-item, and, $t(17) = 2.928, p < .01$, for six-item sequences.

To illustrate the variability in the LDFs, the anticipation and postponement slope estimates of individual participants are shown graphically in Figure 4-6. As for Experiment 1, the majority of participants contributed steep negative slope estimates for anticipations and shallow positive slope estimates for postponements.

Discussion

The results of the current experiment parallel those documented for Experiment 1. Under conditions of articulatory suppression the accuracy and response latency serial position curves and transposition gradients for sequences varying in length exhibited the same qualitative characteristics as witnessed previously. Indeed, a between experiment comparison of performance failed to identify a significant detrimental effect of articulatory suppression, suggesting that if verbal encoding strategies were employed by participants, they contributed little to performance. However, the main outcome of the present experiment was to show that when the opportunity to engage in verbal encoding was obstructed the LDFs still exhibited the same empirical pattern as before – the slopes of the functions for anticipations were steeply negative, whereas the slopes of

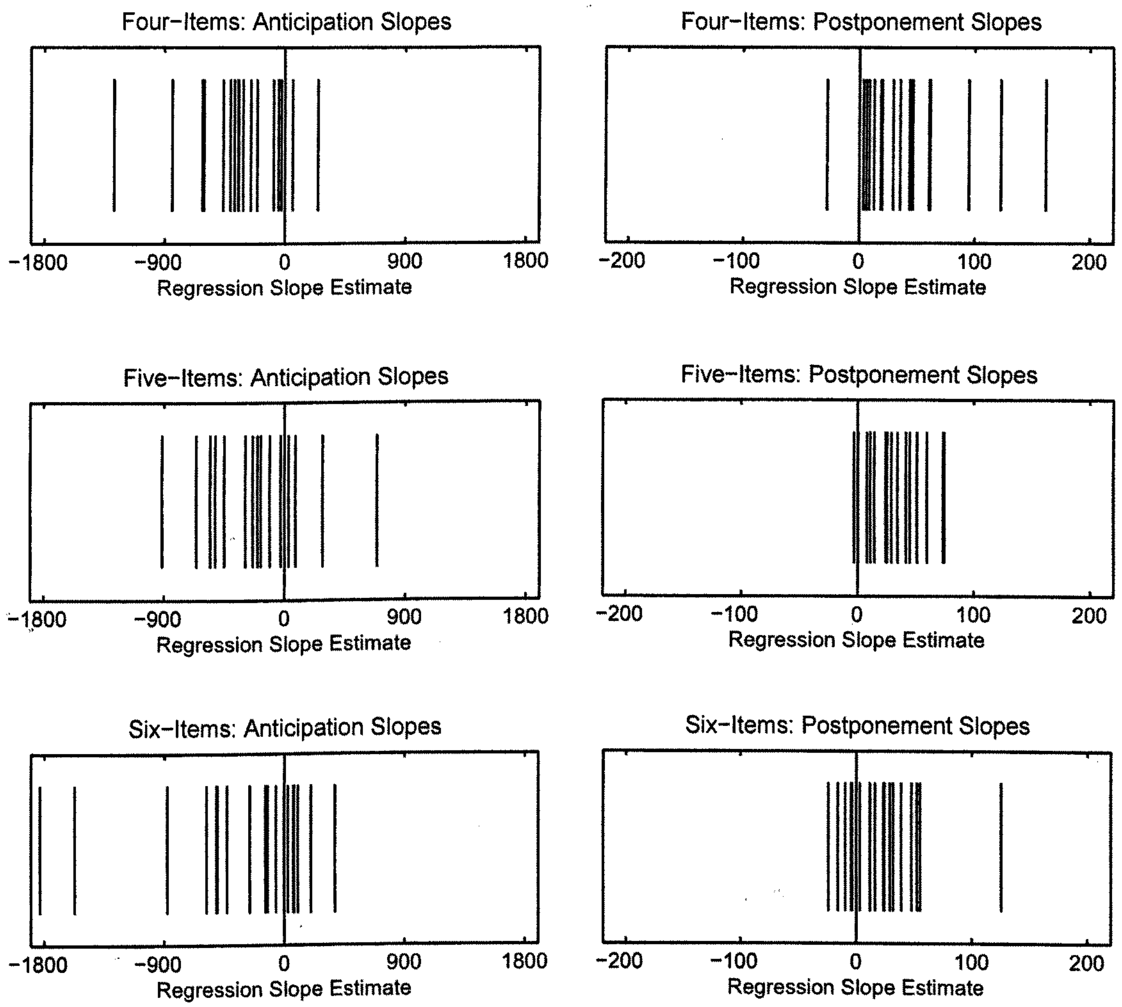

Figure 4-6 *Individual regression slope parameter estimates for the latency-displacement functions of Experiment 2. Each vertical line in a panel represents a regression slope estimate for a participant for a particular experimental condition. The left hand panels show slope estimates for anticipations, whilst the right hand panels show slope estimates for postponements. The top panels show the slope estimates for four-item sequences, the middle panels show the slope estimates for five-item sequences and the bottom panels show the slope estimates for six-item sequences. Note—the scales of the x-axes for the slope estimates for anticipations and postponements are different.*

the functions for postponements were shallowly positive. This outcome nullifies concerns that the LDFs observed in Experiment 1 materialised due to contamination by verbal mediation, and instead lends support to the hypothesis that the serial order of a sequence of visual items is represented by a mechanism combining a primacy gradient with positional marking. The same findings question once more the viability of the four alternative mechanisms for representing serial order.

Experiment 3

One limitation of the methodology of Experiments 1 and 2 is that items were deleted from the reconstruction array once they had been selected, meaning that individuals did not have to keep track of items they had already selected, in order to prevent erroneous repetitions in their responding. This constitutes a notable shortcoming, since three of the four models presented in Chapter 3 invoke a specific mechanism – namely response suppression – to prevent the selection of already emitted items, and to enable recall to evolve. The aim of Experiment 3 was to examine whether the relationship between transposition latency and transposition displacement observed in the previous experiments would still hold under a reconstruction scenario in which this external guidance was removed. That is, an experimental scenario in which the operation of response suppression should be necessary.

Accordingly, two reconstruction conditions were compared: a with-clearing condition and a without-clearing condition. The former reconstruction condition is the same as that employed in the previous experiments in that once an item is selected it is deleted from the reconstruction array. In contrast, in the latter condition items remain on-screen once selected and can therefore be erroneously chosen again. Thus, the without-clearing condition possesses recall dynamics that are functionally more akin to those of the models incorporating response suppression. If response suppression contributes to the serial reconstruction of visual sequences then performance in the without-clearing condition should resemble that in the with-clearing condition. That is, because response suppression fulfils the same goal as removing items from the reconstruction array – preventing erroneous repetitions in responding – serial reconstruction performance should not dramatically differ between the two conditions.

More specifically, in accordance with the repetition constraint identified in verbal serial recall (Henson, 1996), which is considered to be an empirical signature of the operation of response suppression, erroneous repetitions should be scarce in the without-clearing condition. Nevertheless, at least some repetitions are anticipated, since it is expected that this response suppression is imperfect, as is assumed to be the case in verbal serial recall (e.g., Henson, 1998b; Vousden &

Brown, 1998). That said the incidence of such erroneous repetitions in the without-clearing condition should not be large enough to produce a significant difference in accuracy between the two reconstruction conditions.

Method

Participants

Twenty two participants recruited from the student population at the University of York took part in the experiment in exchange for course credit or an honorarium of £4.

Stimuli

These were sequences of six unfamiliar faces of the same gender drawn randomly without replacement from a stimulus ensemble of 1000 faces subject to the constraint that no face was presented more than once throughput the entire experiment.

Design & Procedure

The experiment manipulated a single independent variable: Condition (with-clearing / without-clearing), which was a within-subjects factor. Half of the participants completed the with-clearing condition followed by the without-clearing condition, with the remaining half of participants completing the conditions in the converse order.

The procedure was identical to that for Experiment 1 with the following exceptions: First, in the without-clearing condition items remained in the reconstruction array once they had been selected and repetition errors were therefore possible. Second, in both reconstruction conditions once an item was selected, instead of presenting that item within the response box corresponding to the item's output position (as for Experiment 1 & 2), instead the appropriate response box illuminated green to acknowledge that a response for that position had been made.

Participants attempted 80 trials in each condition, which were preceded by 2 practice trials. Enforced 30-second rest periods were imposed after every 20 experimental trials. The experiment lasted approximately 60 minutes.

Results

Accuracy serial position curves and transposition gradients

The accuracy serial position curves are shown in Figure 4-7A from which it is apparent that performance does not differ between the with-clearing and without-clearing conditions at any serial position. These data were subjected to a 2 (Condition) X 6 (Serial Position) ANOVA, which revealed a significant main effect of Serial Position, $F(5, 105) = 38.114$, $MSE = .967$, $p < .001$, however, neither the main effect of Condition, $F(1, 21) = .834$, $MSE = .010$, $p = .372$, nor the Condition X Serial Position interaction, $F(5, 105) = .073$, $MSE = .000$, $p = .996$, were significant. Figure 3-11B shows the underlying transposition gradients for the two conditions. Corroborating the serial position analysis, the incidence of anticipations and postponements does not diverge between the two reconstruction conditions.

Repetition errors

Repetition errors in the without-clearing condition were extremely rare, accounting for less than 1% of all responses, which is considerably less than that expected by chance. The average lag between the two instances of the repeat was 3.11 positions. Figure 4-7C shows the proportions of repetitions as a function of output position in the without-clearing condition. Inspection of this figure reveals that the probability of repetitions increased across output positions. This was statistically confirmed by a one-way serial position ANOVA performed on the log-odds transformed error proportions, $F(5, 110) = 5.429$, $MSE = 2.199$, $p < .01$.

Latency serial position curves and latency-displacement functions

The response latency serial position curves for correct responses are shown in Figure 4-7D. Mirroring the accuracy serial position curves the latency serial position curves for the two conditions are virtually superimposed on one another. These data were analysed by means of a 2 (Condition) X 6 (Serial Position) ANOVA, which revealed a reliable main effect of Serial Position, $F(5, 105) = 101.336$, $MSE = 3.123E8$, $p < .001$. However, the main effect of Condition, $F(1, 21) =$

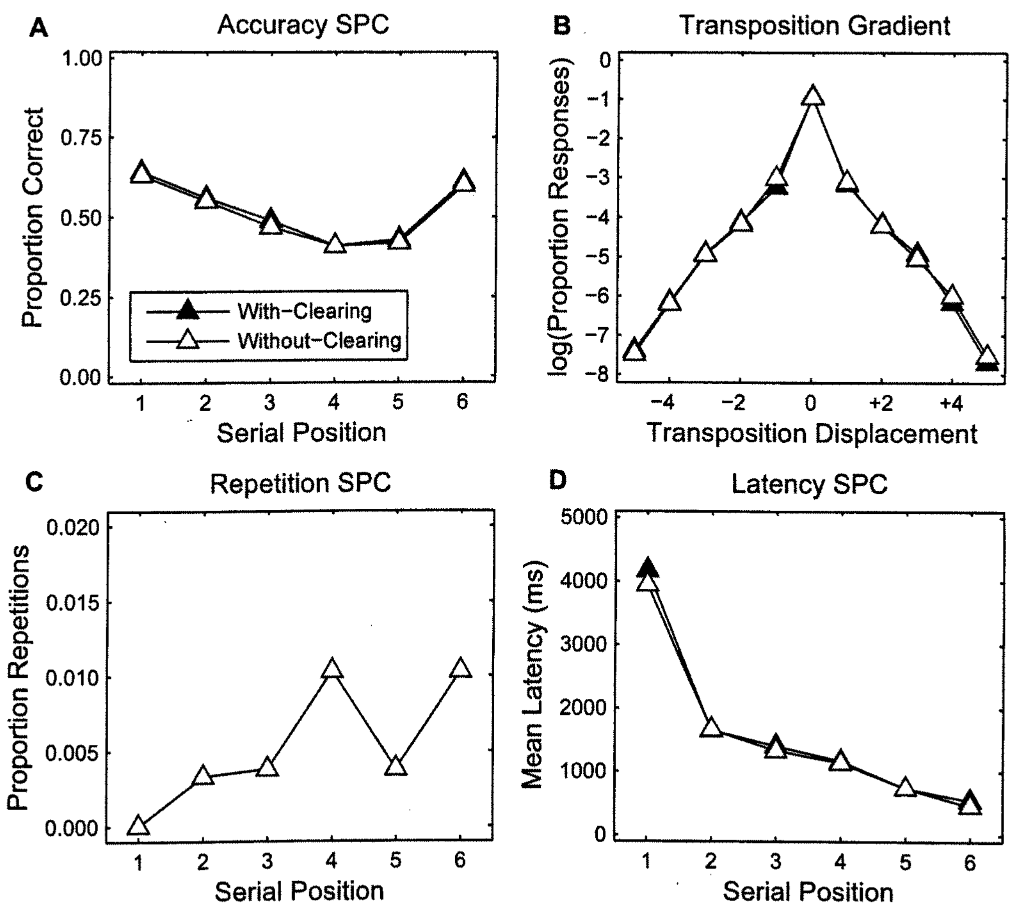

Figure 4-7 *Serial memory performance measures for Experiment 3. Panels show accuracy serial position curves (A), transposition gradients, (B) repetition error serial position curves (C), and latency serial position curves (D).*

1.564, *MSE* = 296633.229, *p* = .225, and the Condition X Serial Position interaction, *F*(5, 105) = 1.208, *MSE* = 147142.118, *p* = .314, both failed to reach significance.

The LDFs with the effects of output position removed are shown in Figure 4-8. The functions for the with- and without-clearing conditions are qualitatively similar, both being characterised by steep negative slopes for anticipations and shallow positive slopes for postponements. The data were analysed by conducting regression analyses on the LDFs for each individual participant examining the relationship between transposition latency and transposition displacements that were anticipations (displacements in the range -5 to 0) and postponements (displacements in the range 0

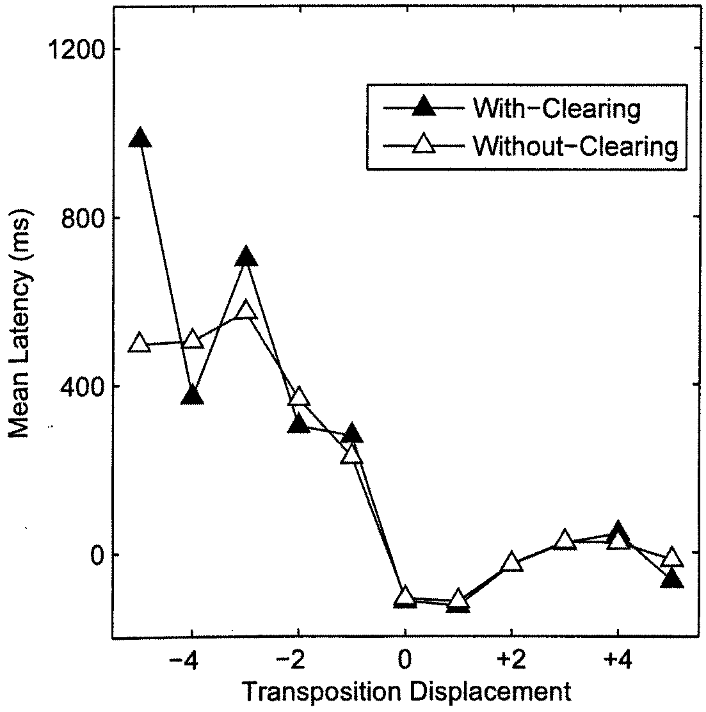

Figure 4-8 *Latency-displacement functions for Experiment 3*

to +5) separately[3]. The regression slope parameter estimates for anticipations and postponements were subsequently pooled together and subjected to one-sample t-tests to establish whether they differed reliably from zero.

The mean regression slope parameter estimates for anticipations and postponements can be scrutinized in Table 4-3. The mean regression slope parameter estimates for anticipations were steeply negative and deviated reliably from zero: $t(21) = -3.276$, $p < .01$, for the with-clearing condition, and, $t(21) = -2.095$, $p = .05$, for the without-clearing condition. In contrast, the mean regression slope parameter estimates for postponements were shallowly positive, but also deviated reliably from zero: $t(21) = 3.322$, $p < .01$, for the with-clearing condition, and, $t(21) = 3.693$, $p = .001$, for the without-clearing condition.

[3] Repetition errors (both occurrences of the repeat), which were only possible in the without-clearing condition, were excluded from the latency-displacement function shown graphically in Figure 4-8 and the associated regression analysis to maintain consistency with the practice employed for generating the model predictions in Chapter 3.

Parameter	Estimate	SE	t	p
With-Clearing				
Anticipation	-176.69	53.94	-3.276	.00
Postponement	22.80	6.86	3.322	.00
Without-clearing				
Anticipation	-104.59	49.92	-2.095	.05
Postponement	24.98	6.77	3.693	.00

Table 4-3 *Mean regression slope parameter estimates for the latency-displacement functions of Experiment 3.*

The individual regression parameter estimates for the slopes of the LDFs for anticipations and postponements can be inspected in Figure 4-9. It is visible from inspection of this figure that most participants contributed negative anticipation slopes, with a small number of participants contributing positive slopes. By contrast most participants contributed positive slope estimates for postponements, although an appreciable number of participants contributed negative slope estimates.

Discussion

The LDFs witnessed in Experiments 1 and 2 are consistent with the predictions of the primacy gradient, positional marking, and response suppression model. However, as already noted, because items were deleted from the reconstruction array in those experiments the data only provide support for the primacy gradient and positional marking components of this model. The aim of the current experiment was to obtain empirical support for the full complement of representational principles underlying this model, by examining the pattern of transposition latencies underlying visual serial reconstruction when this source of external guidance was removed. The without-clearing condition was therefore incorporated to provide an experimental scenario in which the operation of response suppression should be required.

The LDFs obtained for the with-clearing and without-clearing conditions were functionally similar and consistent with the empirical pattern documented in the previous experiments.

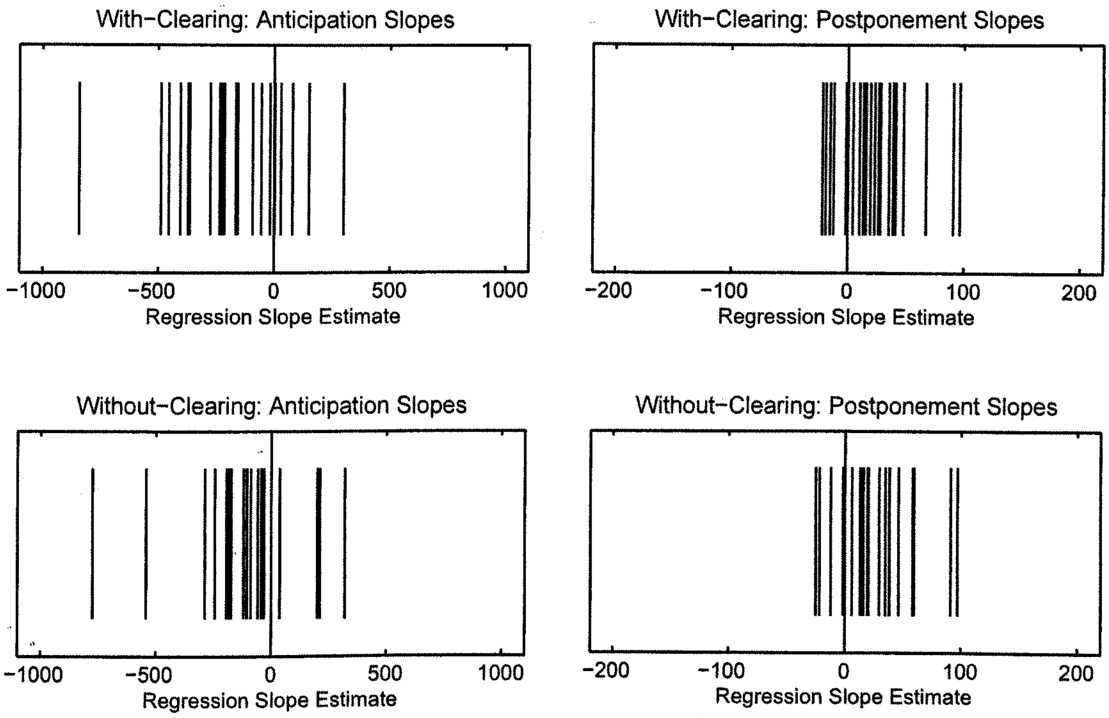

Figure 4-9 *Individual regression slope parameter estimates for the latency-displacement functions of Experiment 3. Each vertical line in a panel represents a regression slope estimate for a participant for a particular experimental condition. The left hand panels show slope estimates for anticipations, whilst the right hand panels show slope estimates for postponements. The top panels show the slope estimates for the with-clearing condition, whilst the bottom panels show the slope estimates for the without-clearing condition. Note—the scales of the x-axes for the slope estimates for anticipations and postponements are different.*

Specifically, the functions exhibited steep negative slopes for transpositions that were anticipations, and shallow positive slopes for transpositions that were postponements, consistent with the error latency prediction of a model combining a primacy gradient, positional marking, and response suppression. That this empirical pattern was observed in the without-clearing condition, where the onus was on participants to prevent erroneous repetitions in their responding, confers support for a contribution of response suppression to the serial reconstruction of visual sequences.

A role for response suppression is further indicated by the absence of a significant difference in serial reconstruction accuracy between the with-clearing and without-clearing conditions. If response suppression did not contribute to serial reconstruction performance in the without-clearing condition then erroneous repetitions should have been frequent in people's reports thereby lowering the accuracy of serial reconstruction in this condition relative to the with-clearing reconstruction condition. However, such erroneous repetitions were extremely scarce, accounting for less than 1% of all responses (an occurrence rate below that expected by chance), with the two occurrences of the repeat being separated on average by a lag of 3.11 positions. The scarcity of erroneous repetitions suggests that once an item was selected in the reconstruction array the representation of that item (or its spatial loci within the reconstruction array) was suppressed, thereby reducing the likelihood that it would be chosen again. That the two occurrences of the repeat tended to be spaced several positions apart, is consistent with the notion that this response suppression wore off gradually over time. It is noteworthy that the incidence of erroneous repetitions, and their average separation across serial positions, is similar to that observed in verbal serial recall. For example, Henson (1996) found that repetitions comprised approximately 2% of all responses, and that the two occurrences of the repeat were separated on average by a lag of 3.34 positions.

In short, the empirical pattern of the LDFs observed in the current experiment provide further support for the role of a primacy gradient of activations and positional marking in visual serial memory. Critically, the pattern of transposition latencies in the without-clearing condition in conjunction with the low incidence of erroneous repetitions confers support for the response suppression component of this model. Subsequent experiments in this chapter explore further the generality of the pattern of transposition latencies witnessed in the experiments so far. Given the absence of significant differences between the with-clearing and without-clearing conditions the following experiments revert back to the use of the former reconstruction method.

Experiment 4

The experiments reported so far have consistently demonstrated LDFs characterised by steep negative slopes for anticipations and shallow positive slopes for postponements. This empirical pattern has been shown to hold across manipulations of sequence length (Experiments 1 and 2), articulatory suppression (Experiment 2), and the dynamics of the reconstruction array (Experiment 3). Experiment 4 sought to establish further the generality of this empirical pattern by incorporating a temporal grouping manipulation. Recall from Chapter 1 that temporal grouping involves inserting one or more extended pauses in an otherwise temporally homogenous sequence, in order to differentiate sequence members into sub-groups, usually of three items. In verbal studies, grouping has been shown to exert a multiplicity of effects on serial memory performance relative to an ungrouped baseline. These effects include an elevation in recall accuracy (Henson, 1999; Hitch et al., 1996; Ryan, 1969), a change in the shape of the accuracy serial position curve reflected by the emergence of mini within-group primacy and recency effects (Hitch et al., 1996), as well as a change in recall latencies typified by peaks in the response latency serial position curve for group-initial items (Anderson & Matessa, 1997; Farrell, 2008; Farrell & Lewandowsky, 2004; Maybery et al., 2002). Grouping also engenders a reduction in adjacent-neighbour transpositions, coupled with an elevation in between-group transpositions that preserve their within-group positions (Henson, 1999; Ng & Maybery, 2002; Ryan, 1969), a class of errors dubbed by Henson (1996) as interpositions. These effects notwithstanding, Farrell and Lewandowsky (2004) showed that the LDFs underpinning verbal serial recall were generally insensitive to manipulations of the temporal grouping of items.

Temporal grouping effects have not previously been examined in visual serial memory so the current experiment is part exploratory. However, some effects of temporal grouping have been demonstrated in visual-spatial (Parmentier et al., 2006) and auditory-spatial (Parmentier et al., 2004) serial memory, providing precedent for an examination of grouping effects in the visual non-spatial domain.

Method

Participants & stimuli

Twenty-four participants recruited from the student population at the University of York took part in the experiment in exchange for course credit or an honorarium of £5. The stimuli were the same as those employed in Experiment 3.

Design & Procedure

The experiment manipulated a single independent variable: Sequence-Type (ungrouped / grouped), which was a within-subjects factor. Half of the participants received the ungrouped sequences followed by the grouped sequences, while the remaining half received the sequences in the converse order.

The procedure was the same as for Experiment 1 with the following exceptions: participants completed only a single experimental session; in the grouped condition the temporal interval separating the third and fourth item in the list was extended from 500ms to 1500ms to divide the sequences into two sub-groups of three items.

Participants attempted 80 experimental trials for each sequence-type, which were preceded by two practice trials. Enforced 30-second rest periods were included after every 20 experimental trials. The experiment lasted approximately 60 minutes.

Results

Accuracy serial position curves and transposition gradients

The accuracy serial position curves can be inspected in Figure 4-10A. A small benefit of temporal grouping is evident, manifesting as an elevation in performance at medial serial positions. These data were subjected to a 2 (Condition) X 6 (Serial Position) ANOVA, which failed to reveal a reliable main effect of Condition, $F(1, 22) = 1.422$, $MSE = .019$, $p = .246$, however, the main effect of Serial Position, $F(5, 110) = 47.300$, $MSE = .502$, $p < .001$, and the Condition X Serial Position interaction, $F(5, 110) = 7.694$, $MSE = .022$, $p < .001$, were both significant. The source of the interaction was an elevation in performance at positions three and four in the grouped condition.

From inspection of the underlying transposition gradients shown in Figure 4-10B, it can be seen that although the proportions of transpositions with an absolute displacement value of one are slightly smaller for the grouped condition (which arose due to a reduction in the probability of adjacent-transpositions spanning the group boundary), there is no evidence for a corresponding

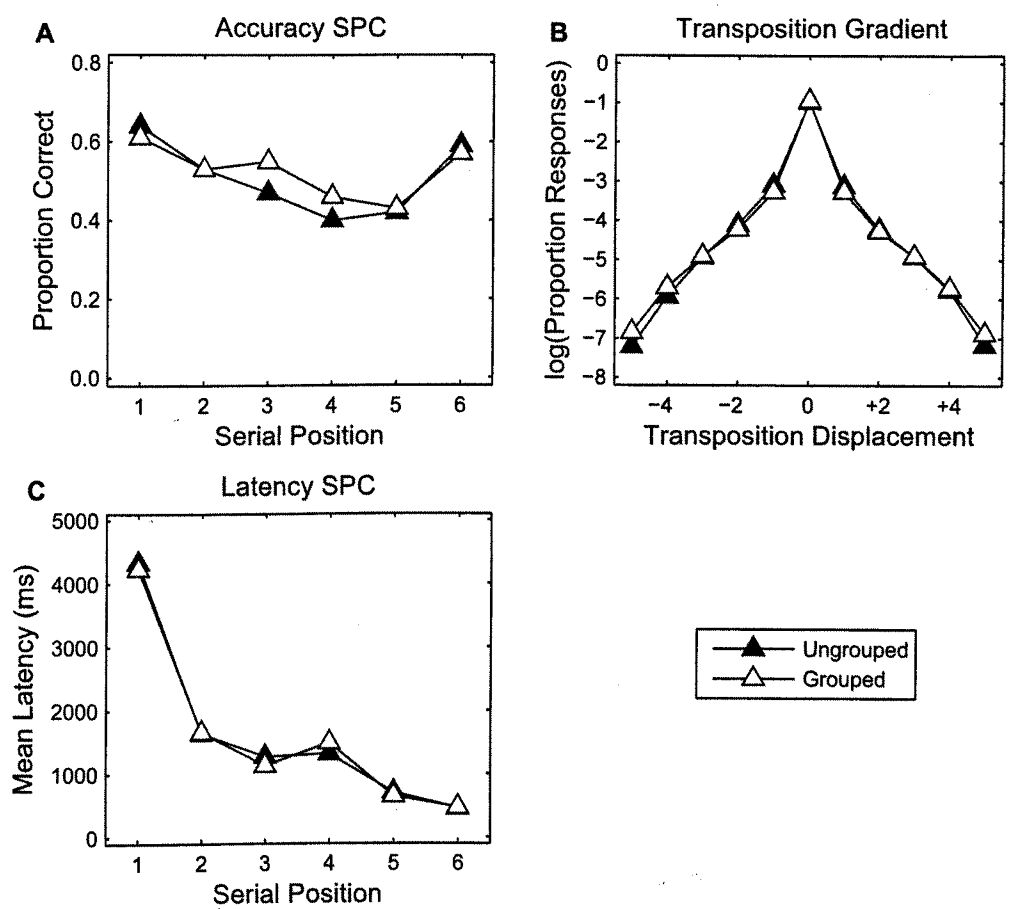

Figure 4-10 *Serial memory performance measures for Experiment 4. Panels show accuracy serial position curves (A), transposition gradients, (B) and latency serial position curves (C).*

increase in the incidence of interposition errors, as reflected by -3 and +3 transposition displacement values.

Latency serial position curves and latency-displacement functions

Turning to the latency data, Figure 4-10C shows the latency serial position curves associated with correct responses. It can be seen that there is a small elevation in the profile of the function for the grouped condition at the fourth serial position relative to positions three and five, which is consistent with an effect of grouping on response timing. However, although a 2 (Condition) X 6 (Serial Position) ANOVA on these data identified a significant main effect of Serial Position, $F(5, 110) = 186.576$, $MSE = 3.336E8$, $p < .001$, neither the main effect of condition, $F(1, 22) = 0.65$, $MSE = 17872.612$, $p = .80$, nor the Condition X Serial Position interaction, $F(5, 110) = 1.038$, MSE

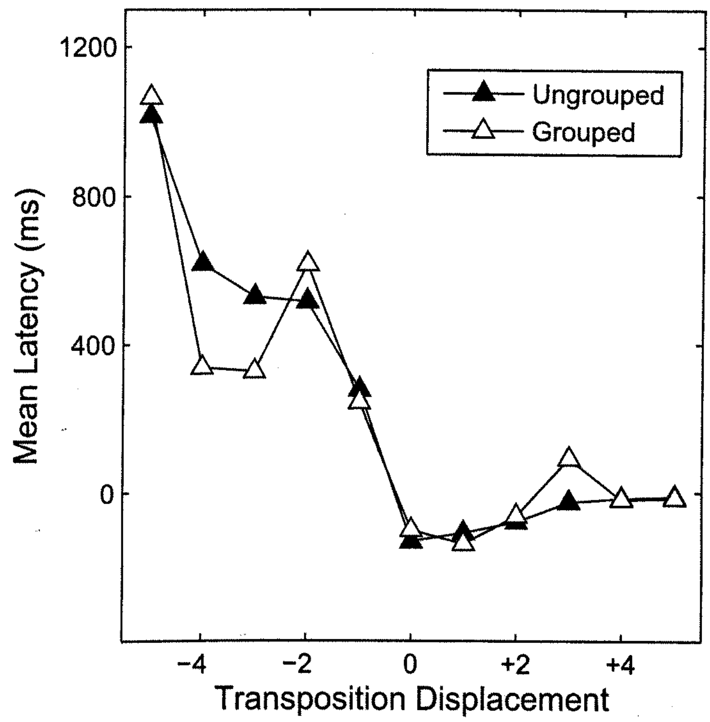

Figure 4-11 *Latency-displacement functions for Experiment 4.*

= 334896.176, p = .366, reached significance. The absence of a significant interaction is due to a slight elevation in the profile of the ungrouped latency serial position curve, also at the fourth serial position.

The LDFs with the effects of output position subtracted can be inspected in Figure 4-11. The slopes of the functions for anticipations are steeply negative, whereas the slopes of the functions for postponements are shallowly positive. It is evident that the LDF for the grouped condition exhibits local discontinuities for -3, -4, and +3 transposition displacement values that are not present for the ungrouped condition. The data were analysed by performing regression analyses on the LDFs for each individual participant examining the relationship between transposition latency and transposition displacements that were anticipations (displacements in the range -5 to 0) and postponements (displacements in the range 0 to +5) separately. The regression slope parameter estimates for anticipations and postponements were subsequently pooled together and subjected to one-sample t-tests to determine whether they deviated reliably from zero.

Parameter	Estimate	*SE*	*t*	*p*
Ungrouped				
Anticipation	-220.74	75.35	-2.930	.01
Postponement	23.32	7.71	3.026	.01
Grouped				
Anticipation	-189.00	58.73	-3.218	.00
Postponement	30.85	7.37	4.188	.00

Table 4-4 *Mean regression slope parameter estimates for the latency-displacement functions of Experiment 4.*

The mean regression slope parameter estimates associated with anticipations and postponements for ungrouped and grouped sequences can be inspected in Table 4-4. The mean regression slope parameter estimates for anticipations were steeply negative and differed reliably from zero: $t(23) = -2.930$, $p = .001$, for ungrouped sequences, and, $t(23) = -3.218$, $p < .01$, for grouped sequences. In comparison, the mean regression slope parameter estimates for postponements were shallowly positive, but also differed reliably from zero: $t(23) = 3.026$, $p < .01$, for ungrouped sequences, and, $t(23) = 4.188$, $p < .001$, for grouped sequences.

Figure 4-12 shows the regression slope parameter estimates for anticipations and postponements for individual participants for ungrouped and grouped sequences. It is apparent from inspection of this figure that the majority of participants contributed steep negative slope estimates for anticipations and shallow positive slope estimates for postponements.

Discussion

The qualitative pattern of the LDFs once again parallels that documented in the previous experiments. Transposition latency was an overall negative function of transposition displacement, but with a slight elevation in the slopes of the functions for postponements. However, there was some heterogeneity in the LDFs for ungrouped and grouped sequences. Specifically, the function for grouped sequences exhibited discontinuities at -4, -3, and +3 transposition displacement values that were not seen in the function for ungrouped sequences. From the perspective of positional

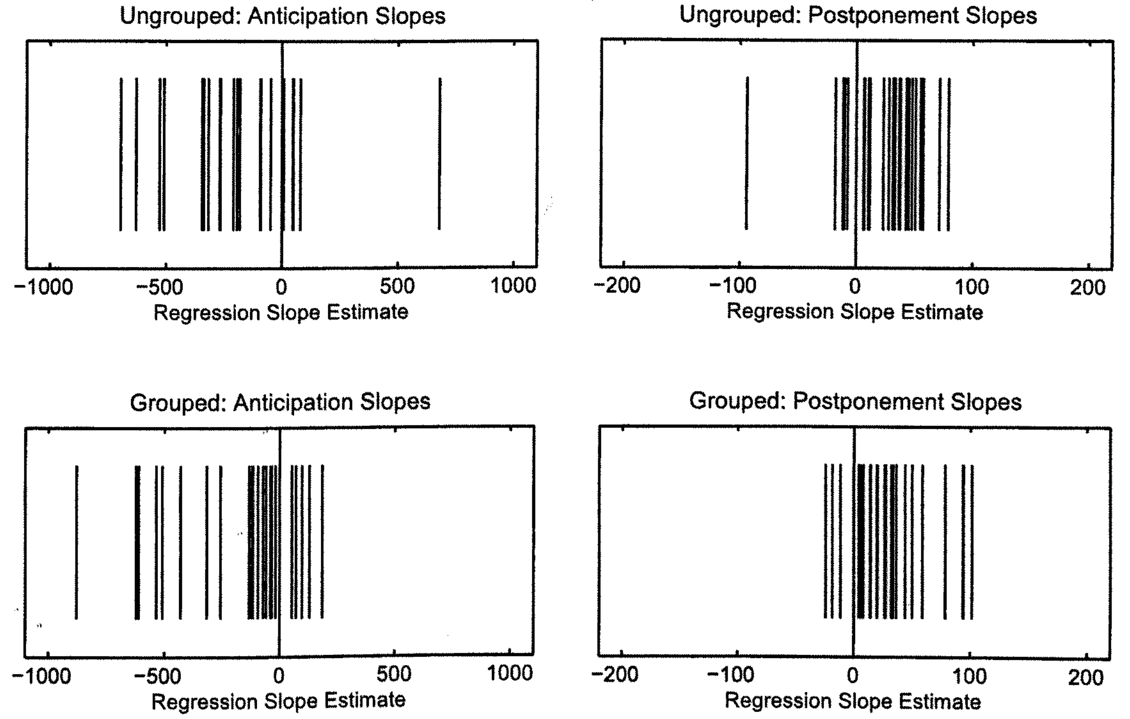

Figure 4-12 *Individual regression slope parameter estimates for the latency-displacement functions of Experiment 4. Each vertical line in a panel represents a regression slope estimate for a participant for a particular experimental condition. The left hand panels show slope estimates for anticipations, whilst the right hand panels show slope estimates for postponements. The top panels show the slope estimates for the ungrouped condition, whilst the bottom panels show the slope estimates for the grouped condition. Note—the scales of the x-axes for the slope estimates for anticipations and postponements are different.*

models of serial recall (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998a; Lewandowsky & Farrell, 2008), grouping would be expected to accelerate latencies for interposition errors, reflected by -3 and +3 transpositions. This is because for grouped sequences, these models assume that one dimension of ordering represents the positions of items within-groups, which necessarily increases the positional overlap of items in different groups that occupy the same within-group positions. However, although latencies were indeed accelerated for -3 transpositions, so too were those for -4 transpositions. Moreover, the latencies for +3 transpositions were decelerated, contrary to what positional models would predict.

The temporal grouping manipulation exerted only a limited impact on the remaining serial memory performance measures. Relative to an ungrouped baseline, grouping engendered an increase in serial position accuracy, as in studies of verbal (Henson, 1996, 1999; Hitch et al., 1996; Ryan, 1969) and spatial (Parmentier et al., 2004, 2006) serial memory. However, in the current experiment this increase was restricted to a within-group recency effect at serial position three and a within-group primacy effect at serial position four, which prevented the usual main effect of grouping from materializing. Although there is some suggestion of an effect of grouping on the response latency serial position curve owing to an elevation in the latency for the first item in the second group, this was much smaller in magnitude than that observed in verbal (Farrell, 2008; Farrell & Lewandowsky, 2004; Maybery et al., 2002) and spatial (Parmentier et al., 2004, 2006) studies. Finally, although grouping fostered a reduction in adjacent-neighbour transpositions there was no evidence for a corresponding increase in the incidence of interpositions. Potential reasons for the absence of more robust effects of temporal grouping are explored in the general discussion.

Experiment 5

The limitation of the temporal grouping manipulation employed in Experiment 4 was that it exerted only a small impact on performance. Accordingly, in Experiment 5 a manipulation was employed that has already been shown to exhibit robust and reliable effects on visual serial memory performance – that of the visual similarity of items. Across three experiments employing randomly filled visual matrices as stimuli, Avons and Mason (1999) found that serial memory performance for sequences composed of similar matrices was poorer than for sequences of dissimilar matrices. This visual similarity effect mirrors the well established phonological similarity effect found in verbal serial memory studies in which phonologically similar items are recalled with reduced accuracy relative to phonologically dissimilar items (Baddeley, 1966; Conrad 1964). Of more direct relevance to the current study, the visual similarity effect reported by Avons and Mason has been replicated by Smyth et al. (2005) using sequences of unfamiliar faces as stimuli.

To maximise the chance of obtaining a robust visual similarity effect, in the current study a small set of highly similar and dissimilar faces was employed as stimuli. This also permitted an examination of whether the LDFs obtained in Experiments 1 to 4 using unique (or almost unique) items on each trial generalized to conditions involving repeated sampling from a small item set. In addition, an articulatory suppression manipulation was incorporated to prevent participants from verbally encoding the faces. The possibility of verbal encoding is inflated with a small item set, because repeated presentations of the same faces renders it easier for participants to extract verbal descriptors for those stimuli that can then be rehearsed as a sequence of verbal tokens.

In Experiment 5, the sequence length was also reduced to five-items to avoid potential floor effects for similar sequences that might otherwise obscure the interpretation of the LDF for this condition.

Method

Participants

Twenty-four participants recruited from the student population at the University of York took part in the experiment in exchange for course credit or an honorarium of £4.

Stimuli

The stimuli were sequences of five unfamiliar faces drawn randomly from one of four stimulus ensembles. Two of the ensembles each contained six similar faces, while the remaining two ensembles each contained six dissimilar faces. The similar ensembles consisted of faces of the same gender that were defined by the experimenter to share similar skin and hair colour, face shape, and hairstyle. The dissimilar ensembles each contained three male and three female faces that varied along the remaining dimensions used to categorize the similar faces.

Design & Procedure

The experiment manipulated a single independent variable: Sequence-Type (dissimilar / similar), which was a within-subjects factor. Dissimilar and similar sequences were presented in random order.

The procedure was identical to that of Experiment 2. Participants attempted a single practice trial for each sequence-type initially, followed by 120 experimental sequences, half of which were composed of dissimilar items and half of which were composed of similar items. Enforced 30-second rest periods were included after every 20 trials. The experiment lasted approximately 45 minutes.

Results

Accuracy serial position curves and transposition gradients

The accuracy serial position curves are shown in Figure 4-13A from which it is clearly visible that serial reconstruction accuracy was higher for dissimilar than for similar sequences. Statistical confirmation of this pattern was obtained by entering the data into a 2 (Sequence-Type) X 5 (Serial Position) ANOVA, which revealed reliable main effects of both Sequence-Type, $F(1, 23) = 40.852$, $MSE = .756$, $p < .001$, and Serial Position, $F(4, 92) = 45.158$, $MSE = .440$, $p < .001$. The interaction between the two factors was also significant, $F(4, 92) = 14.562$, $MSE = .073$, $p < .001$, reflecting the impact of the similarity manipulation was strongest for the first serial position and then gradually decreased across positions. The underlying transposition gradients shown in Figure 4-13B confirm that the effect of the similarity manipulation was to increase the proportions of anticipations and postponements for similar sequences relative to dissimilar sequences.

Latency serial position curves and latency-displacement functions

Turning now to the latency data, Figure 4-13C shows the response latency serial position curves associated with correct responses. It is visible from inspection of this figure that response latencies were longer for the first two serial positions for similar than for dissimilar sequences. These data were subjected to a 2 (Sequence-Type) X 5 (Serial Position) ANOVA, which revealed significant main effects of Sequence-Type, $F(1, 23) = 78.834$, $MSE = 7125349.342$, $p < .001$, and Serial Position, $F(4, 92) = 114.762$, $MSE = 1.898E8$, $p < .001$, as well as a significant interaction between the two factors, $F(4, 92) = 42.267$, $MSE = 8310012.994$, $p < .001$. The source of the

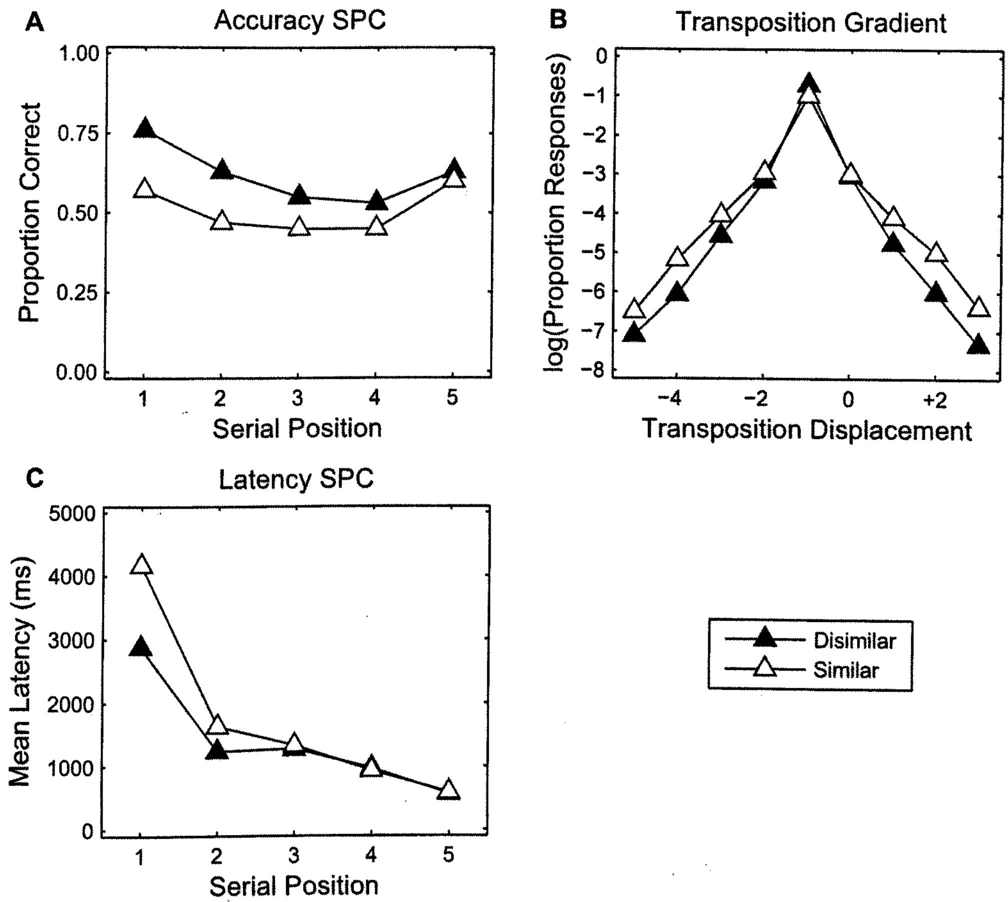

Figure 4-13 *Serial memory performance measures for Experiment 5. Panels show accuracy serial position curves (A), transposition gradients, (B) and latency serial position curves (C).*

interaction was the longer latencies associated with responses at positions one and two for similar sequences.

Figure 4-14 shows the LDFs with the effects of output position removed. The empirical pattern is similar to that observed previously – the slopes of the functions for anticipations are steeply negative, whereas the slopes of the functions for postponements are shallowly positive. Interestingly, the latencies accompanying anticipations are considerably longer for dissimilar than for similar sequences, but there is no consistent difference between the two conditions for the latencies accompanying postponements. The LDFs were analysed by regressing transposition latency on transposition displacements that were anticipations (displacements in the range -4 to 0) and postponements (displacements in the range 0 to +4) separately for each individual participant.

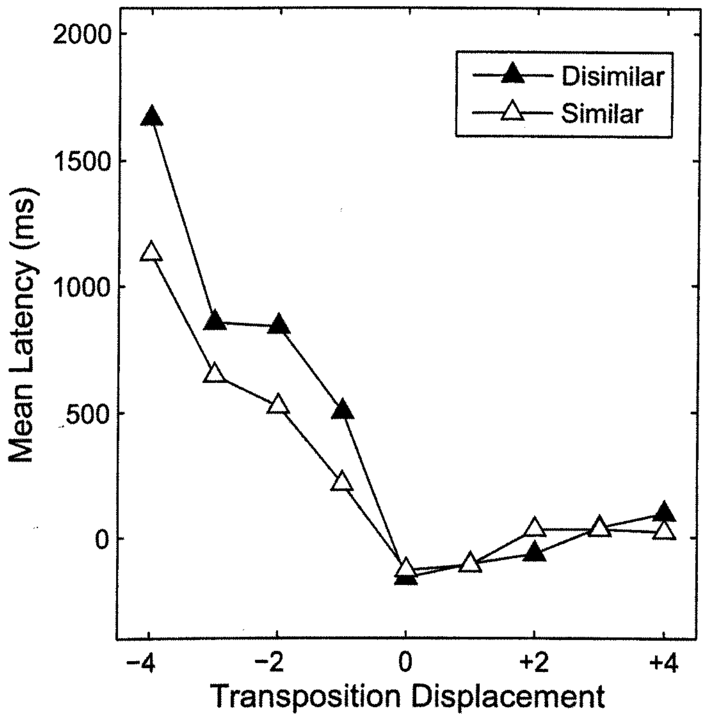

Figure 4-14 *Latency-displacement functions for Experiment 5.*

The regression slope parameter estimates were then pooled together and evaluated using one-sample t-tests to determine whether they differed reliably from zero.

The mean regression slope parameter estimates for anticipations and postponements are summarised in Table 4-5. As can be seen from inspection of this table, the mean regression slope parameter estimates for anticipations were steeply negative and differed reliably from zero: $t(23) = -6.116, p < .001$, for dissimilar sequences and, $t(23) = -4.138, p < .001$, for similar sequences. By contrast, the mean regression slope parameter estimates for postponements were shallowly positive, but also differed reliably from zero: $t(23) = 3.933, p = .001$, for dissimilar sequences and, $t(23) = 4.732, p < .001$, for similar sequences.

Figure 4-15 shows the regression slope parameter estimates of individual participants for anticipations and postponements for dissimilar and similar sequences. For both sequence-types, the majority of participants contributed steep negative slope estimates for anticipations and shallow positive slope estimates for postponements.

Parameter	Estimate	*SE*	*t*	*p*
Dissimilar				
Anticipation	-394.75	64.55	-6.116	.00
Postponement	76.48	19.44	3.933	.00
Similar				
Anticipation	-295.62	71.45	-4.138	.00
Postponement	43.75	9.25	4.732	.00

Table 4-5 *Mean regression slope parameter estimates for the latency-displacement functions of Experiment 5.*

Discussion

Consistent with previous studies (Avons & Mason, 1999; Smyth et al., 2005), the current experiment demonstrated robust effects of item similarity on visual serial order memory. Sequences containing visually similar items were reconstructed with reduced accuracy relative to sequences composed of visually dissimilar items. Critically, the LDFs for dissimilar and similar sequences were qualitatively similar and consistent with the empirical pattern documented in previous experiments providing further support for the role of a primacy gradient and positional marking in visual serial memory. The current experiment additionally shows that the empirical pattern of the LDFs observed in the previous experiments using sequences drawn from an open stimulus set, generalizes to the use of sequences drawn from a closed stimulus set.

One intriguing property of the LDFs observed in the current experiment is the longer latencies associated with anticipation, but not postponement errors, for dissimilar sequences. To facilitate interpretation of this outcome, I begin by considering some aspects of the latency data that were not mentioned in the results section. Inspection of Figure 4-13C shows that the latencies accompanying correct responses were longer for similar than for dissimilar sequences for the first two serial positions, but that no difference between the two conditions was observed for the remaining positions. However, the empirical pattern changes when the serial position data are plotted in terms of errors, rather than correct responses. Specifically, the latencies accompanying errors for dissimilar sequences are considerably longer for the first three serial positions than those for similar

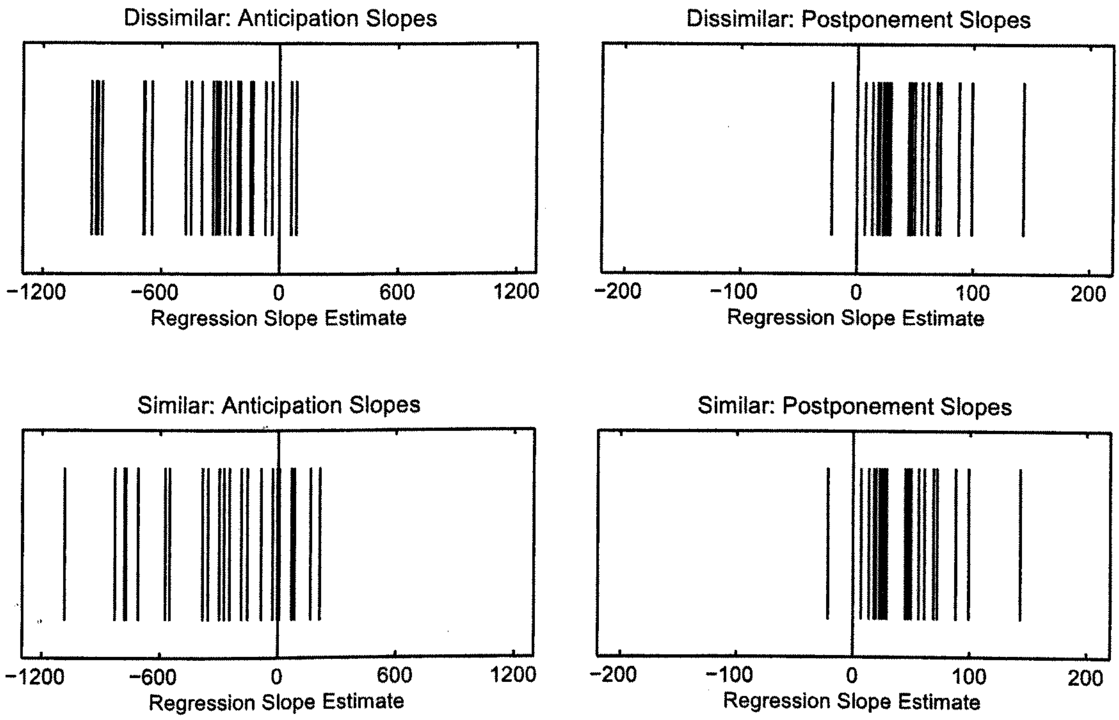

Figure 4-15 *Individual regression slope parameter estimates for the latency-displacement functions of Experiment 5. Each vertical line in a panel represents a regression slope estimate for a participant for a particular experimental condition. The left hand panels show slope estimates for anticipations, whilst the right hand panels show slope estimates for postponements. The top panels show the slope estimates for dissimilar sequences, whilst the bottom panels show the slope estimates for similar sequences. Note—the scales of the x-axes for the slope estimates for anticipations and postponements are different.*

sequences, whereas only a negligible difference is observed for the final two positions (data not shown). Recall that errors at early output positions will tend to be anticipations, whereas errors at late output positions will tend to be postponements. Since the error latencies only diverge considerably for the first few output positions this accounts for the selective effect of the similarity manipulation on the latencies for anticipation errors.

The question which arises is why the error latencies were generally longer for dissimilar than for similar sequences. I now seek an explanation for this result by considering how the similarity mechanisms underpinning theories of verbal short-term memory might accommodate it. A number of theories explain item similarity effects by postulating a two-stage recall mechanism (e.g., Burgess & Hitch, 1999, 2006; Henson, 1998; Page & Norris, 1998). In the first stage sequence items undergo an order-based competition determined by the item activations elicited by the ordering mechanism driving recall. The winning item from this first stage is subsequently forwarded to a further stage where it undergoes a second, similarity-based competition. Critically, similar items encounter more competition in this second stage than do dissimilar items. In terms of the response selection architecture delineated in Chapter 2, this two-stage mechanism could be modelled by adding a second lateral inhibition layer in which items recalled from the first layer enter into several additional iterative competitive selection cycles before they are recalled.

It is easy to see that this mechanism would necessarily predict longer error latencies for anticipations and postponements for similar sequences. This is because the greater competition encountered between similar items compared to dissimilar items in the second similarity-based confusion stage will take longer to resolve. This prediction is incompatible with the data reported here showing that error latencies are longer for dissimilar than for similar sequences, at least for the first few output positions, and that consequently the latencies for anticipation errors are longer for dissimilar sequences. Of course, this two-stage mechanism was developed to explain item similarity effects in verbal short-term memory and so these results only compromise the viability of such a mechanism when applied to the visual domain. Nevertheless, recent studies have also called into question the feasibility of a two-stage similarity mechanism applied to verbal short-term memory (see e.g., Farrell, 2006; Farrell & Lewandowsky, 2003).

It is possible to sketch out an alternative interpretation of the results, based upon a single stage similarity mechanism, as employed in the C-SOB model of Farrell and Lewandowsky (Farrell, 2006; Lewandowsky & Farrell, 2008). Unlike the two-stage mechanism described above, C-SOB explains item similarity effects by assuming a single memorial stage within which item similarity interacts with the encoding of serial order information. C-SOB is a model of short-term memory in which serial order is represented by forming associations between items and positional markers. The strength of these associations across positions is governed by a mechanism known as *novelty-sensitive encoding*. To explain, the strength of the association between each study item and its corresponding position marker is determined by comparing that item with the current contents of memory. If the item is novel with respect to the stored contents of memory then it will be encoded strongly, whereas if the item is similar to the stored contents of memory then it will be encoded weakly. Novelty-sensitive encoding naturally gives rise to a primacy gradient in association strength across positions, because each item entering memory will bear some approximate resemblance to pre-encoded items. Critically, under novelty-sensitive encoding the primacy gradient will be steeper for sequences of similar items than for sequences of dissimilar items, because the overlap between items on similar sequences means they will be encoded with less strength than their counterparts on dissimilar sequences. As well as incorporating a primacy gradient and positional markers, C-SOB incorporates response suppression during recall.

Because C-SOB uses a dynamic recall architecture it generates both response probability and latency predictions. Lewandowsky and Farrell (2008) have shown that the model generates qualitatively similar LDFs to the PG+PM+RS model. Although the authors did not compare the LDFs predicted for dissimilar and similar sequences, C-SOB would appear to qualitatively predict longer error latencies for anticipations and postponements on dissimilar sequences. This is because the stronger associations between items and their positional markers on dissimilar sequences means that when those markers are reinstated at recall the difference in the relative degree by which the target item and neighbouring items are cued will be greater than would be the case for items on similar sequences. Accordingly, for an item to be transposed on a dissimilar sequence it will have to overcome a stronger set of competitor items than a corresponding item on a similar sequence and this competition process will consequently take longer to resolve.

This prediction was borne out for anticipation errors, but not for postponement errors. However, the failure to observe a difference in error latencies for postponements is due to the limited impact of the similarity manipulation on error latencies for the last two output positions (remember that errors at these positions will mostly consist of postponements). Why a stronger effect of the similarity manipulation on error latencies was not observed at these positions is not clear, but it may reflect that some of the effects of similarity were absorbed in the latencies for responses at earlier positions, as might be expected if participants planned their recalls several positions ahead. The current data are therefore tentatively consistent with the view that similarity effects in visual serial memory are the result of a single-stage rather than a dual-stage similarity mechanism. Moreover, they lend some credibility to the hypothesis that the primacy gradient identified as contributing to the representation of serial order in the current experiments is an emergent property of a novelty-sensitive encoding mechanism, as posited by the C-SOB model.

Experiment 6

The preceding experiments have consistently demonstrated LDFs characterised by steep negative anticipation slopes and shallow positive postponement slopes. Although this empirical pattern is consistent with that observed in the verbal studies of Farrell and Lewandowsky (2004) the experiments of these authors utilised a serial recall paradigm, whereas the experiments documented here employed a serial reconstruction paradigm. Furthermore, the current experiments have generally used unique items across trials, whereas Farrell and Lewandowsky used a closed item set of digits or consonants. It follows that a better assessment of the similarities and dissimilarities between the LDFs for visual and verbal stimuli can be obtained via a direct comparison between the two domains when the task requirements are equated as far as possible.

Accordingly, Experiment 6 compared the LDFs underpinning visual and verbal serial memory when the same recall paradigm – serial reconstruction – was employed in a within-subjects design. The stimuli for the visual task were sequences composed of five unfamiliar faces, whereas the stimuli for the verbal task were sequences composed of six words. Both tasks employed unique stimulus items across trials. The slightly longer sequence length for the verbal task (combined with the use of long words) was chosen to try and equate overall reconstruction accuracy in the two tasks. Given the absence of an output mechanism for rehearsing visual non-spatial stimuli it was anticipated that reconstruction performance for the visual task would otherwise be considerably lower than for the verbal task. Although the use of different sequence lengths induced a discrepancy between the possible ranges of transposition displacements in the two tasks, this was deemed acceptable, since interest centred on the similarities in the qualitative rather than the quantitative characteristics of the LDFs.

Method

Participants

Eighteen participants recruited from the student population at the University of York took part in the experiment in exchange for course credit or an honorarium of £4.

Stimuli

The stimuli for the verbal task were sequences of six nouns drawn randomly without replacement from an ensemble of 550 nouns. These were obtained from the MRC Psycholinguistic Database and were selected according to the constraints that each contained between 6 and 8 letters and had a Kucera and Francis (1968) written word frequency of between 10 and 40. The nouns were presented in point 18 Arial lower-case font.

The stimuli for the visual task were sequences of five unfamiliar faces drawn randomly without replacement from an ensemble of 550 faces. Half of the sequences were composed of male faces and half were composed of female faces.

Design & Procedure

The experiment manipulated a single independent variable: Sequence-Type (Words / Faces), which was a within-subjects factor. Half of the participants received the word-sequences followed by the face-sequences; with the remaining half of participants receiving sequences in the converse order.

Participants were tested individually in the presence of the experimenter. A trial was initiated when the participant selected a 'begin trial' icon located in the centre of the computer display using the computer mouse. A fixation cross then appeared for 1000ms in the central screen position and was replaced by a sequence of items presented in succession. For sequences composed of faces each item was presented for 500ms followed by a 500ms inter-stimulus interval. For sequences composed of words, each item was also presented for 500ms, but the inter-stimulus interval was 300ms. The different inter-stimulus intervals were chosen to ensure that the encoding times for the two sequence-types were the same (4500ms) even though they contained different numbers of items. For both sequence-types, the final item was followed by a 1000ms interval.

Once sequences had been presented the items simultaneously appeared on-screen in a noisy circular array, the centre of which contained a question mark. Situated at the bottom of the screen were a series of initially empty response boxes representing output positions. Participants were required to select the items in the order they remembered them being presented using the mouse-driven pointer. Once an item had been selected it was cleared from the reconstruction array and the appropriate response box illuminated green indicating that a response had been registered for that position. If participants were unsure of their response for a given position they could select the question mark, which recorded a 'don't know response'. Once all the response boxes had been filled participants selected a 'validate input' icon located in the bottom right hand corner of the screen. This was followed by the presentation of the reconstruction time for the sequence in the central screen position for 3000ms, which was subsequently replaced by the 'begin trial' icon for the next trial.

Participants completed 80 trials for each sequence-type, which were preceded by two practice trials. Enforced 30-second rest periods were included after every 20 experimental trials. The experiment lasted approximately 50 minutes.

Results

Accuracy serial position curves and transposition gradients

The accuracy serial position curves are shown in Figure 4-16A. These were similar for sequences composed of words and faces, both being characterised by pronounced primacy and restricted recency. However, the mean proportion of correct responses was significantly greater for sequences composed of words (.72), than faces (.64), $t(17) = -2.412$, $p < .05$, despite the use of longer sequences in the former task to try and equate performance. The transposition gradients underlying the serial position curves, which can be inspected in Figure 4-16B confirm the presence of more anticipations and postponements for sequences made up of faces.

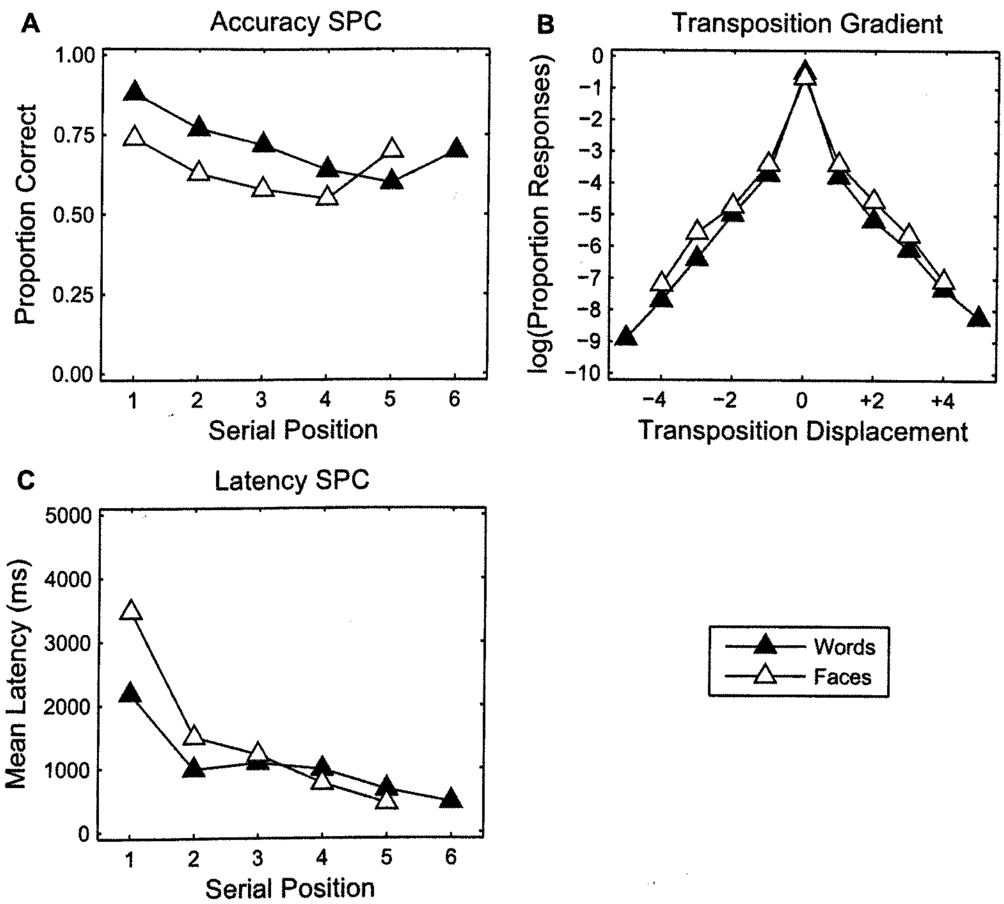

Figure 4-16 *Serial memory performance measures for Experiment 6. Panels show accuracy serial position curves (A), transposition gradients, (B) and latency serial position curves (C).*

Latency serial position curves & latency-displacement functions

The response latency serial position curves associated with correct responses can be inspected in Figure 4-16C. The serial position curves for both sequence-types exhibit a long latency for the first item; however, the relationship between latency and serial position is different in the two cases. Specifically, for sequences composed of faces latency is a monotonically decreasing function of serial position, whereas for sequences of words when the first response latency is ignored the latency serial position curve exhibits an inverted U shape trend.

The LDFs with the effects of output position subtracted are shown in Figure 4-17 and are qualitatively similar for sequences of words and faces. Both functions exhibit a steep negative slope for anticipations and a shallow positive slope for postponements. The data were analysed as in previous experiments by conducting regression analyses on the LDFs for each individual

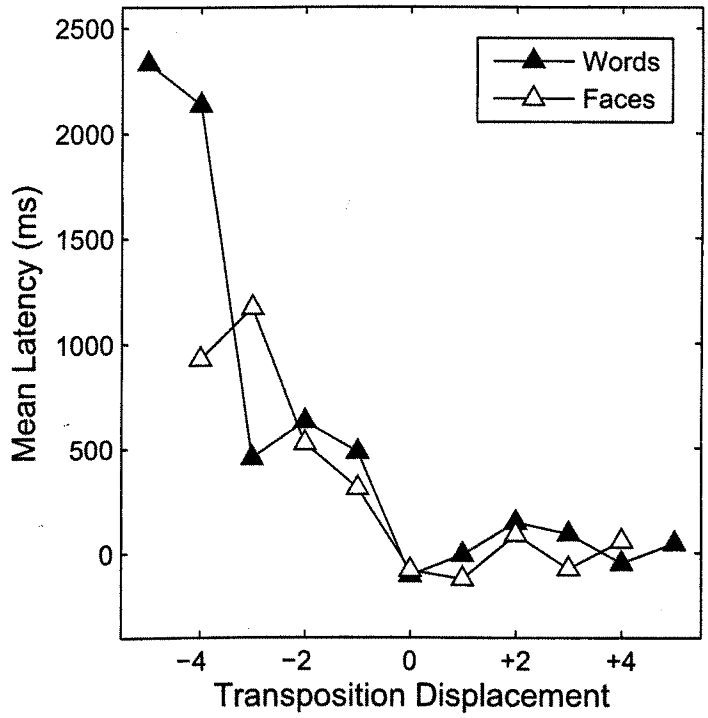

Figure 4-17 *Latency-displacement functions for Experiment 6.*

participant examining the relationship between transposition latency and transposition displacements that were anticipations (displacements in the range -4 to 0 for sequences of faces and -5 to 0 for sequences of words) and postponements (displacements in the range 0 to +4 for sequences of faces and 0 to +5 for sequences of words) separately. The regression slope parameter estimates were then pooled together and evaluated using one-sample t-tests to determine whether they deviated reliably from zero.

The mean regression slope parameter estimates for anticipations and postponements can be inspected in Table 4-6. Consistent with the empirical pattern that has now become familiar the mean regression slope parameter estimates for anticipations were steeply negative and differed reliably from zero: $t(17) = -4.200, p < .001$, for sequences of words, and $t(17) = -3.131, p < .01$, for sequence of faces. In contrast, the slopes of the functions for postponements were shallowly positive, but also differed reliably from zero: $t(17) = 2.309$, $p < .05$, for sequences of words, and $t(17) = 3.148$, $p < .01$, for sequences of faces. Figure 4-18 shows that for both sequence-types the majority of participants contributed negative slopes for anticipations and positive slopes for postponements.

Parameter	Estimate	*SE*	*t*	*p*
Words				
Anticipation	-424.91	101.18	-4.200	.00
Postponement	33.19	14.38	2.309	.03
Faces				
Anticipation	-338.11	108.00	-3.131	.00
Postponement	44.19	14.04	3.148	.01

Table 4-6 *Mean regression slope parameter estimates for the latency-displacement functions of Experiment 6.*

Discussion

In brief, the LDFs for serial reconstruction of sequences composed of visual and verbal stimuli were qualitatively similar, both being characterised by a steep negative slope for anticipations and a shallow positive slope for postponements. This empirical outcome confirms that when the recall methodology is equated comparable LDFs are observed for both types of material. The original experiments examining transposition latencies, conducted by Farrell and Lewandowsky (2004), employed digits and consonants as verbal stimuli (necessarily a closed stimulus set), and a serial recall paradigm. The current experiment therefore extends the generality of the findings of these authors by showing that the LDFs observed in their study extend to a serial reconstruction of order task in which the stimuli were words drawn from an open stimulus set. Importantly, the current experiment provides joint evidence for the role of a primacy gradient of activation and positional marking in visual and verbal serial memory, based upon a direct comparison between the two domains using a repeated measures design.

Summary of experiments

The experiments reported in this chapter have consistently revealed LDFs for visual stimuli characterised by steep negative anticipation slopes and shallow positive postponement slopes. This outcome is consistent with the error latency prediction of a model in which serial order is represented via a primacy gradient of activation levels, associations between items and positional

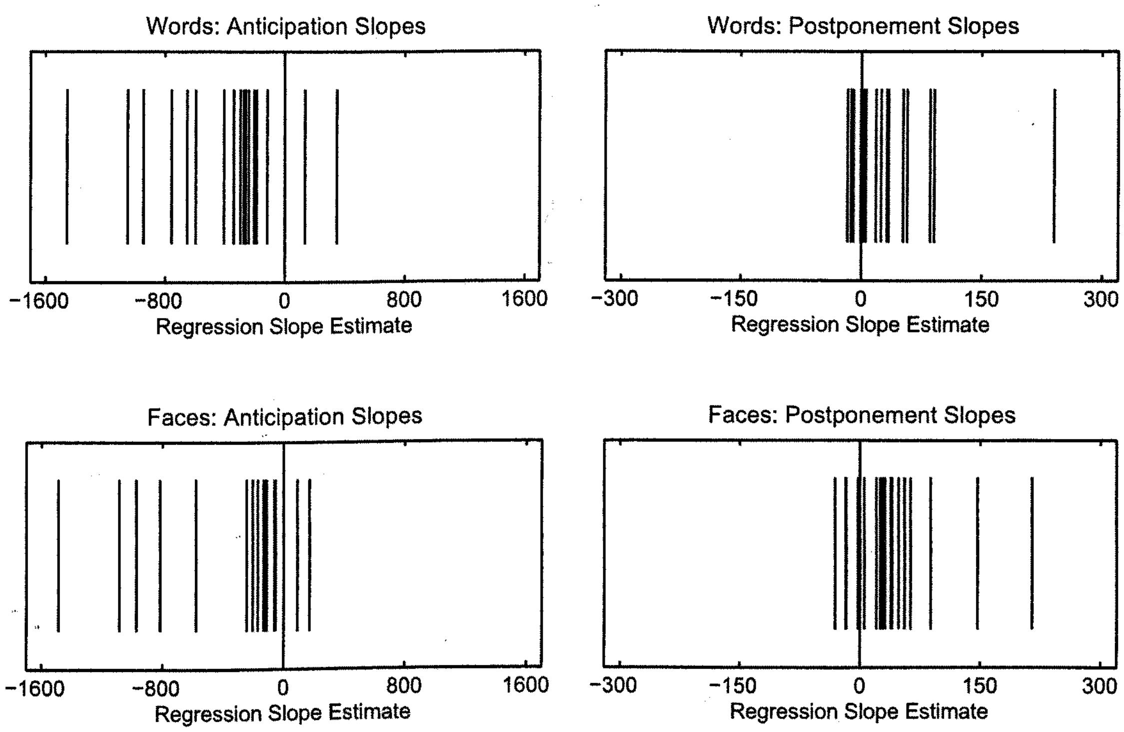

Figure 4-18 *Individual regression slope parameter estimates for the latency-displacement functions of Experiment 6. Each vertical line in a panel represents a regression slope estimate for a participant for a particular experimental condition. The left hand panels show slope estimates for anticipations, whilst the right hand panels show slope estimates for postponements. The top panels show the slope estimates for sequences of words, whilst the bottom panels show the slope estimates for sequences of faces. Note—the scales of the x-axes for the slope estimates for anticipations and postponements are different.*

markers, and suppression of recalled items. The generality of this empirical pattern is highlighted by the fact that it was insensitive to manipulations of sequence length (Experiments 1 & 2), articulatory suppression (Experiment 2 & 5), temporal grouping (Experiment 4), visual similarity (Experiment 5), the nature of the stimulus materials to-be-remembered (Experiment 6), and was found to hold for sequences composed of unique (or almost unique) items (Experiments 1, 2, 3, 4, & 6), as well as non-unique items (Experiment 5).

Quantitative model fitting

Although the LDFs observed in the preceding experiments are consistent with the error latency prediction of a model combining a primacy gradient with positional marking and response suppression, it is important to ascertain whether this models better account of the empirical pattern persists when the competitor models of Chapter 3 are fit to actual behavioural data. One concern is that the model predictions generated in that chapter to guide the interpretation of the data presented here were based upon parameter values that were arbitrarily chosen to accentuate the differences in their predicted LDFs, whilst still generating realistic and comparable serial position curves and transposition gradients. It remains possible therefore that the alternative models and mechanisms for representing serial order might predict the empirical pattern of the LDF when model parameters are estimated from behavioural data.

To establish whether this is the case three of the models from Chapter 3 were fit to the accuracy serial position curves and transposition gradients for the different sequence length conditions of Experiment 2. The LDFs predicted by the models were then compared with reference to the empirical data. The models that were fit were the position marking and response suppression (PM+RS) model, the primacy gradient and response suppression (PG+RS) model, and the primacy gradient, position marking, and response suppression (PG+PM+RS) model. The position marking (PM) model was excluded, because it predicts repetition errors that are absent in the empirical data owing to the deletion of selected items from the reconstruction array. The position marking, output interference, and response suppression (PM+OI+RS) model was also excluded for reasons delineated below.

Fitting procedure

The fitting procedure was the same as that employed to fit the Farrell and Lewandowsky (2004) data in Chapter 3. The models were fit to the data for individual participants (18 in total) using the simplex algorithm (Nelder & Mead, 1965), minimizing the summed root mean square deviation (RMSD) between the predicted and observed accuracy serial position curves and transposition gradients for the different sequence length conditions. The parameters for the competitive queuing response selection network, which includes the response threshold (T) and the excitatory (w^+) and inhibitory weights (w^-), were set to constant values of 0.8, 1.1 and -0.1, respectively. The parameters that were free to vary during the fitting process for the PM+RS model were the distinctiveness of the position markers (ϕ), and the standard deviation of noise (σ). The parameters that were free to vary for the PG+RS model were the steepness of the primacy gradient (γ), and the standard deviation of noise (σ). The parameters that were free to vary for the PG+PM+RS model were the steepness of the primacy gradient (γ), the distinctiveness of the position markers (ϕ), and the standard deviation of noise (σ).

The model parameters frozen during the fitting process included the weighting of activation of the position markers (λ), for the PM+RS and PG+PM+RS models, the weighting of activation of the primacy gradient ($a_1 = 1$), for the PG+RS and PG+PM+RS models, the attentional weight given to the primacy gradient and position markers ($\omega = .5$), for the PG+PM+RS model, and the amount of response suppression ($\alpha = 1$), which was universal to all models. Notice that the response suppression parameter was fixed to its maximum value. This was to ensure that the models were incapable of generating repetition errors, which as noted above were not possible in the data due to the serial reconstruction paradigm employed.

In summary, the number of free model parameters was two for the PG+RS and PM+RS models, whereas the PG+PM+RS model incorporated three free parameters. To increase the chances of finding the global minimum of the goodness-of-fit functions parameter estimates were obtained using multiple starting points for the search algorithm. The points were chosen by selecting two values for each free parameter and then factorially crossing these to create a grid of starting values. Each parameter vector explored by the search algorithm involved 10,000 simulation trials.

Model predictions

Before comparing the predictions of the models generated under their best fitting parameter values, consideration is first given to differences in their goodness-of-fits to the accuracy serial position curves and transposition gradients. The minimized RMSD values of the models for the different sequence length conditions can be scrutinized in Appendix 1. For four-item sequences (Table A1-2) the RMSDs averaged across fits to individual participants were 0.07 (PM+RS), 0.06

(PG+RS), and 0.04 (PG+PM+RS); for five-item sequences (Table A1-3) the corresponding RMSDs were 0.13 (PM+RS), 0.07 (PG+RS), and 0.07 (PG+PM+RS); whilst for six-item sequences (Table A1-4) the RMSDs were 0.15 (PM+RS), 0.08 (PG+RS), and 0.07 (PG+PM+RS). In brief, the PM+RS model consistently provided the worst fits, whilst the PG+PM+RS model generally provided the best fits. Nevertheless, the improvement in fit of this model relative to the PG+RS model was only small. Remember that these goodness-of-fits only take into consideration the descriptive accuracy of the models for the accuracy serial position curves and transposition gradients, which the simulations of Chapter 3 have already shown are generally insufficient to distinguish between the models.

The predictions of the three models for the accuracy and latency serial position curves and transposition gradients, averaged across fits to individual participants, are shown in Figure 4-19, alongside the target data from Experiment 2. Considering first the accuracy serial position curves (leftmost column of panels), it is evident that the PM+RS model predicts effects of primacy, but fails to capture the extent of recency observed for four-item sequences, whilst also failing to predict a recency effect whatsoever for five- and six-item sequences. The inability of this model to capture the recency effect is the reason for excluding the PM+OI+RS model from the fitting process, because adding output interference to the combination of position marking and response suppression will only exacerbate this problem. The PG+RS and PG+PM+RS models, by contrast, both capture the extent of primacy and recency observed in the data. From inspection of the corresponding transposition gradients (middle column of panels), it can be seen that the PM+RS model predicts a constant asymmetry reflected by more postponement than anticipation errors, which is at odds with the symmetrical error gradients for anticipations and postponements observed empirically. The PG+RS model fares better, but under-predicts the extent of transpositions for long positive displacements. In contrast, the PG+PM+RS model predicts symmetrical error gradients for anticipations and postponements in accordance with the data. Turning to the latency serial position curves (rightmost column of panels), these have had constants of 20, 30, and 40 iterations added to the average latency for the first output position for four-, five-, and six-item sequences, respectively, to increase graphical comparability with the data. It is apparent that unlike the latency

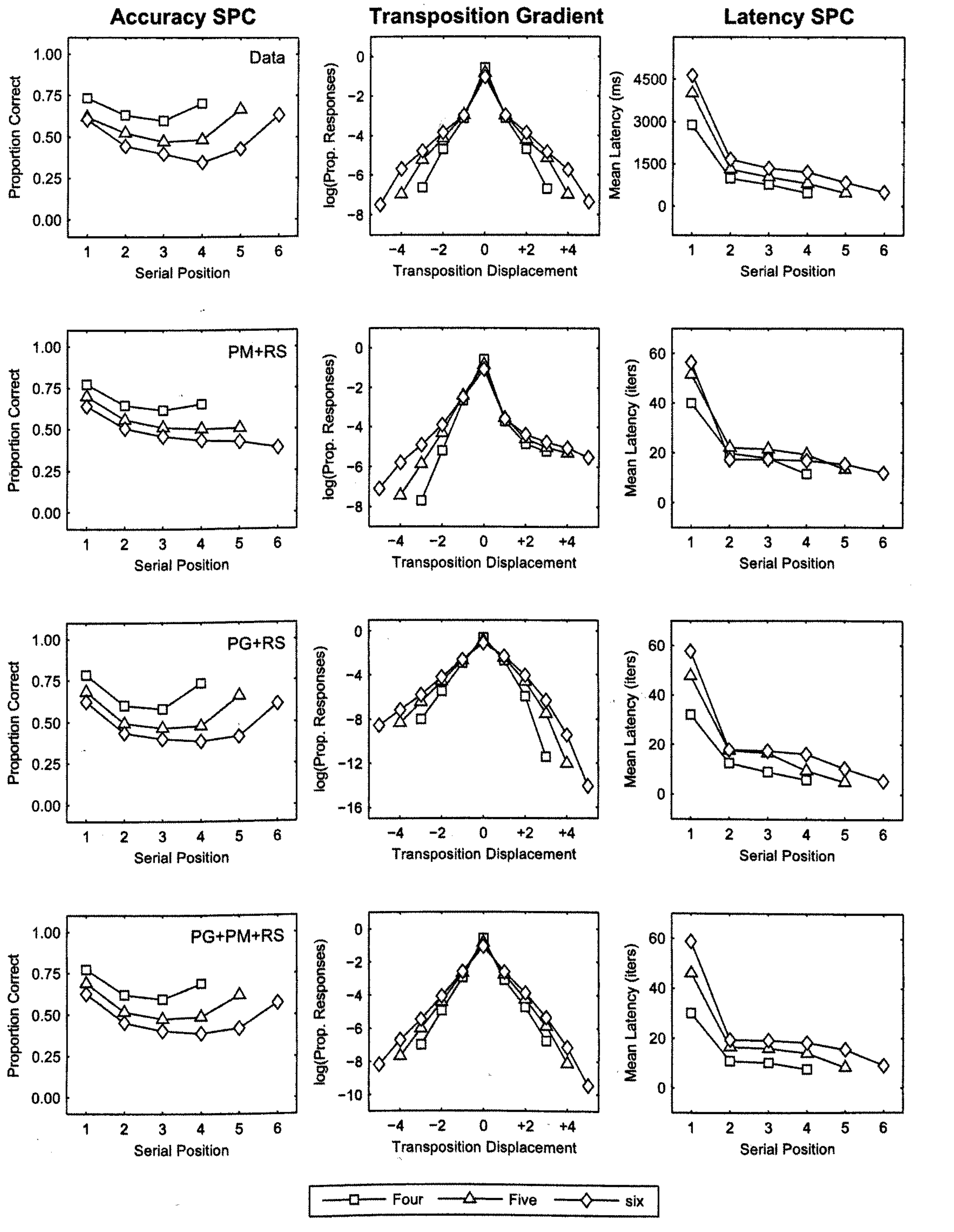

Figure 4-19 *Serial memory performance measures for Experiment 2 and fits of three models of serial order. Left-hand column of panels show accuracy serial position curves, middle column of panels show transposition gradients, and right-hand column of panels show latency serial position curves. The top row of panels show the data, the second row of panels show the predictions of the PM+RS model, the third row of panels show the predictions of the PG+RS model, whilst the fourth row of panels show the predictions for the PG+PM+RS model.*

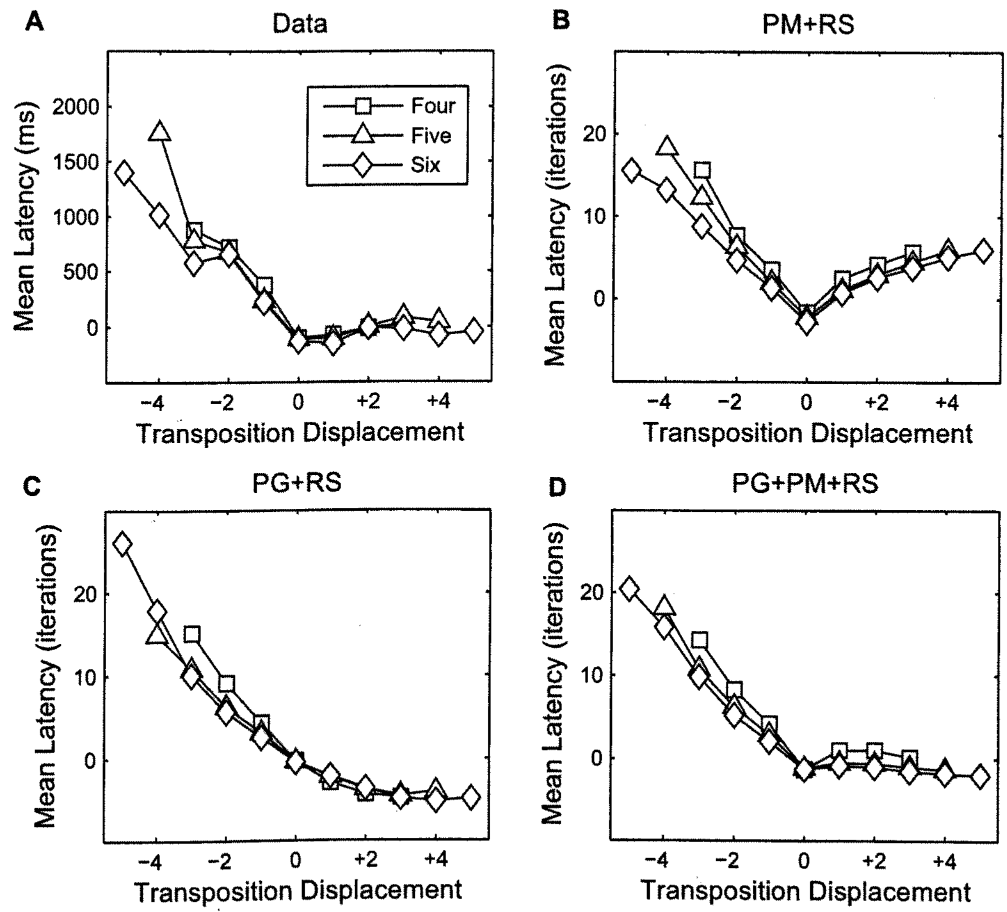

Figure 4-20 *Latency-displacement functions for Experiment 2 and fits of three models of serial order. Panels show the data (A), predictions of the PM+RS model (B), predictions of the PG+RS model (C), and predictions of the PG+PM+RS model (D).*

serial position curves generated by the models in Chapter 3, in which latency varied inversely with serial position accuracy (see Figure 3-1C), the models now predict that latency is a monotonically negative function of output position, consistent with the data.

Attention is now directed to the central predictions of interest. Figure 4-20 shows the LDFs, averaged across fits to individual participants, predicted by the models alongside the LDFs observed empirically for Experiment 2. Note that the effects of output position have been removed from the model predictions in exactly the same manner as described for the data. As can be seen from inspection of this figure, all the models predict negative anticipation slopes in accordance with the data. However, there is considerable heterogeneity in their predicted postponement slopes. The PM+RS model predicts steeply positive postponement slopes (Figure 4-20B), whereas the

PG+RS model predicts shallow negative postponement slopes (Figure 4-20C), both of which are contrary to the empirical pattern. In contrast, the PG+PM+RS model predicts flat postponement slopes (Figure 4-20D), which are most compatible with the data (Figure 4-20A). Note that although the postponement slopes observed empirically actually exhibit shallow positive rather than flat slopes, the disparity between the data and the predictions of this model are only negligible. Indeed, the close correspondence between the PG+PM+RS model and the data is striking, given that it was not fitted directly to the LDFs.

In summary, the main outcome of the current modelling exercise has been to show that when a sub-set of the models from Chapter 3 are fitted directly to representative behavioural data, only the combination of a primacy gradient, position marking, and response suppression can satisfactorily reproduce the qualitative pattern of the observed LDFs. This outcome suggests that both a primacy gradient and position marking must coexist in any adequate model of serial order in visual short-term memory. The same outcome compromises the viability of a model in which serial order is represented through positional marking or a primacy gradient alone.

General Discussion

The aim of this chapter was to identify a preferred combination of principles for representing serial order in visual short-term memory. In service of this goal, the error latency predictions of the five models and associated mechanisms for representing serial order presented in Chapter 3 were tested across six experiments, using a task requiring serial reconstruction of sequences of unfamiliar faces. These experiments consistently revealed that transposition latency is a negative function of transposition displacement, but with a shallow positive slope for postponement errors. The generality of this result is underscored by the fact that it was observed across manipulations of sequence length, articulatory suppression, temporal grouping, visual similarity, as well as a direct comparison of serial reconstruction for verbal and visual stimuli. This empirical outcome is consistent with the error latency prediction of a model in which serial order is represented through the combination of a primacy gradient, positional marking, and response suppression. It is incompatible with the error latency predictions of the four alternative mechanisms for representing serial order. Qualified support for the PG+PM+RS model was provided by the results of the quantitative model fitting exercise, which confirmed that it was the only model of those considered that could accommodate the pattern of the observed LDF when model parameters were estimated from the behavioural data. The outcomes of this chapter combined with the results of Chapter 3, implicate a role for a primacy gradient, positional marking, and response suppression in visual and verbal serial short-term memory, and support the hypothesis that at least some common principles represent serial order in the two domains.

Consideration is now given to potential limitations of the current experiments, as well as exceptions to the above claim that common principles represent serial order in the visual and verbal domains. The first limitation concerns the decision to remove items from the reconstruction array once they had been selected, which was the reconstruction method employed for the majority of the experiments. Under such conditions serial reconstruction proceeds under external guidance, thereby preventing the need for the operation of response suppression. Thus, although the empirical LDFs observed across the different experiments are consistent with the error latency prediction of the PG+PM+RS model, the above design choice means that for the majority of the experiments only a role for the operation of the primacy gradient and positional marking components of this model can be inferred from the data. The without-clearing condition of Experiment 3 was the only condition in which serial reconstruction proceeded without external guidance, thereby providing an opportunity to test the contribution (or lack thereof) of response suppression to the serial reconstruction of visual sequences. That the LDF for this experimental condition still conformed to the error latency prediction of the PG+PM+RS model provides empirical support for a contribution of response suppression. The operation of response suppression is further buttressed by the finding that erroneous repetitions were both rare and widely separated in the without-clearing condition, in accordance with the repetition constraint identified in verbal serial recall, which is considered to be a signature of the operation of response suppression. This empirical support for response suppression notwithstanding, the shortcoming of this evidence is that it was only observed in a single condition of one of the six experiments. Thus, the generality of the results of the without-clearing condition of Experiment 3 and their robustness to experimental manipulations of the kind examined here remains to be seen[4].

A second limitation concerns the absence of more robust effects of temporal grouping in Experiment 4. Grouping effects have been considered as the evidence *par excellence* for positional marking in short-term memory. This is because positional information can be organised on multiple dimensions, so that for example one dimension of ordering represents the positions of groups in sequence, whilst a second dimension represents the positions of items within groups. This multidimensional coding scheme has been shown to be both necessary and sufficient to explain the various effects of grouping, including the prevalence of interposition errors (see e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1999; Lewandowsky & Farrell, 2008). Such effects cannot be accounted for by primacy gradient models, which can only order information along a single dimension of activation strength. That the effects of grouping typically observed in verbal studies either did not materialise in Experiment 4 (interpositions), or were not fully expressed (elevation in serial position accuracy and change in response latencies) implies that positional information in visual short-term memory can only be organised along a single dimension of within-sequence position. This constitutes a fundamental discrepancy between verbal and visual serial memory, but it is one that must be considered in light of another distinction between the two domains. Unlike verbal stimuli, which can be rehearsed sub-vocally (Baddeley, 1986) and spatial stimuli, which can be rehearsed via shifts of covert spatial attention (Awh & Jonides, 2001) or eye-movements (Postle, Idzikowski, Della Sala, Logie, & Baddeley, 2006), there is no corresponding output mechanism for rehearsing visual, non-spatial stimuli. This discrepancy is important because there is evidence

[4] Given that response suppression was motivated in Chapter 1 as a core mechanism of serial order, one might reasonably ask why the without-clearing condition was not employed in all of the experiments reported here. Two points are noteworthy with respect this question. The first point is that the removal of items from the reconstruction array is the standard protocol in serial reconstruction experiments. The second point is that the experiments in this chapter pre-dated the writing of Chapter 1 and the modelling of Chapter 2, and at the time of running the experiments interest centred chiefly on the mechanisms that represent the relative priority of a sequence of visual items and so testing a role for response suppression was, at the time, an ancillary objective.

indicating that for visually presented verbal sequences effects of grouping are contingent upon rehearsal for their expression. For example, Frankish (1989) showed that effects of grouping increase as a function of the length of the temporal pause demarcating groups, with longer intervals presumably conferring greater opportunities for inter-group rehearsal. Moreover, a study by Hitch et al. (1996), as well as an as yet unpublished study by Hurlstone (2006), both demonstrated that the effects of temporal grouping are abolished when rehearsal is prevented by having participants engage in articulatory suppression during encoding. Thus, whilst the absence of robust effects of temporal grouping in visual serial memory might be interpreted as indicating fundamentally different forms of organisation in the verbal and visual domains, the above considerations suggest this discrepancy may have more to do with rehearsal related factors.

A further discrepancy between the experiments reported here and studies of verbal serial memory corresponds to the form of the latency serial position curve. In verbal studies, when responses for the first output position are ignored latency typically varies inversely with serial position accuracy (Farrell, 2008; Farrell & Lewandowsky, 2004; Maybery et al., 2002), whereas in the current experiments latency was a monotonically negative function of output position. How can this discrepancy be explained? The interpretation advanced here is that at each output position individuals scan the reconstruction array of remaining items before making a decision and committing to a response. Because the number of items in the reconstruction array decreases as output position increases (either through the overt clearing of items from the reconstruction array or by the suppression of selected items, or some analogous process) it takes progressively less time to scan the array and select a response. It is hypothesised that this visual search process is the consequence of the absence of an output mechanism for rehearsal and response generation in the visual, non-spatial domain, the consequence of which is that the representations of serial order will be impoverished and difficult to access without external guidance and support. Individuals therefore have to inspect the array of candidate items at each output position to probe the internal representations of order and recover the item representation with the largest relative priority. Notwithstanding the apparent difference between the latency serial position curves for verbal and visual serial memory, it is important to emphasise that this discrepancy does not compromise the interpretation of the LDFs presented here, nor those reported by Farrell and Lewandowsky (2004). This is because in both instances the effects of output position on latency were subtracted from the LDFs, both for models and data.

Before concluding this chapter one final issue merits comment, this concerns the extent to which the LDFs predicted by the models are representative of their general behaviour. The conclusion that the best account of the data presented here is provided by the PG+PM+RS model is based upon comparisons of the observed LDFs with the error latency predictions of the models presented in Chapter 3, combined with comparisons of the LDFs of Experiment 2 with the error latency predictions of the models after fitting to the same data. One limitation of these model predictions is that they are based upon a restricted range and number of local model parameter value combinations. That is, they only provide information about the behaviour of the models at a small number of points within a narrow range of their parameter space, but do not provide information about their predictions at other points in their parameter space. Thus, it is unclear whether the predictions of the models observed thus far are representative of their more general behaviour, and hence attributable to their core underlying principles. Exploring the wider predictive behaviour of models is important, because as noted in Chapter 2, a model can provide a good fit to data for reasons other than it being a good approximation of the underlying cognitive process. Specifically, a complex model can provide a good fit simply because it has the flexibility to accommodate a range of different data patterns. This problem is pertinent, because the model advocated here is indeed more complex than its rivals, incorporating five parameters, which is two more than both the PM+RS and PG+RS models. What is required then is an analysis of the how the predictions of the models (the PG+PM+RS model in particular) vary across a wide range of parameter setting combinations. Such an analysis is reported in Chapter 6. However, for present purposes two points are noteworthy. First, in fitting the models to the data from Experiment 2 (as well as the verbal serial recall data of Farrell and Lewandowsky in Chapter 3), the number of free model parameters was minimized so that the PG+PM+RS model contained only one free parameter more than its rivals. Second, the models were not fit directly to the latency data, but instead to the accuracy serial position curves and transposition gradients, constraints that contain no information about the temporal dynamics of recall. It is noteworthy therefore that the PG+PM+RS model predicted the empirical pattern of the LDF, despite being given no information about it during the fitting process. This strongly suggests that the predictions of the model follow from its core representational principles.

In conclusion, the novel contribution of the experiments and computational modelling presented in this chapter is to show that the similarities observed between verbal and visual short-term serial order memory can be traced to their reliance on at least some common principles for the representation of serial order. The analyses reported here, and in the previous chapter, indicate that in both domains serial order information is represented by a competitive queuing system equipped with a primacy gradient of activation, associations between items and positional markers, and suppression of recalled items. The next step is to establish whether an analysis of LDFs in spatial serial memory provides support for the same combination of representational principles. This issue is addressed in Chapter 5.

Chapter summary

This chapter tested the error latency predictions of five alternative mechanisms for representing serial order, using a task requiring serial reconstruction of sequences of unfamiliar faces. The results of six experiments, employing a range of experimental manipulations, consistently revealed that transposition latency is a negative function of transposition displacement, but with a shallow positive slope for postponement errors. This empirical outcome is consistent with the error latency prediction of a model in which serial order is represented by a primacy gradient of activation, associations between items and positional markers, and suppression of recalled items.

5

Transposition latencies in spatial serial memory

Abstract

This chapter presents three experiments that tested the error latency predictions of the five models and mechanisms for representing serial order presented in Chapter 3, using a task requiring serial reconstruction of sequences of spatial locations. These experiments consistently revealed latency-displacement functions that were most compatible with the error latency predictions of a representational mechanism combining a primacy gradient, positional marking, and response suppression. Quantitative fits of the models to representative data confirmed that only this mechanism could accommodate the pattern of the observed latency-displacement functions when model parameters where estimated from the behavioural data. Combined with the results of Farrell and Lewandowsky (2004), as well as those documented in Chapter 4, these findings are consistent with the hypothesis that verbal, visual, and spatial serial memories rely on common principles for representing serial order.

Introduction

The preceding chapter evaluated the latency-displacement functions (LDFs) underpinning serial reconstruction of sequences composed of visual non-spatial items, with reference to the error latency predictions of the five models and mechanisms for representing serial order presented in Chapter 3. The outcomes of that chapter indicate that serial order information in visual short-term memory is represented by a competitive queuing system equipped with a primacy gradient of activation, associations between items and positional markers, and suppression of recalled items. Nevertheless, any attempt to explain the representation of serial order in visuospatial short-term memory more generally must also engage the spatial component of the visuospatial short-term memory system.

The main vehicle for exploring the properties of spatial serial memory has been the Corsi-Blocks Task (Corsi, 1972), in which participants observe sequences of spatial locations, which they are subsequently required to imitate. Numerous studies utilising this task, or basic variants of it, have shown that spatial serial memory exhibits a wealth of characteristics that are similar to those observed for verbal serial memory. These include, but are not restricted to: effects of serial position on accuracy, including primacy and recency effects (Jones et al., 1995; Smyth & Scholey, 1996; Guerard & Tremblay, 2008), effects of serial position on latency (Parmentier, Elford, & Maybery, 2005; Parmentier et al., 2006), effects of sequence length (Jones et al., 1995), similar transposition error gradients (Parmentier et al., 2006; Smyth & Scholey, 1996), as well as similar distributions of item and order errors (Guerard & Tremblay, 2008). Although these findings are important in highlighting parallels between the functional characteristics of verbal and spatial serial memory, as noted previously such empirical constraints cannot be tied unambiguously to a specific combination of principles for representing serial order.

A major focus of recent research in this domain has been to determine the competencies involved in the encoding, rehearsal, and maintenance of spatial sequences in short-term memory. Accordingly, studies have examined the nature of the reference frame used to represent spatial sequences (Avons, 2007; Avons & Oswald, 2008), the importance of factors such as the length (Parmentier et al., 2006), complexity (Kemps, 1999; Parmentier et al., 2005), spatial organisation (De Lillo, 2004), and spatio-temporal organisation (Parmentier et al., 2006) of sequences and how this modulates recall performance, as well as the role of covert shifts of spatial attention (Awh & Jonides, 2001; Awh, Jonides, & Reuter-Lorenz, 1998) and eye movements (Guerard, Tremblay, & Saint-Aubin, in-press; Postle et al., 2006; Tremblay, Saint-Aubin, & Jalbert, 2006) in the rehearsal and maintenance of spatial information. Nevertheless, the more fundamental question of how serial order is represented in this domain has been overlooked. This neglect is noteworthy because this question has been the driving force of much of the theoretical developments in the verbal short-term memory literature (see e.g., Farrell, 2006; Farrell & Lewandowsky, 2003, 2004; Henson, 1996, 1998a, b, 1999; Henson et al., 1996).

The purpose of the current chapter is to bridge this evidential gap via an empirical investigation of the LDFs underpinning a spatial serial reconstruction task, in order to identify a preferred combination of principles for representing serial order in the spatial domain. The task employed to probe spatial serial memory in the experiments reported herein was a computerized version of the Corsi-Blocks Task that afforded the recording of response latencies. Note that there is no universally recognised, standardised instantiation of this task (see e.g., Berch, Krikorian, & Huha, 1998), so before moving on it is worth spelling out the key characteristics of the version employed here. The task utilized a fixed display of locations denoted by nine grey icons haphazardly arranged on a white background. Unlike most versions of this task, the first two experiments in this series (Experiments 7 & 8) employed a sequential spatial presentation array, as opposed to a simultaneous presentation array. To explain, in typical instantiations of the computerised Corsi-Blocks Task the locations are always simultaneously visible and the presentation order of a sequence is indicated by a transitory change in colour of a sub-set of locations. The sequential presentation format adopted here involved displaying locations in isolation by having each briefly appear and then disappear in succession. This format was chosen to increase the correspondence between the spatial task and the verbal immediate serial recall task, as well as the visual serial reconstruction task employed in Chapter 4, both of which use a sequential presentation format. To ensure that this modification did not engender any idiosyncratic performance characteristics, a final experiment in the series (Experiment 9) used the conventional simultaneous presentation array format.

A fundamental difference between the task employed here (and instantiations of the Corsi-Task more generally) and a popular alternative known as the Dots-Task (Jones et al., 1995) is that the latter task utilises unique spatial locations across trials, thereby engendering greater spatial uncertainty in the positioning of locations. It has been suggested that the principal advantage of this task over the Corsi-Blocks Task is that it is less vulnerable to verbal encoding strategies (Couture & Tremblay, 2006; Jones et al., 1995). The Dots-Task was not employed here for several reasons: First, the task is less widely employed in the research literature, having been associated with the work of a specific research group, that of Jones and his colleagues (e.g., Tremblay, Guerard, Parmentier, Nicholls, & Jones, 2006). Second, the majority of verbal studies utilize a closed experimental vocabulary, which is potentially of import for exploring the generality of certain experimental effects associated with a restricted item set, which includes the various effects of the manipulation of temporal grouping; a manipulation employed in two of the experiments reported here. Third, contrary to the claims of Jones et al. (1995), studies have shown that articulatory suppression has only a negligible (Meisser & Klauer, 1999), if any effect (Smyth & Scholey, 1992; Smyth, Pearson, & Pendleton, 1988) on spatial serial reconstruction accuracy, suggesting that verbal encoding strategies contribute little to performance.

A final noteworthy feature of the spatial serial reconstruction task employed here is that items could be selected on more than one occasion at recall, meaning that repetition errors were possible. In conventional serial reconstruction tasks repetition errors are not possible, because once an item is selected its shading changes and the item becomes refractory, thereby preventing it from being chosen again. In the current task, once an item was chosen its shading only changed transitorily to indicate the response had been registered, after which it was available for re-selection. Again, this feature of the task was incorporated to increase correspondence with the verbal immediate serial recall task and to provide an experimental scenario in which response suppression (a process hypothesised by four of the five models) might be recruited to support spatial serial reconstruction. The contribution (or lack thereof) of response suppression was determined (amongst other ways) by examining the incidence of erroneous repetition errors. If response suppression contributes to spatial serial reconstruction then the incidence of such errors should be lower than would be expected by chance.

The organization of the rest of this chapter is as follows: First, three experiments are reported examining the LDFs underpinning spatial serial reconstruction across manipulations of sequence length, temporal grouping (Experiments 7 & 9), and post-sequence interference (Experiment 8). Quantitative fits of the models from Chapter 3 to representative behavioural data are subsequently reported. To foreshadow the main conclusions, the empirical LDFs combined with the results of the quantitative model fitting exercises provide consistent support for a representational mechanism combining a primacy gradient of activation, positional marking, and response suppression.

Experiment 7

The first experiment examined the LDFs underpinning ungrouped and temporally grouped spatial sequences. As noted in previous chapters, differentiating a sequence into sub-groups by inserting extended pauses after every few items is a diagnostic manipulation that has been shown in verbal studies to exert a multiplicity of effects on serial recall performance relative to an ungrouped baseline. These effects include an elevation in recall accuracy (Frankish, 1985, 1989), a change in the shape of the accuracy serial position curve characterised by the emergence of primacy and recency effects within groups (Hitch et al., 1996), as well as a change in recall latencies typified by pronounced peaks in the latency serial position curve for group-initial positions (Farrell & Lelievre, 2009; Farrell & Lewandowsky, 2004; Maybery et al., 2002; Parmentier & Maybery, 2009). Grouping also engenders a change in recall errors reflected by a reduction in the tendency for items to transpose between groups (Henson, 1996; Ryan 1969a). However, one type of between-group error actually increases for grouped sequences (Farrell & Lelievre, 2009; Henson, 1999; Ng & Maybery, 2002; Ryan, 1969a); these interposition errors are between-group transpositions that preserve their within-group positions.

It is generally accepted that grouping constitutes a parade case for the important role of positional information in short-term memory (Henson, 1996, 1998a). To explain the various effects of grouping it has been necessary to assume that order information is encoded along multiple dimensions, with one encoding the positions of groups in sequence and another encoding the positions of items within-groups (Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998a; Lewandowsky & Farrell, 2008). That one dimension represents the positions of items within-groups, as opposed to some alternative dimension of ordering, is necessary to accommodate the pattern of between-group interposition errors that are a hallmark of grouped sequences. Grouping effects are beyond the scope of primacy gradient models, which can only order information along a single dimension.

Recently, Parmentier et al. (2006; Experiment 3) have shown effects of temporal grouping in spatial serial memory using nine-item sequences organised into three groups of three locations.

These effects included an elevation in recall accuracy, a change in recall latencies, and a reduction in transpositions between groups. This outcome supports the hypothesis that temporally grouped spatial sequences are encoded along multiple dimensions of ordering and by extension provides some support for a role for positional marking in spatial short-term memory. Nonetheless, there was one fundamental discrepancy between the effects of grouping reported by Parmentier et al. (2006) and those observed in verbal studies that the authors failed to draw attention to; this concerns the failure to detect an increase in interposition errors for grouped relative to ungrouped sequences. This discrepancy does not negate the claim that positional representations contribute to the encoding of temporally grouped spatial sequences, nor does it negate the claim that such information is organised along multiple dimensions of ordering. However, it does fail to support the hypothesis that one of those dimensions represents the positions of items within-groups.

The chief reason for employing the temporal grouping manipulation in the present experiment was to provide an experimental scenario in which the contribution of positional marking should be strong, and accordingly the slopes of the LDF for postponements would be expected to be steeply positive, consistent with the predictions of the models implementing positional marking in the absence of a primacy gradient. However, if the slopes of these functions are weakly positive or negative then this would point to a role for principles other than positional marking in the representation of serial order in spatial short-term memory (i.e., a primacy gradient). A further motivation for the grouping manipulation was to establish the generality of the results of Parmentier et al. (2006), most notably the failure to observe an elevation in interposition errors for temporally grouped spatial sequences.

Although many verbal studies employing a temporal grouping manipulation have used six-item sequences organised into two groups of three (Farrell, 2008; Farrell & Lewandowsky, 2004; Maybery et al., 2002; Parmentier & Maybery, 2009), this approach was not adopted here, because piloting revealed that performance was too high even for ungrouped sequences, yielding insufficient errors for a reliable transposition latency analysis. Instead, seven-item sequences were employed and grouped sequences were divided into a group of four items followed by a group of three items, a grouping pattern that has also been employed previously in verbal studies (Farrell & Lelievre, 2009; Henson, 1999).

One implication of employing an irregular grouping pattern is that interposition errors can fall into two categories: absolute and relative (Henson, 1999). Absolute interpositions are errors in which items migrate between groups, but preserve their position with respect to the beginning of the group. For example, the item at the first position of group one migrates to the first position of group two. For a 4-3 grouping pattern, absolute interpositions are reflected by transpositions with an absolute displacement value of four. Relative interpositions are errors in which items migrate between groups, but preserve their position relative to the beginning and end of the group. For a 4-3 grouping structure, relative interpositions are restricted to positional exchanges between items at the fourth and seventh positions. A change in the incidence of either of these two categories of errors would provide evidence for an effect of grouping on interposition rates.

Note that positional models, in addition to predicting an increase in the probability of interpositions for grouped sequences, also predict that the latencies accompanying these errors should be faster than the corresponding latencies associated with interpositions for ungrouped sequences. This is because the LDFs predicted by the models involving positional marking are inversely related to the transposition error gradients, as was demonstrated in Chapter 3. Thus, any peaks in the transposition gradient for grouped sequences at transposition displacements that are interpositions should materialize as troughs in the mean latencies at corresponding transposition displacements on the LDF.

Method

Participants

Twenty members of the campus community at the University of York took part in the experiment in exchange for course credit or payment of £4.

Stimuli & apparatus

The stimuli were sequences composed of seven seen spatial locations. The locations were nine grey icons (measuring 2.5cm X 2.5cm each) arranged haphazardly on a white background. The minimum and maximum distances between pairs of locations (measured from the centre of each square), were 3cm and 9cm, respectively. Stimulus presentation and response collection were controlled using software developed in-house using a Dell Optiplex (Intel Core 2 Duo, 2.13 GHz processor) PC equipped with a 19" monitor and a Razer Copperhead high precision mouse. The same apparatus was used for all subsequent experiments reported in this chapter.

Design & procedure

The experiment manipulated a single independent variable: Sequence-Type (Ungrouped / Grouped), which was a within-subjects factor. Participants always received the ungrouped sequences prior to the grouped sequences to reduce the likelihood of subjective grouping of the ungrouped sequences.

Participants were tested individually in a quiet room. They initiated each trial by selecting a 'begin trial' icon, located in the centre of the computer display, using the mouse-driven pointer. Following a 1000ms blank interval, seven locations selected at random by the computer programme controlling stimulus presentation were displayed individually on-screen in random order. For ungrouped sequences, locations were presented for 500ms each with a 250ms blank inter-stimulus interval separating locations. For grouped sequences, the presentation rates were the same except that the inter-stimulus interval separating presentation of the fourth and fifth locations was 1000ms creating the impression of two groups of locations, the first containing four locations and the second containing three.

Following the final location there was a 1000ms blank interval, after which all nine locations appeared simultaneously on-screen prompting participants to reconstruct the sequence in forward serial order using the mouse-driven pointer. Once an item was selected its colour changed transitorily from grey to green for 50ms to acknowledge that the response had been registered by the computer. Locations could be selected on more than one occasion, meaning that repetition errors were possible, as were intrusion errors, because the reconstruction array included two locations that had not been presented in the to be remembered sequence[1]. Following each response, a counter located in the bottom right hand corner of the screen incremented in value to inform the participant of the number of responses made so far. Participants were encouraged to guess if they were uncertain of the location for a given position, but if no location came to mind a 'don't know response' could be registered by selecting a question mark, which was located adjacent to the response counter. Once seven responses had been recorded by the computer the display cleared and the reconstruction time for the sequence was presented in the central screen position for 3000ms after which it was replaced by the 'begin trial' icon for the subsequent trial.

Participants were instructed to encode sequences visually, without deploying supplementary verbal or gestural encoding strategies. All participants reported compliance with these instructions. The experiment consisted of two practice and 80 experimental trials for each sequence type. Sessions lasted approximately 60 minutes.

Results

The data were analysed using a strict serial reconstruction scoring procedure: an item was scored as correct only if it was reported in its presentation serial position. The results are organised into five sections: (1) accuracy serial position curves, (2) transposition gradients, (3) repetition errors, (4) latency serial position curves, and (5) latency-displacement functions.

Accuracy serial position curves

The accuracy serial position curves for ungrouped and grouped sequences can be inspected in Figure 5-1A. As can be seen from inspection of this figure, grouping caused an elevation in recall performance and a scalloping of the serial position curve. Statistical confirmation of these observations was obtained by means of a 2 (Sequence-Type) X 7 (Serial Position) ANOVA, which revealed significant main effects of both Sequence-Type, $F(1, 19) = 8.451$, $MSE = .258$, $p < .01$,

[1] These were the locations remaining (from the possible nine) once the seven locations that formed the to-be-remembered sequence had been randomly chosen.

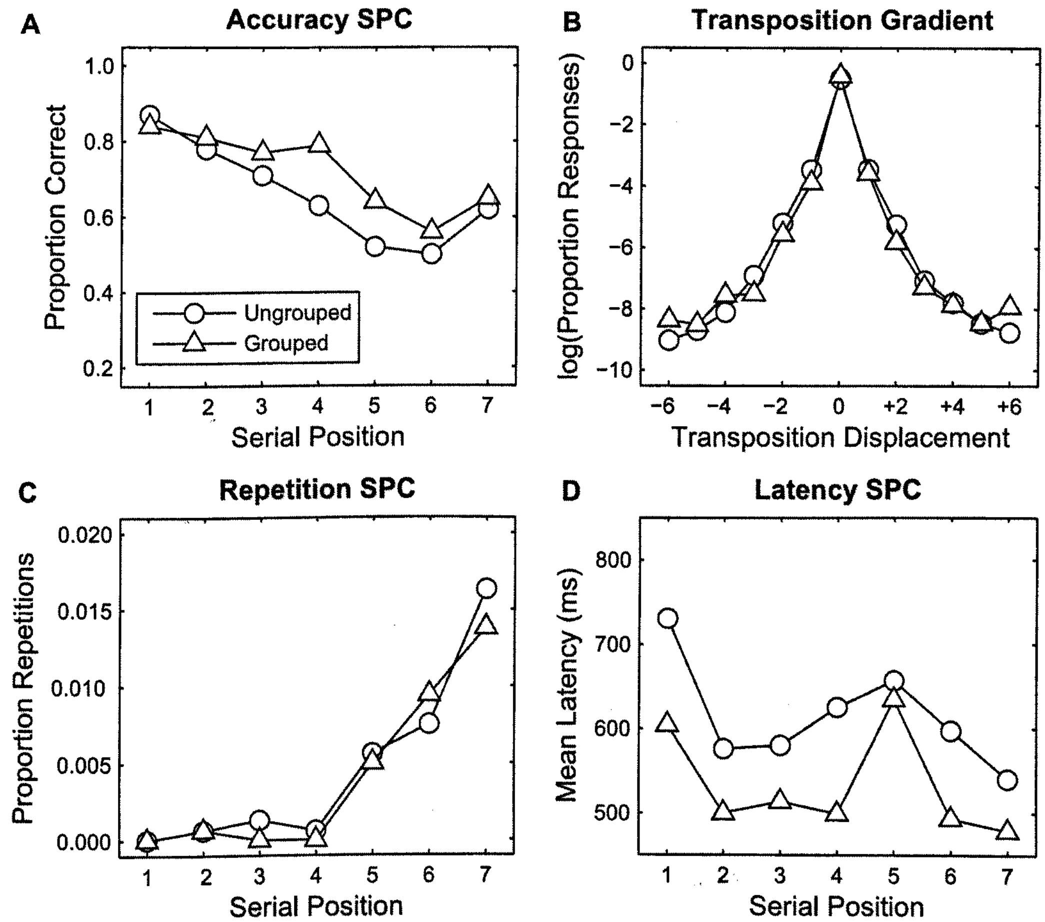

Figure 5-1 *Serial memory performance measures for Experiment 7. Panels show accuracy serial position curves (A), transposition gradients (B), repetition error serial position curves (C), and latency serial position curves (D).*

and Serial Position, $F(6, 114) = 59.814$, $MSE = 1.547$, $p < .001$, as well as a significant interaction between the two factors, $F(6, 114) = 11.262$, $MSE = .073$, $p < .001$.

Transposition gradients

The transposition gradients are illustrated in Figure 5-1B and exhibit three major characteristics. First, the gradients peak at displacement zero, reflecting that the majority of responses are correct. Second, the distribution of transpositions conforms to the locality constraint: the proportion of transpositions is greatest for displacements with an absolute value of one, and subsequently decreases gradually as the absolute value of the displacement increases (one unique exception being the slight elevation for transpositions with an absolute displacement value of six

Sequence-Type	Within Groups	Between Groups	
		Interpositions	Other
Ungrouped	.71	.05	.24
	(.07)	(.03)	(.07)
Grouped	.87	.04	.09
	(.09)	(.03)	(.06)

Table 5-1 *Proportion of transpositions within and between groups for Experiment 7. Values in brackets represent the standard deviation of the mean.*

for grouped sequences). Third, the transposition gradients are symmetrical, indicating that the incidence of transpositions for each absolute displacement is generally unaffected by whether the error is an anticipation or postponement. Note that if grouping had resulted in a tendency for items in different groups to exchange their absolute and/or relative within-group positions then peaks in the transposition gradient for displacements with an absolute value of three and four would have been expected. However, with the exception of a slight elevation in responses for -4 displacements, there is little indication that grouping engendered an increase in such interposition errors.

To scrutinize the error patterns further, transpositions were classified as occurring within or between groups, with the latter error-type being further subdivided into interpositions and other between group order errors. Because the incidence of interposition errors was low, absolute and relative interpositions were pooled together. The proportion of errors of each type for ungrouped and grouped sequences can be observed in Table 5-1. It is apparent that grouping increased the incidence of within-group errors, but decreased the incidence of between-group errors, without however engendering an increase in interpositions. The incidence of each error-type was compared between ungrouped and grouped sequences using t-tests performed on the log-odds transformed error proportions, a transformation which makes some allowance for potential floor and ceiling effects. These comparisons revealed a greater incidence of transpositions within groups for grouped sequences, $t(19) = 6.943$, $p < .001$, and a greater incidence of transpositions between-groups that were not interpositions for ungrouped sequences, $t(19) = 7.399$, $p < .001$. The incidence of interpositions did not differ reliably between the two sequence conditions, $t(19) = 2.044$, $p = .220$.[2]

Repetition errors

Repetition errors were extremely rare, accounting for less than 1% of all responses for both ungrouped and grouped sequences, which is considerably below that expected by chance[3]. The average lag between the two instances of the repeat was 4.4 positions for ungrouped sequences and 4.76 positions for grouped sequences. Figure 5-1C shows the proportions of repetitions as a function of output position. Inspection of this figure reveals that the probability of repetitions increased across serial positions and that the incidence of repetitions at each position was generally unaffected by the grouping manipulation. These data were analysed by means of a 2 (Sequence-Type) X 7 (Serial Position) ANOVA performed on the log-odds transformed error proportions. This revealed a significant main effect of Serial Position, $F(1, 119) = 13.206$, $MSE = 3.986$, $p < .001$, however neither the main effect of Sequence-Type, $F(1, 19) = .372$, $MSE = .065$, $p = .549$, nor the Sequence-Type X Serial Position interaction, $F(6, 114) = .393$, $MSE = .111$, $p = .757$, reached significance.

Latency serial position curves

The serial position curves associated with correct responses are shown in Figure 5-1D. Immediately apparent is that the serial position curves for ungrouped and grouped sequences are characterised by a long initial latency, as well as a second peak occurring at the fifth serial position,

[2] One concern here is that the failure to detect an increase in interpositions for grouped sequences might be the consequence of the comparison of frequencies of interpositions expressed as a proportion of order errors, rather than responses as a whole. When expressed as a proportion of all responses interpositions had an occurrence frequency of 1.4% for ungrouped sequences, compared to 1% for grouped sequences, a significant difference between conditions, $t(19) = 2.0768$, $p = .05$. Thus, interpositions were actually more frequent for ungrouped than for grouped sequences.

[3] Following the precedent set in Chapter 3 repetition errors (both occurrences of the repeat) were retained in the transposition gradients, but excluded from the LDFs for all of the experiments reported in this chapter as well as the model simulations presented later.

which is considerably more pronounced for grouped sequences. When the first position is ignored the serial position curve for ungrouped sequences exhibits an inverted U shaped profile, whereas the serial position curve for grouped sequences exhibits a shallow negative trend (notwithstanding the peak in latency at position five). These data were analysed by means of a 2 (Sequence-Type) X 7 (Serial Position) ANOVA, which revealed significant main effects of both Sequence-Type: $F(1, 19) = 33.489$, $MSE = 486999.294$, $p < .001$, and Serial Position: $F(6, 114) = 12.452$, $MSE = 402301.548$, $p < .001$, as well as a significant interaction between the two factors, $F(6, 114) = 3.781$, $MSE = 24711.981$, $p < .05$.

Latency-displacement functions

Turning now to the LDFs, Figure 5-2 shows the mean latencies associated with transpositions as a function of transposition displacement.[4] As before, the effects of output position have been removed from these data by subtracting each participant's mean latency for each output position from the individual latencies at corresponding output positions. It is apparent from inspection of this figure that the slopes of the LDFs for anticipations are negative, although the continuity of the functions is interrupted by accelerated latencies for -5 displacements for ungrouped sequences, and -6 displacements for grouped sequences. Also apparent is that the slope of the function for anticipations is shallower for grouped than for ungrouped sequences. In contrast, the slopes of the LDFs for postponements appear flat, albeit with a slight acceleration in the latencies for +5 and +6 displacements.

[4] As per the experiments reported in Chapter 4 exploratory analyses were performed initially to examine the sensitivity of the LDFs to potential outliers. These analyses once again confirmed that the qualitative form of the LDFs (the sign and steepness of the anticipation and postponement slopes) was generally unaffected by whether all latencies were included, or only responses within the range of 2.5, or 3 standard deviations from the mean. The qualitative form of the LDFs was also similar when the mean of the median latencies was used as the dependent measure instead of the mean latency. Given the insensitivity of the qualitative form of the LDF to these different approaches to dealing with response time outliers all responses were retained for the transposition latency analysis and the mean latency was once again employed as the dependent measure.

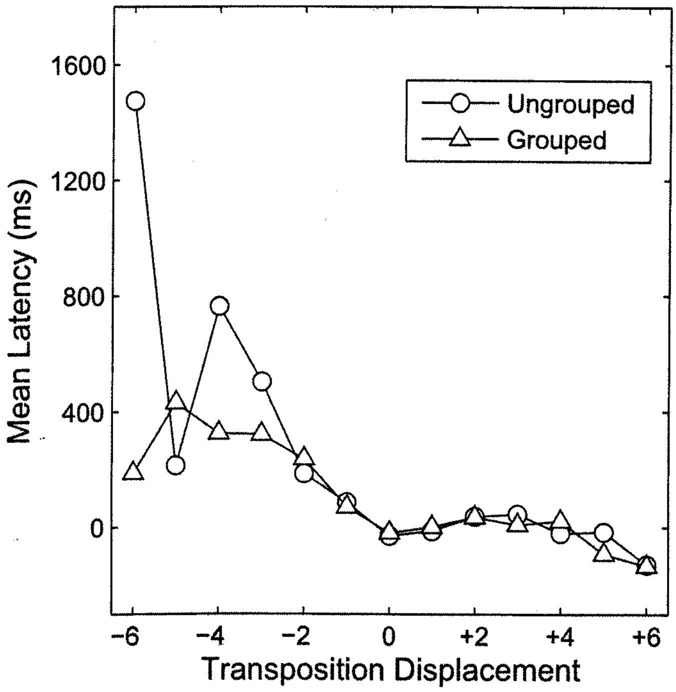

Figure 5-2 *Latency-displacement functions for Experiment 7.*

The LDFs were analysed using the same two-stage procedure employed in Chapter 4. In the first stage, regression analyses were performed examining the relationship between transposition latency and transposition displacement for each individual participant. One set of analyses examined the relationship between transposition latency and transposition displacements that were anticipations (displacements in the range -6 to 0), whilst a second set of analyses examined the relationship between transposition latency and transposition displacements that were postponements (displacements in the range 0 to +6).

In the second stage, the regression slope parameter estimates resulting from these analyses were pooled together and evaluated using one-sample t-tests to determine whether they deviated reliably from zero. The mean regression parameter estimates for anticipation and postponement slopes for ungrouped and grouped sequences can be scrutinized in Table 5-2. The mean regression parameter estimates for anticipation slopes were negative and deviated reliably from zero: $t(19) = -3.676$, $p < .01$, for ungrouped sequences, and $t(19) = -3.091$, $p < .01$, for grouped sequences. In contrast, the mean regression parameter estimates for postponement slopes did not deviate reliably from zero: $t(19) = -.480$, $p = .636$, for ungrouped sequences, and $t(19) = .246$, $p = .808$, for grouped sequences.

Parameter	Estimate	*SE*	*t*	*p*
Ungrouped				
Anticipation	-169.22	46.04	-3.676	.00
Postponement	-4.07	8.48	-.480	.64
Grouped				
Anticipation	-120.66	39.03	-3.091	.01
Postponement	1.79	7.30	.246	.81

Table 5-2 *Mean regression slope parameter estimates for the latency-displacement functions of Experiment 7.*

To give some indication of the variability in LDF slopes, Figure 5-3 shows the anticipation and postponement slope estimates for individual participants for ungrouped and grouped sequences. It is apparent from inspection of this figure that for both sequence-types the majority of participants contributed steep negative slopes for anticipations, whereas for postponements there was an approximately equal distribution of negative and positive postponement slopes, and these slopes were exclusively shallow slopes. This confirms that the empirical pattern of the averaged LDFs is not the consequence of a small number of participants exerting undue influence on the data.

Discussion

Although the central findings of interest are the LDFs, I begin first by considering the impact of the temporal grouping manipulation on the other serial memory performance measures. Consistent with the spatial memory study of Parmentier et al. (2006), as well as numerous verbal studies (Farrell & Lewandowsky, 2004; Farrell & Lelievre, 2009; Henson, 1999; Hitch et al., 1996; Maybery et al., 2002), temporal grouping engendered an elevation in serial reconstruction performance, a change in the shape of the accuracy and latency serial position curves, as well as a reduction in the tendency for items to transpose between groups. These findings support the hypothesis that positional information contributes to the encoding of spatial sequences, because such effects cannot be explained on the basis of a simple primacy gradient of activations. On the contrary, the presence of grouping effects implies that order information is organised along multiple dimensions, at least one of which represents the positions of groups in sequence. On the

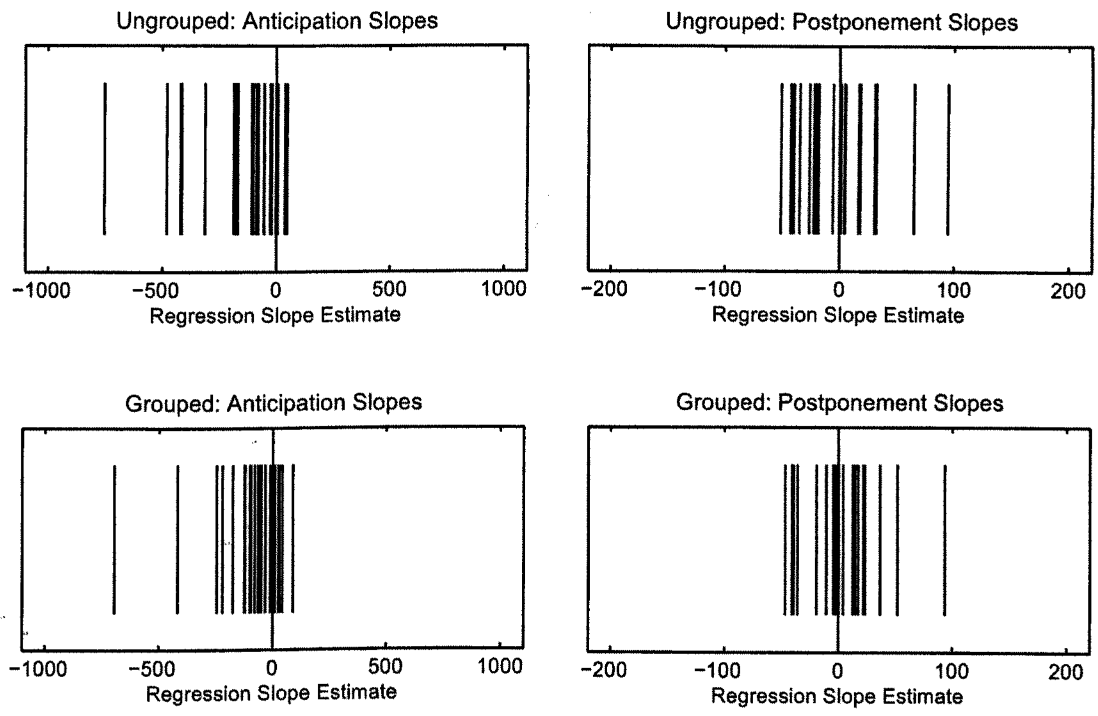

Figure 5-3 *Individual regression slope parameter estimates for the latency-displacement functions of Experiment 7. Each vertical line in a panel represents a regression slope estimate for a participant for a particular experimental condition. The left hand panels show slope estimates for anticipations, whilst the right hand panels show slope estimates for postponements. The top panels show the slope estimates for ungrouped sequences, whilst the bottom panels show the slope estimates for grouped sequences. Note—the scales of the x-axes for the slope estimates for anticipations and postponements are different.*

basis of the ubiquitous tendency for items to transpose between groups, but preserve their within-group positions, it has been hypothesised that for temporally grouped verbal sequences a second dimension represents the positions of items within-groups (Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998a; Lewandowsky & Farrell, 2008). However, consistent with the results of Parmentier et al. (2006), the present experiment failed to detect an elevation in the incidence of interposition errors for temporally grouped spatial sequences. The absence of such errors appears to imply that items in grouped spatial sequences are not coded for their positions within-groups, but perhaps instead in terms of their positions within the sequence overall. This possibility, as well as others, is considered further in the general discussion.

Attention is now directed to the empirical findings of chief interest. The LDFs for ungrouped and grouped spatial sequences exhibited an overall negative trend, but with a reduction in the slope of the functions for postponements compared to anticipations. This empirical pattern is generally consistent with that observed in the verbal serial recall experiments of Farrell and Lewandowsky (2004), as well as the visual serial reconstruction experiments of Chapter 4. However, one minor discrepancy is that the LDFs for verbal and visual stimuli exhibited a statistically significant shallow positive postponement slope, whereas the current LDFs for spatial serial reconstruction exhibited statistically flat postponement slopes. This discrepancy notwithstanding, the LDFs observed in the current experiment, like those witnessed with verbal and visual stimuli, are most compatible with the error latency prediction of a representational mechanism combining a primacy gradient of activation, positional marking, and response suppression (see Figure 3-2B of Chapter 3 for a reminder of the predictions of the different representational mechanisms). Note that although the qualitative LDF predicted by this model at the outset of Chapter 3 exhibits a shallow positive postponement slope, the difference between the prediction of the model and the current data is negligible. Moreover, the primacy gradient, position marking, and response suppression model does sometimes predict flat postponement slopes, as exemplified by the LDFs predicted by this model after fitting to the visual data of Experiment 2 in Chapter 4 (see Figure 4-20).

Thus, although the effects of grouping reported above point to a critical role for positional marking in the representation of spatial sequences, the LDFs also point to a necessary role for a primacy gradient and response suppression, since combinations of position marking that do not invoke a primacy gradient universally predict steep positive postponement slopes that were not observed empirically. Indeed the insensitivity of the postponement slope of the LDF to the grouping manipulation – which was also observed in the experiments of Farrell and Lewandowsky (2004) – suggests that even temporally grouped spatial sequences recruit a primacy gradient. Qualified support for the complementary role of response suppression was provided by the scarcity of erroneous repetition errors, which occurred at a level below that expected by chance for both ungrouped and grouped sequences. One final noteworthy feature of the LDFs is the failure to observe faster latencies for interposition errors for grouped relative to ungrouped sequences. This should not be considered surprising given the failure of the current experiment to detect an increase in the probability of interpositions for grouped sequences. Nevertheless, it is noteworthy that even Farrell and Lewandowsky (2004) failed to observe faster latencies for interpositions in the grouped condition of their Experiment 3, which nevertheless yielded an elevation in the probability of such errors.

Experiment 8

One limitation of Experiment 7 was that insufficient errors were observed for certain transposition displacements. This is reflected by the percentage of missing data cells for the LDF analyses, which was 36% for ungrouped sequences and 41% for grouped sequences. The majority of these missing data cells represented transpositions with absolute displacement values in the range of four and six. It follows that a more reliable assessment of the LDFs can be obtained by increasing the number of observations for the transposition latency analysis. The aim of Experiment 8 was to increase the frequency of transposition errors by incorporating an attention demanding secondary interference task. The specific secondary task employed involved making parity judgements for two centrally presented digits that followed presentation of each seven-item spatial sequence.

Method

Participants

Twenty members of the campus community at the University of York took part in the experiment in exchange for course credit or payment of £5.

Design & procedure

The experiment manipulated a single independent variable: Sequence-Type (Control / Interference), which was a within-subjects factor. Half the participants received the control-sequences followed by the interference-sequences, while the remaining half of participants received the sequences in the converse order.

The procedure was identical to that for the ungrouped sequence condition of Experiment 1 with two exceptions. First, the inter-stimulus interval was increased from 250ms to 500ms. Second, for the interference-sequences, the 1000ms interval after presentation of the final location was followed by two digits presented individually in the central screen position. Participants were required to make parity judgements for each digit, pressing the left mouse button for odd digits and the right mouse button for even digits. Participants were told that it was important that they classified each digit correctly. Following the parity judgement task the locations appeared on screen simultaneously, prompting forward serial order reconstruction of the sequence.

Participants were instructed to encode sequences visually, without deploying supplementary verbal or gestural encoding strategies. All participants reported adherence to these instructions. The experiment consisted of two practice and 80 experimental trials for each sequence-type. Sessions lasted approximately 75 minutes.

Results

The data for a single participant were excluded from all following analyses because this individual's mean accuracy scores for the control and interference sequences (97% and 94%, respectively) were almost at ceiling yielding insufficient errors for the transposition latency analysis.

Accuracy serial position curves

The accuracy serial position curves for control and interference sequences can be inspected in Figure 5-4A. Immediately apparent is that performance is lower for interference sequences than for control sequences, confirming that the secondary task manipulation had the desired effect of lowering performance. The mean proportion of correct responses, averaged across serial positions, for control and interference sequences was .59 and .46, respectively. It is noteworthy that performance for control sequences was lower than in Experiment 7, where the mean proportion of correct responses was .66. The data were subjected to a 2 (Sequence-Type) X 7 (Serial Position) ANOVA, which revealed significant main effects of Sequence-Type, $F(1, 19) = 36.668$, $MSE = 1.231$, $p < .001$, and Serial Position, $F(6, 114) = 58.885$, $MSE = 1.528$, $p < .001$, however, the

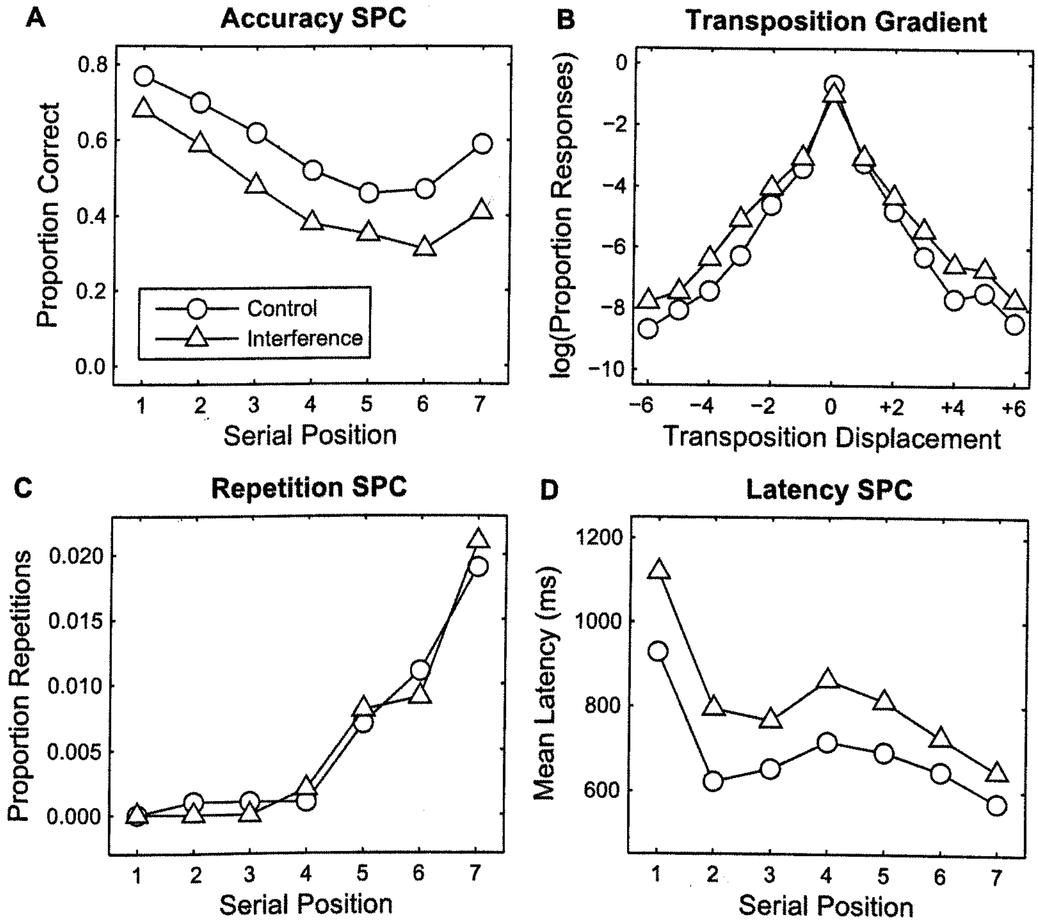

Figure 5-4 *Serial memory performance measures for Experiment 8. Panels show accuracy serial position curves (A), transposition gradients (B), repetition error serial position curves (C), and latency serial position curves (D).*

interaction between the two factors marginally fell short of significance, $F(6, 114) = 2.540$, $MSE = .017$ $p = .055$.

Transposition gradients

The transposition gradients underlying the serial position curves are portrayed in Figure 5-4B. They exhibit a steep peak at displacement value zero, a locality constraint on the distribution of transpositions, and symmetrical error gradients for anticipations and postponements. Corroborating the serial position analyses, the proportions of anticipations and postponements is greater for interference than for control sequences.

Repetition errors

As for Experiment 7 repetitions were scarce, accounting for approximately 1% of all responses for both control and interference sequences, which is below that expected by chance. The average lag between the two instances of the repeat was 4.68 positions for control sequences and 4.84 positions for interference sequences. The distribution of repetitions across output positions can be observed in Figure 5-4C. It is apparent from inspection of this figure that the probability of repetitions increased across positions, but the incidence of repetitions at each position was insensitive to the secondary-task manipulation. Statistical confirmation of these trends was obtained via a 2 (Sequence-Type) X 7 (Serial Position) ANOVA performed on the log-odds transformed error proportions, which revealed a reliable main effect of Serial Position, $F(6, 114) = 22.856$, $MSE = 32.148$, $p < .001$, however neither the main effect of Sequence-Type, $F(1, 19) = .160$, $MSE = .050$, $p = .693$, nor the Sequence-Type X Serial Position interaction, $F(6, 114) = .654$, $MSE = .305$, $p = .604$, reached significance.

Latency serial position curves

Figure 5-4D shows the latency serial position curves associated with correct responses. As can be seen from inspection of this figure, the latencies were longest for the first position after which they followed an inverted U shaped profile. Also apparent is that the latencies were longer at all serial positions for interference than for control sequences. These data were analysed using a 2 (Sequence-Type) X 7 (Serial Position) ANOVA, which revealed reliable main effects of both Sequence-Type, $F(1, 19) = 12.606$, $MSE = 1143394.652$, $p < .01$, and Serial Position, $F(6, 114) = 8.326$, $MSE = 2689942.336$, $p < .01$, but the interaction between the two factors failed to reach significance, $F(6, 114) = 1.066$, $MSE = 18949.118$, $p = .387$.

Latency-displacement functions

The LDFs with the effects of output position subtracted are shown in Figure 5-5. As can be seen from inspection of this figure, the slopes of the functions for anticipations are negative, but with a shallower slope for interference than control sequences largely owing to a levelling off of

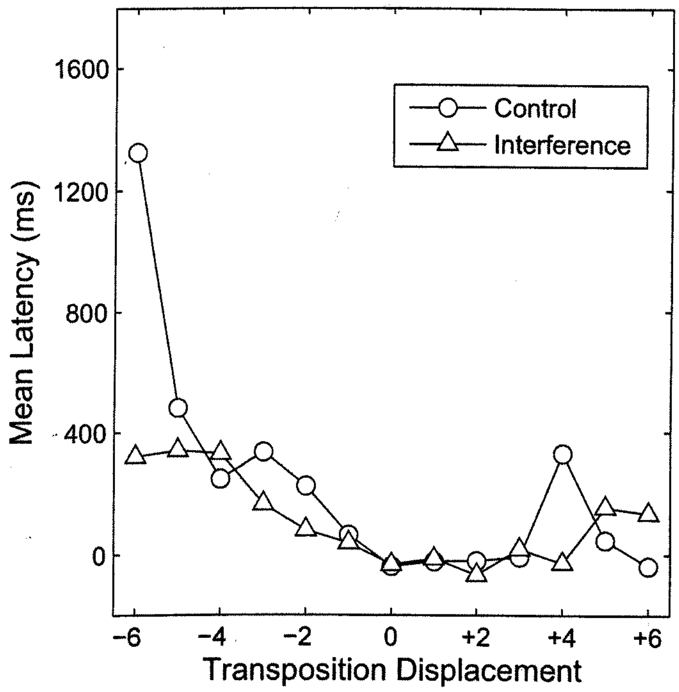

Figure 5-5 *Latency-displacement functions for Experiment 8.*

Parameter	Estimate	*SE*	*t*	*p*
Control				
Anticipation	-152.06	44.37	-3.427	.00
Postponement	57.02	40.80	1.398	.18
Interference				
Anticipation	-89.51	30.29	-2.955	.01
Postponement	30.47	16.92	1.801	.09

Table 5-3 *Mean regression slope parameter estimates for the latency-displacement functions of Experiment 8.*

the mean latencies for -4, -5, and -6 transposition displacements. In contrast, with some unique deviations (e.g., the peak at displacement +4 for control sequences and the peaks at displacements +5 and +6 for interference sequences) the slopes of the functions for postponements appear flat.

As in Experiment 7, the data were analysed by performing regression analyses on the LDFs for each individual participant, which examined the relationship between transposition latency and transposition displacements that were anticipations (displacements in the range –6 to 0) and

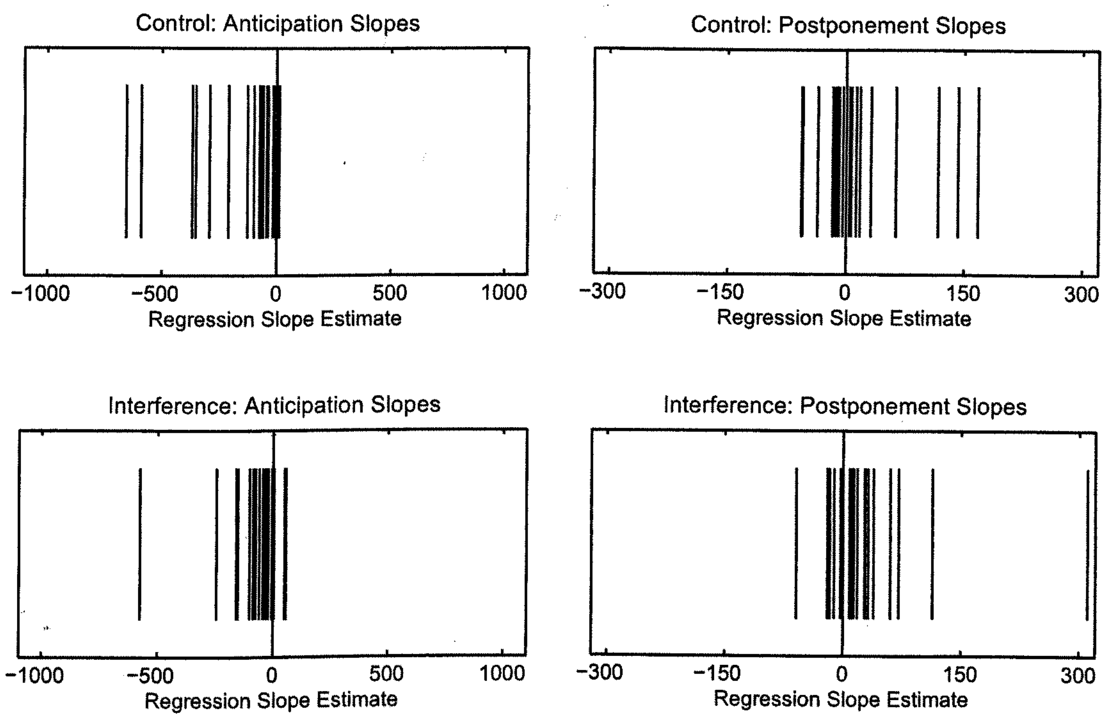

Figure 5-6 *Individual regression slope parameter estimates for the latency-displacement functions of Experiment 8. Each vertical line in a panel represents a regression slope estimate for a participant for a particular experimental condition. The left hand panels show slope estimates for anticipations, whilst the right hand panels show slope estimates for postponements. The top panels show the slope estimates for control sequences, whilst the bottom panels show the slope estimates for interference sequences. Note—the scales of the x-axes for the slope estimates for anticipations and postponements are different.*

postponements (displacements in the range 0 to +6) separately. The regression slope parameter estimates were subsequently pooled together and evaluated using one-sample t-tests to determine whether they deviated reliably from zero. The mean regression parameter estimates for anticipation and postponement slopes for each sequence-type can be scrutinized in Table 5-3. The mean slope estimates for anticipations were negative and deviated reliably from zero: $t(19) = -3.427, p < .01$, for control sequences, and $t(19) = -2.955, p < .01$, for interference sequences. In contrast, the mean slope estimates for postponements were weakly positive, but did not differ reliably from zero: $t(19) = 1.398, p = .178$, for control sequences, and $t(19) = 1.801, p = .088$, for interference sequences.

To provide a sense of the variability in the LDF slopes, Figure 5-6 shows the individual regression slope parameter estimates for anticipations and postponements. It can be seen that for both control and interference sequences the majority of participants contributed steep negative slope estimates for anticipations, whereas for postponements there was an approximately equal distribution of negative and positive postponement slopes, and with one unique exception (a single individual in the interference condition who contributed a steep positive slope) these slopes were exclusively shallow slopes. This indicates that the empirical patterns of the aggregate LDFs are an accurate reflection of the individual LDFs from which they are composed.

Discussion

The requirement to make parity judgements to two digits presented at the end of spatial sequences significantly lowered the accuracy of recall and led to a corresponding increase in the frequency of transpositions, as was desired. Indeed, even the incidence of transpositions for control sequences was increased relative to corresponding sequences in Experiment 7, presumably due to the counterbalancing of conditions employed in the current experiment. The resulting LDFs are generally comparable to those witnessed previously, except that the slopes of the functions for postponements in the current experiment although statistically flat once more, nevertheless exhibited a positive trend that almost reached conventional significance levels. The empirically observed LDFs are once again most compatible with the error latency prediction of a representational mechanism combining a primacy gradient of activation, positional marking, and response suppression. Additional evidence for a contribution of response suppression was once more provided by the scarcity of erroneous repetitions in participants' recalls, which occurred at a level below chance expectancy.

Experiment 9

The preceding experiments have highlighted the generality of LDFs characterised by negative anticipation slopes and flat postponement slopes in the serial reconstruction of spatial sequences, by demonstrating the insensitivity of this empirical trend to manipulations of temporal grouping and post-sequence interference. The aim of Experiment 9 was to further examine the generality of this empirical pattern in three ways. First, longer sequences of nine spatial locations were employed to further lower recall accuracy, thereby increasing the frequency of transposition errors without resorting to a secondary interference task. This also enabled an assessment of whether the relationship between latency and transposition displacement hitherto observed holds when transposition errors could span a larger number of displacements. Second, a temporal grouping manipulation was incorporated once more, but this time employing the presentation format of grouping into threes, which has been the most widely employed grouping pattern in verbal serial recall studies (Farrell, 2008; Farrell & Lewandowsky, 2004; Frankish, 1989; Hitch et al., 1996; Maybery et al., 2002; Parmentier & Maybery, 2008; Ryan, 1969a, b). Third, a simultaneous spatial presentation array was employed instead of a sequential presentation array, consistent with typical instantiations of the Corsi-Blocks Task. During presentation of the sequence all nine locations were simultaneously visible and their presentation order was indicated by having each change colour temporarily from grey to yellow and then back to grey. An additional feature of Experiment 9 was that ungrouped and grouped sequences were administered to different groups of participants. This design choice was made to reduce the length of experimental sessions and to circumvent potential order artefacts that may have arisen in Experiment 7 due to the constant administration of ungrouped sequences prior to grouped sequences.

Method

Participants

Fifty two members of the campus community from the University of York took part in the experiment in exchange for course credit or an honorarium of £3.

Design & procedure

The experiment manipulated a single independent variable: Sequence-Type (Ungrouped / Grouped), which was a between-subjects factor. Half the participants received the ungrouped sequences, whilst the remaining half received the grouped sequences.

The procedure was identical to that of Experiment 1, with the following exceptions. First, the sequence length was increased from seven to nine locations. Second, following the 'begin trial' icon all nine locations appeared simultaneously on-screen. After a two second delay, the locations temporarily changed colour from grey to yellow, one location at a time, according to a random sequence determined by the computer programme controlling stimulus presentation. Each location was highlighted yellow for 500ms and followed by a 250ms inter-stimulus interval during which all locations remained grey. Following the change in colour of the final item the locations disappeared for 1000ms after which they re-appeared prompting forward serial order reconstruction. Third, for grouped-sequences the inter-stimulus intervals separating the third and fourth and the sixth and seventh locations were 1250ms, creating the impression of three groups of three locations.

Participants were instructed to encode sequences visually, in the absence of supplementary verbal or gestural encoding strategies, and all participants reported compliance with these instructions. Each sequence condition involved two practice trials followed by 80 experimental trials. Sessions lasted approximately 40 minutes.

Results

Accuracy serial position curves

The accuracy serial position curves for ungrouped and grouped sequences can be inspected in Figure 5-7A. Immediately apparent is that grouping fostered an elevation in recall accuracy and induced a scalloping of the serial position curve characterised by the emergence of mini within-group primacy and recency effects. The data were analysed via a 2 (Sequence-Type) X 9 (Serial Position) ANOVA, which revealed reliable main effects of Sequence-Type, $F(1, 50) = 8.795$, $MSE = 1.260$, $p < .01$, and Serial Position, $F(8, 400) = 81.502$, $MSE = 1.283$, $p < .001$, as well as a significant interaction between the two factors, $F(8, 400) = 9.227$, $MSE = .145$, $p < .001$.

Transposition gradients

The transposition gradients for the two sequence-types can be observed in Figure 5-7B and exhibit the same hallmark characteristics as documented in the preceding experiments. Of interest is whether grouping engendered an increase in the tendency for items to exchange groups, but preserve their within-group positions. Such an outcome would be reflected by discontinuities in the transposition gradient for grouped sequences, with local peaks in responses for absolute

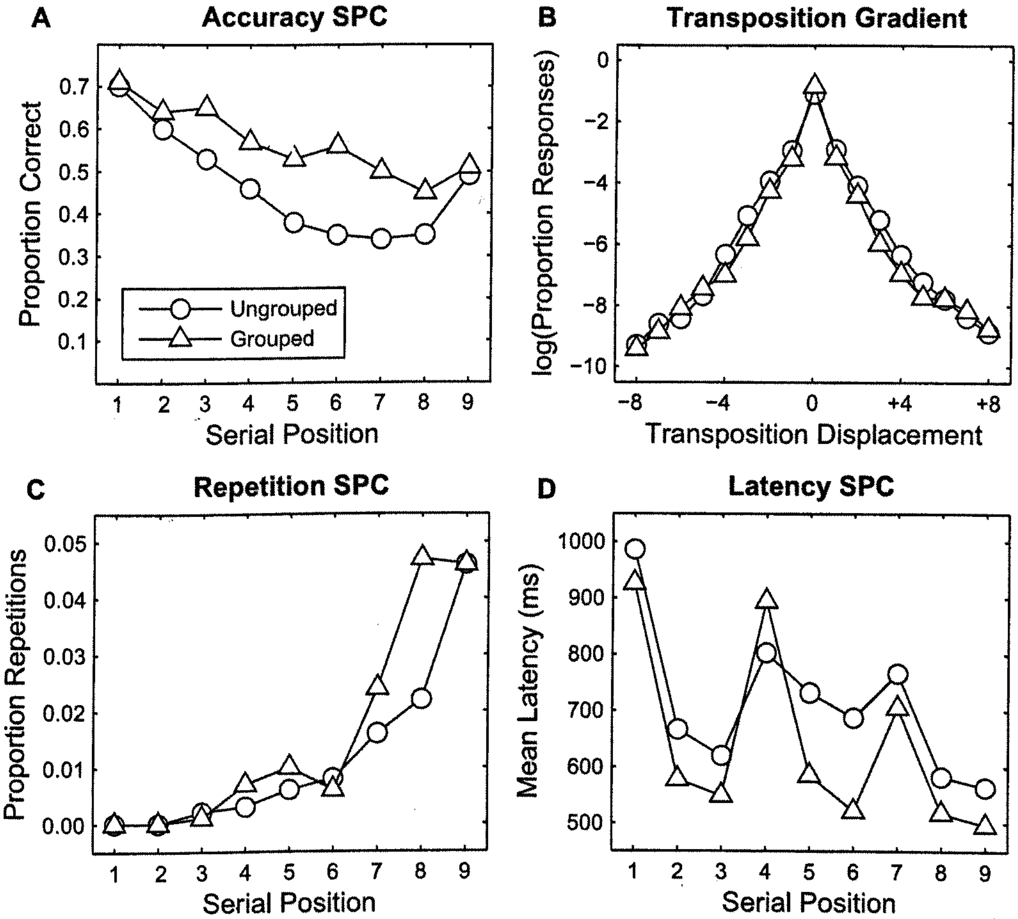

Figure 5-7 *Serial memory performance measures for Experiment 9. Panels show accuracy serial position curves (A), transposition gradients (B), repetition error serial position curves (C), and latency serial position curves (D).*

transposition displacement values of three and six. It is readily apparent that such peaks are absent in the data.

Transpositions were scrutinized further by classifying errors as occurring within or between groups, with the latter errors being further sub-divided into interpositions and other between group order errors. The proportion of errors of each type for ungrouped and grouped sequences can be inspected in Table 5-4. It is apparent that grouping increased the proportion of errors within groups, but reduced the proportion of errors between groups, including interpositions. Statistical confirmation of these observations was sought by comparing the incidence of each error-type between ungrouped and grouped sequences using t-tests performed on the log-odds transformed error proportions. These comparisons revealed a reliable increase in within group errors for

Sequence-Type	Within Groups	Between Groups	
		Interpositions	Other
Ungrouped	.56 (.10)	.13 (.04)	.31 (.08)
Grouped	.70 (.13)	.11 (.06)	.19 (.08)

Table 5-4 *Proportion of transpositions within and between groups for Experiment 9. Values in brackets represent the standard deviation of the mean.*

grouped sequences, $t(50) = 4.350, p < .001$, accompanied by reliable decreases in the incidence of interpositions, $t(50) = 2.089, p < .05$, and other between group errors, $t(50) = 5.445, p < .001$.[5]

Repetition errors

As in the previous experiments, the incidence of repetitions was well below that expected by chance. Repetitions accounted for approximately 1% of all responses for ungrouped sequences and approximately 1.5% of responses for grouped sequences. The average lag between the two instances of the repeat was 6.32 positions for ungrouped sequences and 6.23 positions for grouped sequences. Figure 5-7C shows their distribution across output positions. As before, the probability of repetitions increased sharply across serial positions for both sequence-types, however the probability of repetitions was greater for grouped than for ungrouped sequences at serial positions four, five, seven, and eight. This empirical pattern was confirmed by a 2 (Sequence-Type) X 9 (Serial Position) ANOVA performed on the log-odds transformed error proportions. This revealed a reliable main effect of Serial Position, $F(8, 400) = 54.568$, $MSE = 34.736$, $p < .001$, and although

[5] As for Experiment 7 the frequencies of interpositions when expressed as a proportion of all responses was also contrasted between ungrouped and grouped sequences. Interpositions accounted for 7% of responses for ungrouped sequences, compared to 5% of responses for grouped sequences, a significant difference between conditions, $t(50) = 2.6786$, $p = .01$. Thus, consistent with the above analysis, interpositions were more frequent for ungrouped than for grouped sequences.

the main effect of Sequence-Type was non-significant, $F(1, 50) = .155$, $MSE = .283$, $p = .695$, the Sequence-Type X Serial Position interaction, $F(8, 400) = 2.273$, $MSE = 1.477$, $p = .056$, only marginally fell short of the conventional significance level.

Latency serial position curves

The latency serial position curves associated with correct responses are plotted in Figure 5-7D. The serial position curve for grouped sequences is characterised by pronounced peaks for group initial positions with a speed up in latencies for subsequent within-group positions. The serial position curve for ungrouped sequences follows a similar profile, except the peaks for group initial positions are less punctuated, with a slower speed up in responding for the second and third positions of the second group relative to grouped sequences. These discontinuities in the serial position curve for ungrouped sequences suggest participants in this condition organised their recalls into groups of threes, in a manner similar to that for participants in the grouped condition. This observation is noteworthy given the absence of any discontinuities in the accuracy serial position curve for ungrouped sequences. These data were subjected to a 2 (Sequence-Type) X 9 (Serial Position) ANOVA, which revealed a reliable main effect of Serial Position, $F(8, 400) = 31.151$, $MSE = 3226346.364$, $p < .001$, but the main effect of Sequence-Type, $F(1, 50) = 1.902$, $MSE = 580346.722$, $p = .174$, and the Sequence-Type X Serial Position interaction, $F(8, 400) = 1.966$, $MSE = 203667.159$, $p = .129$, were both non-significant.

Latency-displacement functions

Figure 5-8 shows the LDFs with the effects of output position subtracted. The slopes of the functions for anticipations exhibit a steep negative trend, whereas with one unique deviation (the single peak at +5 displacements for ungrouped sequences) the slopes of the functions for postponements are relatively flat. The data were analysed in the same way as for Experiments 7 and 8 by performing regression analyses on the LDFs of each individual participant, which examined the relationship between transposition latency and transposition displacements that were anticipations (displacements in the range -8 to 0) and postponements (displacements in the range 0

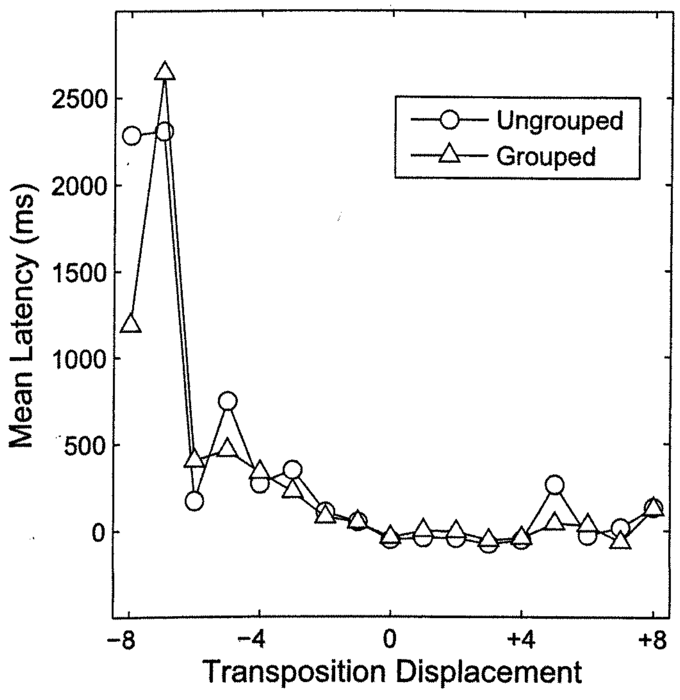

Figure 5-8 *Latency-displacement functions for Experiment 9.*

Parameter	Estimate	*SE*	*t*	*p*
Ungrouped				
Anticipation	-213.48	52.00	-4.106	.00
Postponement	19.41	12.53	1.550	.13
Grouped				
Anticipation	-148.16	35.97	-4.119	.00
Postponement	4.66	5.56	.838	.41

Table 5-5 *Mean regression slope parameter estimates for the latency-displacement functions of Experiment 9.*

to +8) separately. The resulting regression slope parameter estimates were then pooled together and subjected to one-sample t-tests to ascertain whether they differed reliably from zero.

The mean regression parameter estimates for anticipation and postponement slopes for ungrouped and grouped sequences are shown in Table 5-5. The mean regression slope estimates for anticipations were strongly negative and deviated reliably from zero: $t(25) = -4.106$, $p < .001$, for ungrouped sequences, and $t(25) = -4.119$, $p < .001$, for grouped sequences. In contrast, the slope

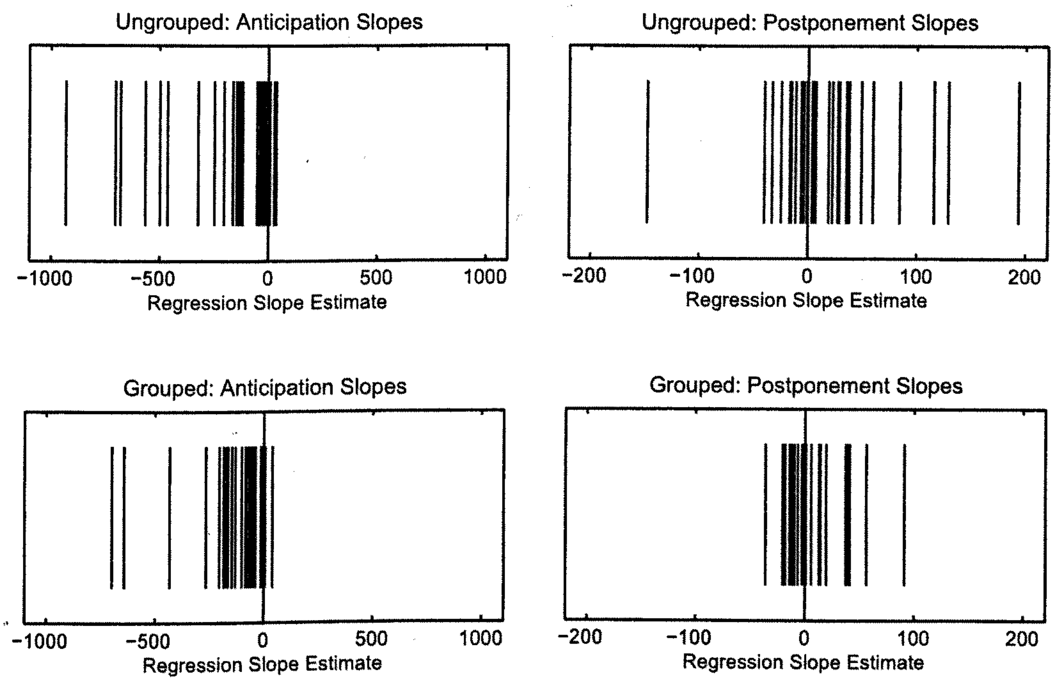

Figure 5-9 *Individual regression slope parameter estimates for the latency-displacement functions of Experiment 9. Each vertical line in a panel represents a regression slope estimate for a participant for a particular experimental condition. The left hand panels show slope estimates for anticipations, whilst the right hand panels show slope estimates for postponements. The top panels show the slope estimates for ungrouped sequences, whilst the bottom panels show the slope estimates for grouped sequences. Note—the scales of the x-axes for the slope estimates for anticipations and postponements are different.*

estimates for postponements were weakly positive, but did not deviate reliably from zero: $t(25) = 1.550$, $p = .134$, for ungrouped sequences, and $t(25) = .838$, $p = .410$, for grouped sequences.

An impression of the variability in the LDF slopes can be obtained by consulting Figure 5-9, which shows the regression slope parameter estimates for anticipations and postponements for individual participants. Consistent with the preceding experiments, the majority of participants contributed steep negative slope estimates for anticipations, whereas for postponements there was an approximately equal distribution of shallow negative and positive postponement slopes, and without exception these were all shallow slopes. This indicates once more that the empirical pattern of the aggregate LDFs is not the consequence of a small number of participants exerting undue influence on the data.

Discussion

Before discussing the LDFs, consideration is given first to the impact of the temporal grouping manipulation on the other serial memory performance measures. Consistent with Experiment 7, reliable effects of temporal grouping on performance were observed, including an elevation in recall accuracy, as well as a scalloping of the accuracy and response latency serial position curves, but this time employing longer spatial sequences and an alternative temporal grouping pattern. Although the pattern of recall latencies for the ungrouped condition is indicative of subjective grouping into threes during recall, it seems unlikely that participants engaged in subjective grouping at encoding, otherwise discontinuities in the accuracy serial position curve would also be expected. Such discontinuities were not present, however, and the overall trend was a monotonically decreasing one with an upturn in the trend line for the penultimate and final serial positions (there was also little evidence of subjective grouping into threes in the accuracy serial position curves for individual participants). Grouping also fostered a reduction in the tendency for items to transpose between groups, but consistent with Experiment 7 there was no corresponding increase in the proportion of transpositions for grouped sequences that were interpositions. The presence of grouping effects provides evidence once more for a role for positional marking in the encoding of temporally grouped spatial sequences.

Turning now to the LDFs, the empirical pattern harmonizes well with that witnessed in the previous experiments. For both ungrouped and grouped sequences, the slopes of the LDF for anticipations were steeply negative, whilst the slopes of the functions for postponements were statistically flat, providing further support for a representational mechanism combining a primacy gradient of activation, positional marking, and response suppression. A contribution of response suppression is also indicated by the scarcity of erroneous repetitions, which once again occurred at a level below that expected by chance. That the slope of the LDF for postponements was again invariant with respect to the temporal grouping manipulation provides further evidence for the role of a primacy gradient and response suppression in the encoding and retrieval of grouped sequences. An additional noteworthy feature of the LDF for grouped sequences is that the latencies for displacements that were interposition errors (transposition displacements with an absolute value of three and six) were not accelerated, consistent with the LDF for grouped sequences in Experiment 7. However, as noted previously, this outcome is not surprising given the failure of the experiments reported here to even detect an increase in the probability of such errors for grouped sequences.

The chief contribution of the present experiment has been to show that the empirical pattern of the LDFs observed in the previous experiments generalizes to the use of longer spatial sequences, a different grouping pattern, as well as the use of a simultaneous (as opposed to sequential) spatial presentation array. The use of longer sequences is particularly diagnostic, because it permitted an assessment of the relationship between latency and transposition displacement when transposition errors could span a greater number of displacements, in order to determine whether the empirical pattern hitherto observed is subject to potential range effects. That the form of the LDF remains unaltered, despite the above manipulations, is particularly telling, and further underscores the generality of this empirical signature of the primacy gradient, position marking, and response suppression model.

Summary of experiments

The empirical pattern observed across the three experiments reported here is unambiguous: the LDFs underlying serial reconstruction of spatial sequences consistently exhibited negative slopes for anticipations and flat slopes for postponements, a pattern which is most consistent with the error latency prediction of a model combining a primacy gradient of activation, associations between items and positional markers, and suppression of recalled items. This empirical outcome was shown to hold across manipulations of temporal grouping (Experiments 7 and 9), post-sequence interference (Experiment 8), sequence length (Experiments 7 and 8, versus Experiment 9), and the type of spatial presentation array employed (Experiments 7 and 8, versus Experiment 9). These findings indicate that the explanatory constructs invoked above apply across a range of settings, suggesting that they are core principles of spatial serial memory.

Quantitative model fitting

Although the LDFs observed in this chapter are most consistent with the error latency predictions of the primacy gradient, position marking, and response suppression model presented in

Chapter 3, it is necessary to determine whether this model still provides a better account of the empirical pattern than its rivals when model parameters are estimated directly from the behavioural data. In service of this requirement, the models from Chapter 3 were fit to the accuracy serial position curves and transposition gradients of the ungrouped condition of Experiment 9, and the LDFs generated under their best fitting parameter values were evaluated with reference to the data.

Fitting procedure

The fitting procedure was the same as that employed to fit the Farrell and Lewandowsky (2004) data in Chapter 3 and the data of Experiment 2 in Chapter 4. The models were fit to the accuracy data of individual participants (26 in total) using the simplex algorithm (Nelder & Mead, 1965), minimizing the summed root mean square deviation (RMSD) between the predicted and observed accuracy serial position curves and transposition gradients. The parameters underlying the competitive queuing response selection network, which includes the response threshold (T) and the excitatory (w^+) and inhibitory weights (w^-), were set to constant values of 1.0, 1.1, and -0.1, respectively. The fits of five models were evaluated: (1) position marking (PM); (2) position marking and response suppression (PM+RS); (3) position marking, output interference, and response suppression (PM+OI+RS); (4) primacy gradient and response suppression (PG+RS); (5) primacy gradient, position marking, and response suppression (PG+PM+RS).

The parameters that were free to vary during the fitting process for the PM and PM+RS models, were the distinctiveness of the position markers (ϕ), and the standard deviation of noise (σ). The PM+OI+RS model included the same free parameters, as well as an additional parameter reflecting the amount of output interference (δ). The parameters that were free to vary for the PG+RS model were the steepness of the primacy gradient (γ), and the standard deviation of noise (σ). The free parameters for the PG+PM+RS model were the steepness of the primacy gradient (γ), the distinctiveness of the position markers (ϕ), and the standard deviation of noise (σ). The model parameters frozen during the fitting process included the weighting of activation of the position markers (λ), for the PM, PM+RS, and PG+PM+RS models, the weighting of activation of the primacy gradient ($a_1 = 1$), for the PG+RS and PG+PM+RS models, the attentional weight given to the primacy gradient and positional markers ($\omega = .5$) for the PG+PM+RS model, and the amount of response suppression ($\alpha = .95$), which was universal to all models, except the PM model.

To summarise, the number of free model parameters was two for the PM, PM+RS, and PG+RS models, whilst the PM+OI+RS and PG+PM+RS models both incorporated three free parameters. To increase the chances of finding the global minimum of the goodness-of-fit functions, parameter estimates were obtained using multiple starting points for the search algorithm. These points were chosen by selecting two values for each free parameter and then factorially crossing these to create a grid of starting values. Each parameter vector explored by the search algorithm involved 10,000 model simulation trials.

Model predictions

Before scrutinizing the predictions of the models generated under their best fitting parameter values, a brief description of the goodness-of-fits of the models to the accuracy serial position curves and transposition gradients is necessary. The minimized RMSD values of the models for the fits to individual participants can be inspected in Table 5-6. It is apparent from inspection of the averaged RMSDs in this table that the PG+PM+RS model provided the best fit to the data, followed jointly by the PG+RS and PM+OI+RS models. The PM+RS model provided the next best fit, whilst the PM model provided the worst fit. Although the improvement in fit of the PG+PM+RS model over its rivals is only small, it is important to remember that these goodness-of-fits only take into consideration the descriptive accuracy of the models for the accuracy data.

Figure 5-10 shows the predictions of the models, averaged across fits to individual participants, generated after fitting to the accuracy serial position curves and transposition gradients. Considering first the predictions of the models for the accuracy serial position curves (Figure 5-10A), it is apparent that the PM model predicts a symmetrical serial position curve, which is at odds with the asymmetrical serial position curve observed empirically (Figure 5-7A). The remaining models all predict more realistic asymmetric serial position curves characterised by

Participant	PM	PM+RS	PM+OI+RS	PG+RS	PG+PM+RS
1	0.07	0.09	0.10	**0.06**	**0.06**
2	0.16	0.12	**0.08**	**0.08**	**0.08**
3	0.13	0.08	**0.07**	**0.07**	**0.07**
4	0.18	0.11	**0.08**	0.09	**0.08**
5	0.08	0.07	0.07	0.07	**0.06**
6	0.15	0.11	**0.08**	0.09	**0.08**
7	**0.06**	0.07	0.08	0.09	0.07
8	0.08	0.07	0.07	0.08	**0.06**
9	0.11	**0.07**	**0.07**	0.08	**0.07**
10	**0.14**	0.15	0.15	**0.14**	**0.14**
11	0.17	0.12	**0.04**	0.07	**0.04**
12	0.09	0.07	0.08	0.07	**0.06**
13	0.26	0.18	0.09	**0.06**	0.08
14	0.13	0.10	0.08	0.10	**0.06**
15	0.11	0.09	0.08	0.07	**0.05**
16	0.18	0.14	**0.05**	0.07	0.11
17	0.09	0.05	0.05	**0.04**	0.06
18	0.11	0.14	0.14	0.12	**0.04**
19	0.11	0.07	0.06	0.10	**0.02**
20	0.12	0.06	**0.05**	0.07	0.06
21	0.13	0.08	0.07	**0.03**	0.06
22	0.13	0.08	**0.06**	0.10	0.10
23	0.08	0.07	**0.06**	0.07	**0.06**
24	0.12	0.10	0.10	0.10	0.10
25	0.12	0.11	0.11	**0.09**	**0.09**
26	0.14	0.09	0.07	**0.06**	**0.06**
Mean RMSD	0.13	0.10	0.08	0.08	**0.07**

Table 5-6 *Minimised RMSDs for the fits of five models of serial order to the accuracy serial position curves and transposition gradients of the ungrouped condition of Experiment 9. The RMSDs of the best fitting model for each individual participant, as well as the best fitting model overall, are indicated in bold.*

greater primacy than recency. From inspection of the underlying transposition gradients (Figure 5-10B), it can be seen that all of the models predict a steep peak for displacement value zero, a locality constraint on the distribution of transpositions, and symmetrical error gradients for anticipations and postponements, consistent with the transposition gradient observed empirically (Figure 5-7B).

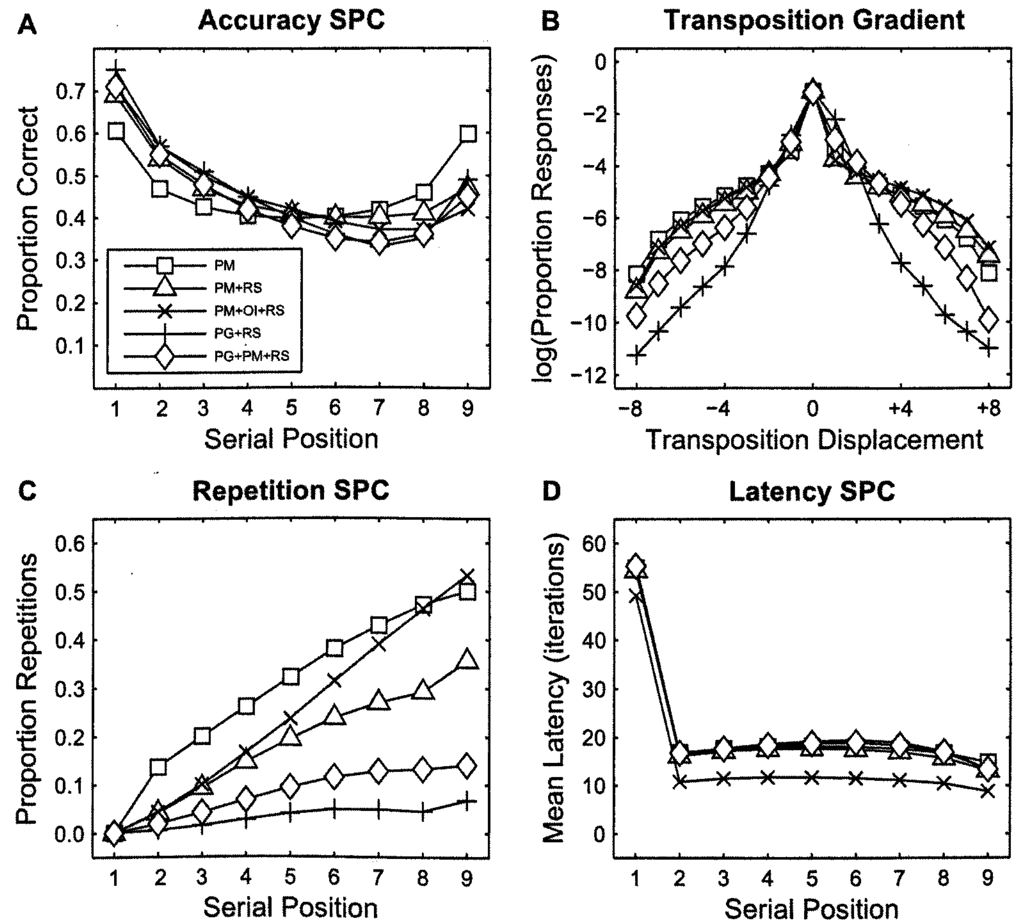

Figure 5-10 *Fits of five models of serial order to the ungrouped data of Experiment 9. Panels show accuracy serial position curves (A), transposition gradients (B), repetition error serial position curves (C), and latency serial position curves (D).*

Figure 5-10C shows the repetition error serial position curves predicted by the models from which it is apparent that all models predict the increase in the probability of repetitions across output positions observed empirically (Figure 5-7C). However, there is considerable heterogeneity in the overall proportions of repetitions predicted by the models. For example, the PM model predicted that repetitions comprised 34% of all responses, which is considerably greater than the 1% observed empirically. The PM+OI+RS model fared little better, predicting that repetitions comprised 28% of responses, whilst the PM+RS model predicted that repetitions comprised 20% of responses, which is considerably less than two abovementioned models, but still well above the empirically observed value. The PG+PM+RS model fared better still, predicting that repetitions comprised 8% of responses, whilst the best account was provided by the PG+RS model, which predicted that repetitions comprised 4% of responses. In brief, the models incorporating response suppression predicted fewer repetitions than the PM model, but any involvement of positional marking generated more repetitions than a primacy gradient with response suppression. The average lag between the two instances of the repeat was approximately 3.96 positions for the PM model, compared to an average lag of approximately 6.32 positions empirically. The remaining models more accurately captured the separation between repetitions, predicting an average lag of approximately 5 positions.

The latency serial position curves for correct responses predicted by the models can be inspected in Figure 5-10D. Note that these have had a constant 40 iterations added to the average latency for the first output position in order to increase graphical correspondence with the data (Figure 5-7D). As can be seen from inspection of this figure, all the models predict an inverted U shaped serial position curve, with faster latencies for early and late serial positions than medial serial positions. This pattern is qualitatively consistent with the overall trend observed empirically, except that the bowing of the latency curve is less marked for the model predictions than seen in the data.

Figure 5-11 shows the LDFs, averaged across fits to individual participants, generated by the models after fitting to the accuracy serial position curves and transposition gradients. Note that the effects of output position have been removed from these predictions in the same manner as described for the empirical data. It is apparent that all the models predict that the slope of the function for anticipations is negative, but there is some heterogeneity in their predicted postponement slopes. Specifically, the PM, PM+RS, and PM+OI+RS models all predict steeply positive postponement slopes, the PG+RS model predicts a negative postponement slope, whereas the PG+PM+RS model predicts a flat postponement slope[6]. Inspection of Figure 5-8 lends empirical support for the PG+PM+RS model – the slope of the LDF for postponements for

[6] A one sample t-test performed on the regression slope estimates for postponements for the individual LDFs predicted by the PG+PM+RS model confirmed that the aggregate slope estimate does not deviate significantly from zero, $t(25) = .023$, $p = .981$, and is therefore statistically flat, as found in the empirical data to which the model was applied.

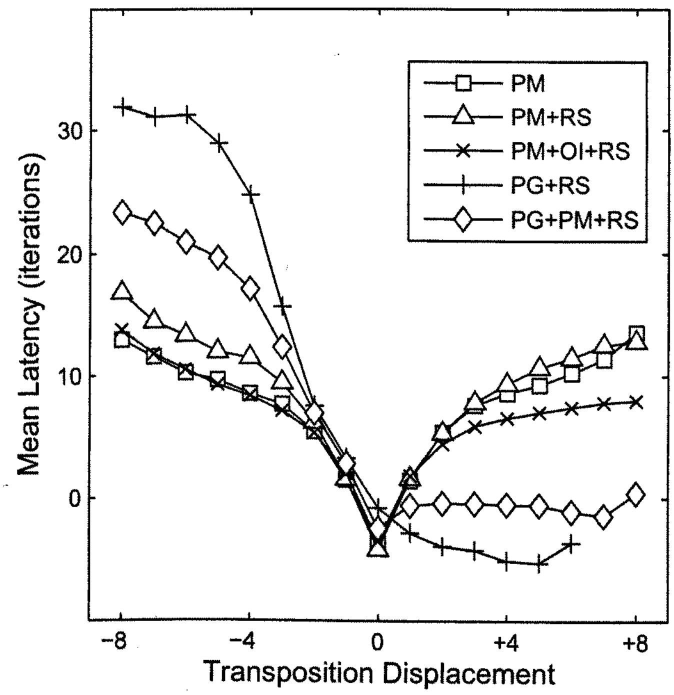

Figure 5-11 *Latency-displacement functions predicted by five models of serial order after fitting to the accuracy serial position curves and transposition gradients of the ungrouped condition of Experiment 9.*

ungrouped sequences is flat. It merits comment that these predictions were generated by fitting the models to measures of performance that do not contain any information about the dynamics of recall. In light of this, the correspondence between the LDF predicted by the PG+PM+RS model and the data is particularly impressive.

To summarize, the main outcome of the current modelling exercise has been to demonstrate that even when the models of Chapter 3 are fitted directly to representative behavioural data, only the PG+PM+RS model can accommodate the pattern of the empirically observed LDFs. This outcome suggests that any adequate model of serial order in spatial short-term memory must incorporate a mechanism combining a primacy gradient of activation, positional marking, and response suppression. The same outcome compromises the viability of the four alternative mechanisms for representing serial order.

General Discussion

The aim of the current chapter was to examine the LDFs underpinning a spatial serial reconstruction task in an attempt to identify a preferred combination of principles for representing serial order in spatial short-term memory. The results of the three experiments reported here are unambiguous in consistently revealing LDFs characterised by negative anticipation slopes and statistically flat postponement slopes (but with a positive trend). This empirical outcome harmonizes well with the pattern observed in the visual serial reconstruction experiments of Chapter 4, as well as the verbal serial recall experiments of Farrell and Lewandowsky (2004), and is most consistent with the prediction of a competitive queuing mechanism in which serial order is represented by a primacy gradient of activation, associations between items and positional markers, and suppression of emitted items. The confluence of these findings buttresses the view that verbal, visual, and spatial short-term memories rely on common principles for representing serial order information.

That the empirical pattern described above was found to hold across manipulations of temporal grouping, sequence length, post-sequence interference, as well as the type of spatial presentation array employed, suggests that these are core principles of spatial short-term memory. Qualified support for the role of these representational constructs was provided by the results of the quantitative model fitting exercise, which revealed that only the PG+PM+RS model predicted the pattern of the observed LDFs. The correspondence between the LDF predicted by this model and the data is striking when one considers that the model was not fit directly to the latency data.

One important difference between the LDFs observed in the current experiments using spatial stimuli and those of Chapter 4 using visual stimuli is that in the current experiments the slopes of the LDFs for postponements were statistically flat, but with a positive trend (with the possible exception of the LDFs for Experiment 7), whereas the slopes of the LDFs for visual stimuli consistently exhibited a statistically significant shallow positive trend. This discrepancy is attributable to greater variability in the LDFs across individual participants in the current experiments compared to the experiments of Chapter 4. Specifically, whilst the majority of participants in the experiments of the previous chapter contributed shallow positive postponement slopes, in the current experiments approximately half of participants contributed shallow positive slopes, whilst the remaining half contributed shallow negative slopes. One interpretation for this disparity between the LDFs for the visual and spatial data is that positional markers are less effective serial recall cues in the visual than in the spatial domain. This interpretation is buttressed by the values of the aggregate parameter estimates returned for the fits of the PG+PM+RS model to the data for the six-item sequence length condition of Experiment 2 employing visual stimuli, and the data for the ungrouped nine-item sequence length condition of Experiment 9 employing spatial stimuli. Specifically, the mean value of the parameter reflecting the distinctiveness of the position markers (ϕ) was considerably higher for the visual than for the spatial data (.77 compared to .59, respectively), meaning that the positional markers were more distinctive, and hence more effective recall cues, in the former case. In contrast, the mean values of the parameter reflecting the decrease in the primacy gradient (γ) were virtually identical for the visual and spatial data (.90 compared to .89, respectively).

Turning now to the impact of temporal grouping on performance, this manipulation was incorporated to provide further diagnostic information about the role of position marking in spatial short-term order memory. Consistent with a wealth of studies of verbal serial recall (Farrell & Lewandowsky, 2004; Farrell & Lelievre, 2009; Frankish, 1985, 1989; Henson, 1999; Maybery et al., 2002; Parmentier & Maybery, 2009, Ryan, 1969a, b), as well as the spatial serial reconstruction study of Parmentier et al. (2006), temporal grouping induced a multiplicity of effects, including an elevation in recall accuracy, as well as a scalloping of the accuracy and latency serial position curves, and a reduction in the tendency for items to transpose between groups. This result suggests that positional information contributes to the encoding of spatial sequences and that this information can be organised along multiple dimensions of ordering, as has been hypothesised necessary to explain grouping effects (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998a; Lewandowsky & Farrell, 2008). Nonetheless, consistent with the experiments of Farrell and Lewandowsky (2004), the qualitative form of the LDFs was generally unaffected by the temporal grouping manipulation. This result is noteworthy, because it suggests that a primacy gradient contributes even to the representation of grouped sequences. Thus, whilst it has long been acknowledged that a primacy gradient is insufficient to explain the typical effects associated with grouping, it is nevertheless necessary to explain the corresponding pattern of error latencies, which is otherwise beyond the purview of purely positional models.

One hallmark of grouping in verbal studies that failed to materialise here is the elevation in interposition errors for grouped relative to ungrouped sequences. This result is of theoretical import, because it suggests that unlike items in temporally grouped verbal sequences, items in temporally grouped spatial sequences are not coded for their within-group positions. This discrepancy points to a fundamental difference in the representations of temporally grouped verbal and spatial sequences. The question is what similarities in representation can explain the common effects of grouping across domains and what differences can accommodate the single discrepancy? The hypothesis advanced here is that in both domains order information for grouped sequences is represented along at least two dimensions, one of which necessarily represents the positions of groups in sequence. Critically, in the verbal domain the second dimension of ordering is assumed to represent the positions of items within-groups, whereas in the spatial domain this dimension is assumed to represent the positions of items in the sequence overall. Common effects of grouping across domains which include the elevation in recall accuracy, scalloping of the accuracy and latency serial position curves, as well as the reduction in between group order errors, can be attributed to the group markers common to both domains. These coarse positional representations modulate order information on the second dimension increasing the discriminability of the positional codes of items occupying different groups. Interposition errors in grouped verbal sequences are of course attributable to the use of within-group positional markers on the second dimension of ordering, which necessarily increases the positional overlap between items in different groups occupying the same positions. The absence of such errors for grouped spatial sequences is attributable to the use of positional markers that vary across sequence position, as opposed to within-group position.

An alternative hypothesis is that the absence of interpositions has nothing to do with differences in the representations of grouped spatial and verbal sequences, but rather is the consequence of the use of different recall methodologies in the current experiments and studies of grouping employing verbal material. Specifically, the recall method employed in the current experiments is serial reconstruction in which only order, but not item information must be recalled, whereas all verbal studies of which I am aware that examined interpositions have used serial recall (e.g., Farrell & Lelievre, 2009; Farrell & Lewandowsky, 2004; Henson, 1996, 1999; Ng & Maybery, 2002; Ryan, 1969a). If the increase in interpositions associated with grouping is contingent upon the explicit recall of item information then no other explanation is needed for their absence in the spatial task. These two hypotheses are examined via a further experiment and set of simulations reported in Chapter 7. Accordingly, firmer conclusions about the implications of the absence of interposition errors in the grouping experiments reported here are deferred until the end of that chapter.

An important property of the serial reconstruction task employed here was that once a location was selected at recall it only changed colour transitorily to indicate that the response had been recorded, after which the location was available for selection again, meaning that repetition errors were possible. Nevertheless, across the three experiments erroneous repetitions were extremely rare, accounting for approximately 1% of responses (an occurrence rate below that expected by chance), with the two occurrences of the repeat being separated on average by a lag of 4.67 positions across the seven-item sequence conditions of Experiments 7 and 8, and 6.23 positions across the nine-item sequence conditions of Experiment 9. The scarcity of erroneous repetitions lends further credibility for a contribution of response suppression to spatial serial memory. This was corroborated by the fits of the models to the ungrouped condition of Experiment 9, which revealed that the PM model – which was the only model not to incorporate response suppression – predicted the greatest proportion of repetitions: 34% compared to the 1% observed empirically. With the exception of the PM+OI+RS model, the remaining models incorporating response suppression predicted considerably fewer repetitions. For example, the PG+PM+RS model predicted that repetitions comprised 8% of responses, which although greater than that seen empirically is considerably less than that predicted by the PM model. Note also that if the PG+PM+RS model were fit to the transposition error gradients for each serial position – which carry more information about the distribution of responses than the aggregate transposition gradients – then it is possible the model would predict even fewer repetitions.

It merits comment that the incidence of repetitions and their average separation across serial positions is slightly different to that documented for the serial recall of sequences of verbal items. For example, Henson (1996) found that repetitions comprised approximately 2% of all responses and were separated by an average lag of 3.34 positions, whilst Vousden and Brown (1998) found that repetitions comprised approximately 5% of responses. The greater incidence of repetitions for verbal material combined with the shorter average lag between the two instances of the repeats tentatively suggests that response suppression is stronger and longer lasting in the spatial than the verbal domain. However, it is important to emphasise that this claim is based upon a comparison of verbal serial recall with spatial serial reconstruction, rendering it possible that these differences are attributable to the different recall methods employed in the two domains.

One criticism that might be levelled against the task employed here to probe spatial serial memory is its reliance on a fixed set of spatial coordinates. As noted in the introduction to this chapter, some commentators have suggested that this renders the task more vulnerable to verbal encoding strategies than a task involving unique locations on each trial (Couture & Tremblay, 2006; Jones et al., 1995). However, it was also noted that studies of spatial memory using fixed locations have shown that articulatory suppression has only a negligible (Meisser & Klauer, 1999), if any effect (Smyth & Scholey, 1992; Smyth et al., 1988) on spatial serial reconstruction accuracy. Added to this, participants in all three experiments reported here were explicitly instructed not to engage in verbal encoding of spatial sequences and all reported compliance with this instruction. Notwithstanding, these considerations, it would seem prudent to establish in the future whether the effects reported here are replicable under conditions where opportunities to engage in verbal encoding are further minimised through the use of unique spatial locations across trials or by incorporation of an articulatory suppression manipulation. An experiment that meets this requirement is presented in Chapter 8.

Before concluding this chapter one further issue merits comment. As noted in the penultimate paragraph of Chapter 4, the claim that the PG+PM+RS model provides the best account of the data rests upon a local analysis of the behaviour of the models. What is not yet clear is whether the error latency predictions of the models hitherto observed, represent the most frequent data patterns those models predict, of all the possible data patterns they can generate. The theoretical significance of the global predictive behaviour of models should not be underestimated. On the one hand, if it can be shown that the PG+PM+RS model predicts flat/shallow-positive postponement slopes across a large portion of its parameter space then we can be confident that this prediction follows from its core principles. On the other hand, if the model only predicts flat/shallow-positive postponement slopes on a minority of occasions then this would indicate that this theoretical prediction is unrepresentative of the model's general behaviour and cannot therefore be tied to its core assumptions. A global analysis of the behaviour of the models is the goal of the next chapter.

Chapter summary

The three experiments in this chapter have consistently shown that the LDFs underlying serial reconstruction of sequences of spatial locations are characterised by negative anticipation slopes and flat postponement slopes. The generality of this pattern is underscored by the fact that it was observed across manipulations of temporal grouping, post-sequence interference, and sequence length. This empirical outcome is most compatible with the error latency prediction of a competitive queuing mechanism in which serial order is represented via a primacy gradient of activation, associations between items and positional markers, and suppression of emitted items. Combined with the results of Chapters 3 and 4, these findings suggest that common principles represent serial order across the verbal, visual, and spatial domains. Nevertheless, the failure to detect an increase in interpositions for temporally grouped spatial sequences may indicate a fundamental difference in how positional information is encoded in spatial and verbal serial memory. This possibility is explored in Chapter 7.

6

Parameter space sensitivity analyses

Abstract

This chapter presents two sets of parameter space sensitivity analyses which examined the latency-displacement functions predicted by the PM+RS, PM+OI+RS, PG+RS, and PG+PM+RS models, across a wide range of their parameter settings, for short sequences of six-items (Analysis 1), and longer sequences of nine-items (Analysis 2). The purpose of these analyses was to establish whether the latency-displacement functions predicted by these models in Chapters 3, 4, and 5, which are based upon local parameter settings, are representative of their wider behaviour. In general, the outcomes of these analyses indicate that the latency-displacement functions predicted by these models in previous chapters are central features of their behaviour, and thus attributable to their core representational principles.

Introduction

The preceding chapters have provided evidence for the role of a primacy gradient of activation, positional marking, and response suppression in verbal, visual, and spatial short-term order memory. The evidence for these three representational principles comes from comparisons of the empirically observed latency-displacement functions (LDFs) with qualitative predictions of the models presented at the outset of Chapter 3, as well as the LDFs predicted by the models under best fitting parameter values.

However, as noted in Chapters 4 and 5, one shortcoming of this evidence is that it stems from a local analysis of the behaviour of the models. That is, the predictions scrutinized so far are based upon a restricted number of each model's parameter settings, and therefore say nothing about the predictions of the models under their wider range of parameter settings. Local model analysis, by means of data fitting, is important for determining whether a model is sufficient to reproduce a data pattern that is a signature of the cognitive process it was designed to explain. Nevertheless, as noted in Chapter 2, a model can provide a good fit to data for reasons other than it being a good approximation of that process (Myung, 2000; Myung et al., 2005; Pitt et al., 2006, 2008; Pitt & Myung, 2002). Specifically, a complex model with many free parameters (and, or an overly friendly functional form) can provide good fits to data simply because this complexity grants the model the flexibility to explain many different data patterns, some of which may be seen empirically, and some of which may not.

As noted by Pitt et al. (2008), model flexibility is a double-edged sword. On the one hand, a model must be sufficiently flexible that it can explain the many different data patterns that are signatures of the cognitive process it was devised to explain. On the other hand, a model should not be so flexible that it can explain almost any data pattern. A cognitive model that predicts many different data patterns not witnessed empirically in addition to the data pattern of central interest is not a useful model, because it is unlikely that the model's overall behaviour can be tied to a coherent set of underlying principles. On the contrary, the variations in the model's predictions are likely to be due to the fact that its underlying assumptions are in fact quite arbitrary. A thorough evaluation of the theoretical adequacy of a cognitive model must therefore take its flexibility into consideration, in addition to its ability to closely mimic data.

The problem of model flexibility is pertinent because the model that consistently provided the best descriptions of the LDFs in previous chapters, the primacy gradient, position marking, and response suppression (PG+PM+RS) model, is the most complex of the field of competitor models under consideration. This model incorporates five parameters, which is one more than the position marking, output interference, and response suppression (PM+OI+RS) model, two more than the position marking and response suppression (PM+RS) and primacy gradient and response suppression (PG+RS) models, and three more than the position marking (PM) model. It therefore follows that the better description of the data provided by this model might be attributable to its greater complexity and flexibility.

Yet there is *a priori* reason to believe that this is not the case. Specifically, in Chapters 3, 4, and 5 the models were not fit directly to the LDFs, but rather to the accuracy serial position curves and transposition gradients. The important point to emphasise here is that these data do not contain any information whatsoever about the dynamics of recall. That the PG+PM+RS model predicted the form of the LDFs despite being given no information about them during the fitting process implies that this prediction stems from its core underlying principles. Nevertheless, a firm conclusion about the theoretical adequacy of this model depends upon an assessment of its flexibility.

In their original work, Farrell and Lewandowsky (2004) addressed the issue of model flexibility by examining the sensitivity of the LDFs predicted by the PM, PM+RS, PM+OI (position marking and output interference), and PG+RS models to variations in their parameter settings; a technique dubbed *parameter space sensitivity analysis* (cf., Li, Lewandowsky, & DeBrunner, 1996). For each model, a broad range of values was specified for each of its parameters and these were factorially combined to create a grid of parameter setting combinations and the predicted LDF for each determined by simulation of the model. This yielded for each model a large number of LDFs covering a broad range of its parameter space. The dependent measure of central interest was the slope of the LDFs for postponements only; since the models all predict negative anticipation slopes, only the slopes of the functions for postponements can differentiate between the models.

Farrell and Lewandowsky examined the behaviour of the models by plotting their predicted postponement slopes as density histograms showing the proportions of simulations falling into bins representing postponement slope estimates covering different slope intervals. Their initial explorations of the predictions of the models, combined with fits of the models to data, had revealed that the PM, PM+RS, and PM+OI models all predict steep positive postponement slopes, whereas the PG+RS model predicts a negative postponement slope. The outcomes of the parameter space sensitivity analyses revealed that the former models consistently predict steep positive postponement slopes and no negative slopes whatsoever, whereas the latter model consistently predicts shallow negative postponement slopes in conjunction with a small number of shallow positive postponement slopes. The outcomes of the parameter space sensitivity analysis therefore confirmed that the initial qualitative predictions obtained for the models are representative of their more general behaviour, and can thus be tied to their core underlying principles. Thus, there was no evidence to indicate that any of the models were unduly flexible.

Of course, Farrell and Lewandowsky (2004) did not perform a parameter space sensitivity analysis of the PG+PM+RS model, since this model was not considered until a later paper (Lewandowsky & Farrell, 2008), in which the authors presented only qualitative predictions of the model. The aim of the current chapter is to replicate the parameter space sensitivity analyses of Farrell and Lewandowsky (2004), but this time incorporate the PG+PM+RS model into the model comparisons. It is important to emphasise at this juncture that the PG+PM+RS model is the preferred model of the data, of the field of competitor models compared, irrespective of the outcomes of these new analyses. This is because the PG+PM+RS model is the only model of those considered that can predict the empirical pattern of the LDF, when its parameters are estimated directly from the data. What these new analyses will determine is the robustness of the predictions generated by this model so far. If it turns out that the model predicts the empirical pattern across a large portion of its parameter space then we can be confident that this prediction stems from its core assumptions. However, if it turns out that the model predicts the empirical pattern on only a minority of occasions this would suggest that it is overly flexible, and that the preferred model of the data is in fact one not considered in the model comparisons conducted so far.

Before proceeding with the analyses, it is necessary to scrutinize the predictions of the PG+PM+RS model and the data in further detail in order to identify the predictions that would be expected of this model if it is a good theory of the data. Recall from Chapters 4 and 5 that in addition to presenting figures showing the aggregate LDFs, figures were also included showing the distribution of anticipation and postponement slope estimates for individual participants. The panels of the latter figures showing the distributions of slopes for postponements exhibit a number of important empirical regularities, which are summarised in Figure 6-1. This figure presents two density histograms, which show the distribution of individual participant postponement slope estimates from a number of experimental conditions. Figure 6-1A shows data from the six-item sequence conditions of Experiments 1 to 4 (128 observations), involving visual serial reconstruction, whilst Figure 6-1B shows the data from the ungrouped and grouped nine-item sequence conditions of Experiment 9 (52 observations), involving spatial serial reconstruction. Each histogram plots the normalized proportions of postponement slopes (represented on the y-

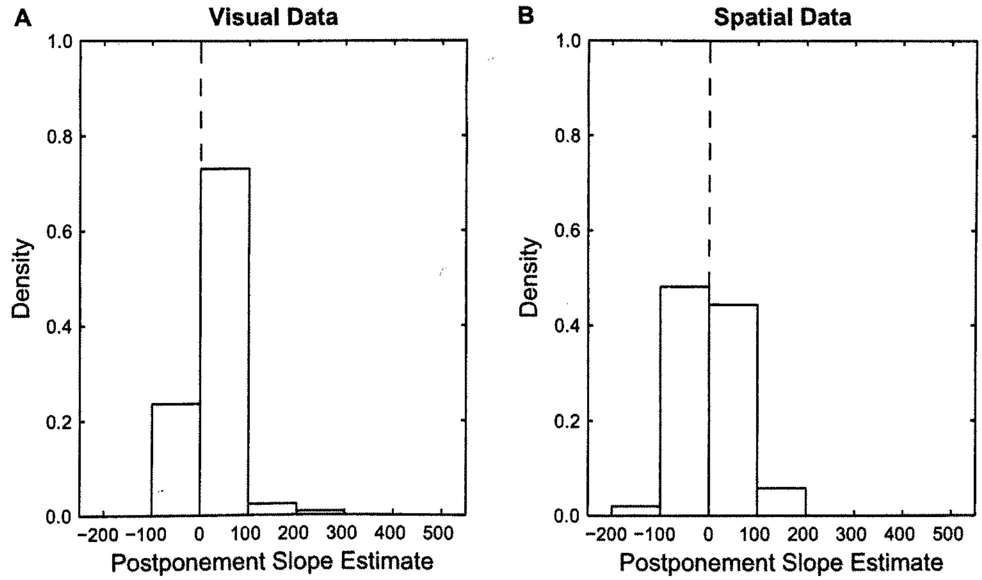

Figure 6-1 *Density histograms showing the distribution of latency-displacement function postponement slopes for individual participants for a sub-set of the visual (A) and spatial (B) serial reconstruction data. See main text for further details.*

axis) falling into bins representing slope estimates covering different millisecond intervals (represented on the x-axis). The broken line in each panel represents a baseline slope value of zero. Accordingly, bars falling to the right of the broken line correspond to positive postponement slope estimates, whilst bars falling to the left correspond to negative postponement slope estimates.

Immediately apparent from inspection of these figures is that the majority of postponement slopes are shallow slopes lying in the range of -100ms to 100ms. Indeed, the proportion of slopes lying within this range is 96% for the visual data and 92% for the spatial data. Within this range the majority of slopes in the visual data (73%) are shallow positive slopes falling in the range of 0ms to 100ms, with a smaller percentage (23%) of shallow negative slopes falling in the range of 0ms to -100ms. The distribution is slightly different for the spatial data, with approximately equal proportions of shallow negative and positive slopes (values equal to 48% and 44%, respectively).

Putting aside for now the slightly different distribution of negative and positive slopes for the visual and spatial data, the crucial aspect of these histograms is that they indicate that the flat/shallow-positive postponement slopes observed in the aggregate LDFs of previous chapters are the consequence of a mixture of shallow negative and positive postponement slopes. In the model fitting of Chapters 3, 4, and 5, the models were applied to the data of individual participants. Inspection of the postponement slope estimates for the fits of the PG+PM+RS model reveals that this model predicts a mixture of shallow negative and positive postponement slopes across participants. For example, in the fits to the data of Experiment 3 of Farrell and Lewandowsky (2004; n = 26) 16 of the slopes were negative and 10 were positive; in the fits of the model to the six-item sequence condition of Experiment 2 (n = 18) 11 of the slopes were negative and 7 were positive; and in the fits of the model to the ungrouped condition of Experiment 9 (n = 26) 16 of the slopes were negative and 11 were positive. In short, although the model predicts a flat/shallow-positive slope on average, this is in fact the result of a combination of shallow negative and positive slopes, as in the empirical data. In sharp contrast to this, the PM+RS and PM+OI+RS models consistently predicted steep positive slopes across participants, whereas the PG+RS model consistently predicted negative slopes across participants.

Bearing the above in mind, the question of interest in the following analyses is how often does the PG+PM+RS model predict shallow negative and positive postponement slopes? More specifically, how often does it predict slopes falling within the range of -100ms to 100ms (which Figure 6-1 clearly indicates is where the majority of empirical observations are found)? This question cannot be directly addressed on the basis of the raw model predictions generated from the parameter space sensitivity analysis. This is because the predictions of the models are measured in iterative cycles, whereas the data are measured in milliseconds. In the analyses that follow this problem is overcome by running linear regression analyses in which each LDF predicted by a model under a specific combination of parameter values is used to predict the aggregate LDF from an appropriate experiment. This generates two scaling parameters for each model LDF: an intercept parameter (measured in milliseconds) and an iteration-to-millisecond slope parameter that can then be used to convert each LDF from measurement in model iterations to milliseconds.

The remainder of this chapter consists of two sets of parameter space sensitivity analyses. The first analysis involves simulations of the models applied to sequences of six-items. The second involves simulations of the models applied to sequences of nine-items. The latter analysis is included to ensure that the results obtained from the first analysis are not specific to the sequence length employed.

Parameter space analysis 1

The first analysis examined the predictions of the models for six-item sequences, the sequence length most frequently employed in the visual serial reconstruction experiments of Chapter 4.

Modelling procedure

The parameters underlying the competitive queuing response selection network, which includes the response threshold (*T*), the excitatory (*w*+) and inhibitory weights (*w*-), and the standard deviation of random Gaussian noise applied to the starting activations (σ), were set to constant values of 1.0, 1.1, -0.1, and .04, respectively. The predictions of four models were examined; these were the PM+RS, PM+OI+RS, PG+RS, and PG+PM+RS models. For each model, a vector of values varying from .05 to .95 in increments equal to .1 was specified for each of its parameters. These parameter values were then factorially combined to create a grid of parameter value combinations (each a vector) to be explored by simulation. The parameters varied for the PM+RS model were the weighting of activation for the position markers (λ), the distinctiveness of the position markers (ϕ), and the amount of response suppression (α). The same parameters were manipulated for the PM+OI+RS model, in addition to the amount of output interference (δ). The parameters varied for the PG+RS model were the weighting of activation for the first item (a_1), the steepness of the primacy gradient (γ), and the amount of response suppression (α). The parameters varied for the PG+PM+RS model were the same as those for the PM+RS and PG+RS models. Note that the weight parameter (ω) governing the attention assigned to the primacy and positional dimensions of ordering in the PG+PM+RS model was set to a fixed value (equal to .5).

In summary, three parameters were varied for the PM+RS and PG+RS models, four for the PM+OI+RS model, and five for the PG+PM+RS model. This resulted in 1,000 parameter value combinations for the PM+RS and PG+RS models (10^3), 10,000 combinations for the PM+OI+RS model (10^4), and 100,000 combinations for the PG+PM+RS model (10^5)[1]. For each parameter vector explored, 1000 simulation trials of six-item sequences were run and the associated model predictions for accuracy and latency serial position curves, transposition gradients, and LDFs were recorded.

The dependent measure of central interest was the slope of the LDF for postponements only. As noted previously, because all of the models predict negative slopes for anticipations, only the slopes for postponements can serve to distinguish between the models.

Model predictions

Although the predictions of the models for the postponement slopes of the LDFs are of chief interest, I begin first by considering the aggregate model predictions for accuracy serial position curves, transposition gradients, and latency serial position curves. This will help to determine the extent to which the models, on average, predict realistic levels and patterns of recall accuracy.

Accuracy serial position curves

The accuracy serial position curves predicted by the models, averaged across simulations, are shown in Figure 6-2A. It is apparent from inspection of this figure that the PG+PM+RS model predicts on average a more realistic serial position curve than the PM+RS model. Whilst the former model predicts an asymmetric serial position curve characterised by greater primacy than recency, the latter model predicts a symmetrical serial position curve. Both models in turn predict on average more realistic serial position curves than the PM+OI+RS and PG+RS models. The latter models predict on average considerably lower levels of recall accuracy, and exhibit an overly extensive effect of primacy combined with no effect of recency.

---

[1] The PM model was excluded from the analysis, because the parameter sampling procedure adopted would have generated only a limited number of observations. Specifically, with only two model parameters and 10 values per parameter this would only have generated 100 model predictions (10^2). However, Farrell and Lewandowsky (2004) have shown that the PM model generates a preponderance of steep positive postponement slopes.

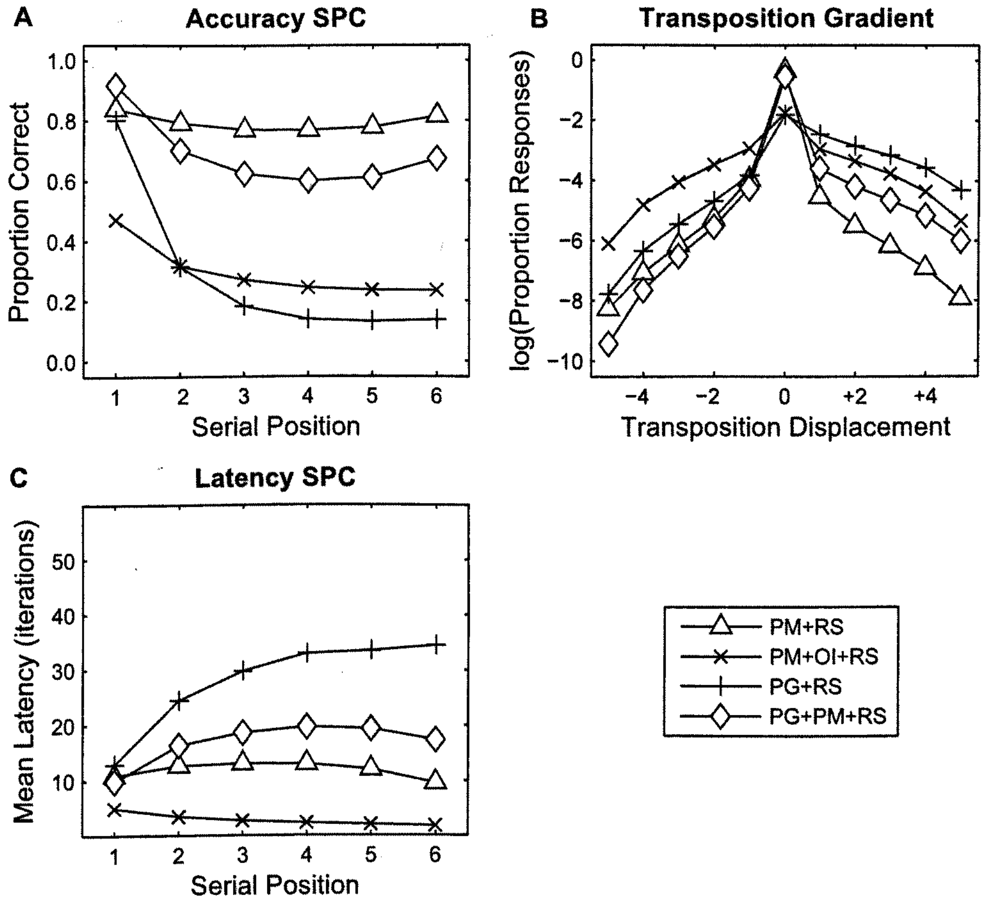

Figure 6-2 *Aggregate predictions of four models of serial order for the first parameter space analysis. Panels show accuracy serial position curves (A), transposition gradients (B), and latency serial positions curves (C).*

The strong effect of primacy and lack of recency in the aggregate serial position curve of the PM+OI+RS model is attributable to those simulations involving moderate to high levels of output interference. Under these conditions recall accuracy will necessarily be extremely poor, particularly at medial and terminal serial positions (given the increasing nature of output interference across serial positions). In contrast, the extensive primacy and absence of recency in the aggregate serial position curve of the PG+RS model is attributable to those simulations involving low settings of the response suppression parameter. When the amount of response suppression is small performance will be low at all but the first couple of serial positions, because response suppression is necessary to prevent perseveration on the same response. Recall of the final item will also be impaired, since response suppression is necessary to produce recency when serial order is represented via a primacy gradient.

Transposition gradients

Figure 6-2B shows the transposition gradients predicted by the models, averaged across simulations. It can be seen that all four models predict a peak at transposition displacement zero, reflecting that the majority of responses are correct responses. The peaks are less pronounced on the transposition gradients for the PM+OI+RS and PG+RS models for the reasons delineated above. It can also be seen that all models predict a locality constraint on the distribution of transpositions, in addition to approximately symmetrical error gradients for anticipations and postponements.

Latency serial position curves

The latency serial position curves for correct responses predicted by the models, averaged across simulations, are shown in Figure 6-2C[2]. It can be seen that both the PM+RS and PG+PM+RS models predict on average inverted U shaped latency serial position curves, consistent with the data from verbal serial recall (Farrell & Lewandowsky, 2004; Maybery et al., 2002; Parmentier & Maybery, 2009) and spatial serial reconstruction (Experiments 7-9 of Chapter 5). In contrast, the PM+OI+RS model predicts on average a monotonically decreasing latency serial position curve, consistent with the data from visual serial reconstruction (Experiments 1-6 of Chapter 4). This negative monotonicity is a consequence of output interference pushing items closer and closer to the decision threshold, as output position increases. Output interference is also the source of the much faster latencies for the PM+OI+RS model than the other three models. Finally, the PG+RS model predicts a monotonically increasing latency serial position curve. The

[2] The reader is reminded that none of the models predict the long initial latency observed empirically in the experiments of previous chapters. As noted in Chapter 3, the modelling of such preparatory latencies requires further ancillary assumptions. Whereas in the simulations of previous chapters a constant was added to the latency for the first output position, in the current chapter the latencies are presented without the addition of this constant.

absence of a dip in the latencies towards the end of the sequence is attributable to simulations involving low settings of the response suppression parameter, which as noted above, will necessarily attenuate the recency effect.

Latency-displacement function postponement slopes

In order to facilitate comparison of the model predictions with the data, the individual postponement slopes predicted by each of the four models were first converted from measurement in model iterative cycles to milliseconds. This was accomplished using the following three stage procedure. In the first stage, regression analyses were performed for each model in which the LDF predicted under each combination of parameter settings (with the effects of output position subtracted) was entered as a predictor variable and the aggregate LDF for the six-item sequence condition of Experiment 1 was used as the dependent variable[3]. For each predicted model LDF this yielded two scaling parameters: an intercept parameter (in milliseconds) and an iteration-to-millisecond slope parameter. In the second-stage, each predicted model LDF was transformed from iterations to milliseconds employing the scaling parameters obtained from the first-stage analyses. In the final stage, another set of regression analyses were performed in order to obtain the postponement slope estimates for the transformed LDFs of each of the models. The distributions of LDF postponement slopes predicted by the models are shown graphically in Figure 6-3 as density histograms similar to those presented in Figure 6-1.

Considering first the predictions of the PM+RS model (Figure 6-3A), it can clearly be seen that this model predicts a majority of positive slopes. Indeed, 98% of the postponement slopes predicted by this model are positive slopes. Furthermore, most of the postponement slopes predicted by this model are steep positive slopes falling in the range of 100ms to 200ms (73% of model predictions). These simulations therefore confirm that the steep positive slopes predicted by this model in previous chapters are a general feature of its behaviour.

[3] Only postponement slopes associated with less than perfect recall accuracy were included in the analysis, since perfect recall accuracy removes postponement errors entirely.

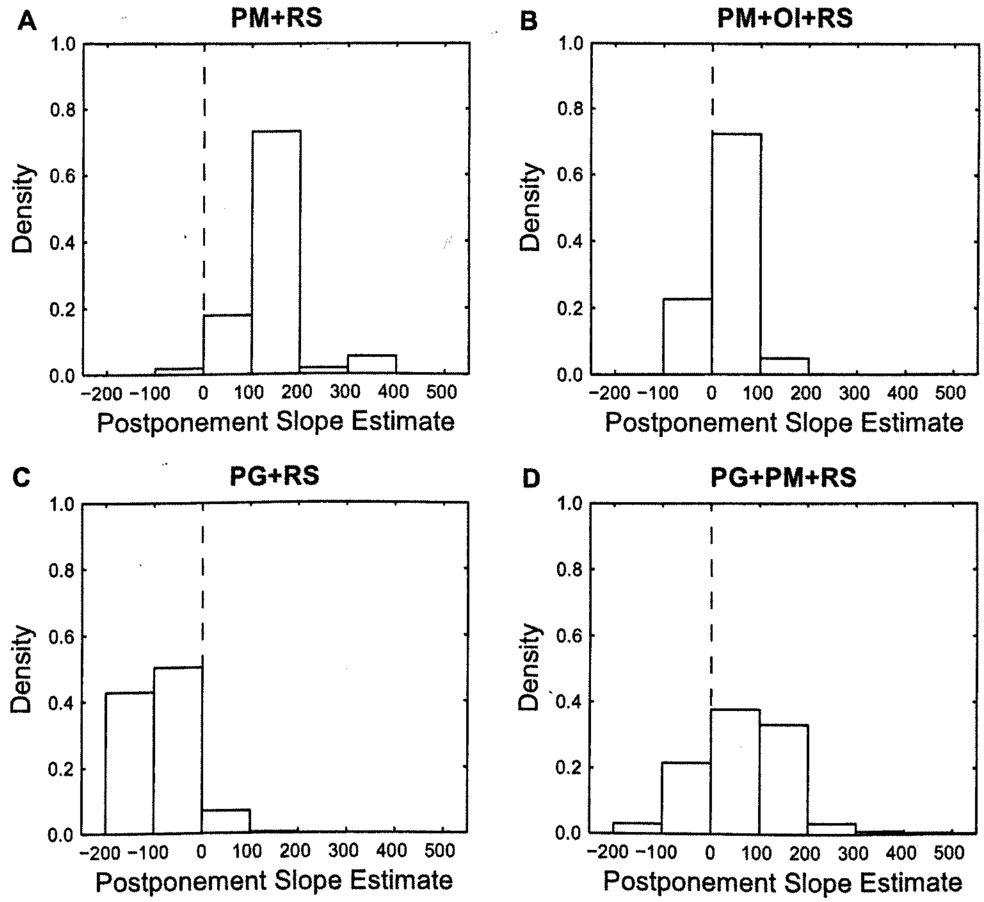

Figure 6-3 *Density histograms showing the distribution of latency-displacement function postponement slopes predicted by four models of serial order for the first parameter space analysis. Panels show predictions for the PM+RS model (A), the PM+OI+RS model (B), the PG+RS model (C) and the PG+PM+RS model (D).*

Turning to the PM+OI+RS model (Figure 6-3B), it can be seen that this model also predicts a majority of positive slopes (76% of model predictions). However, unlike the PM+RS model, this model mostly predicts shallow positive slopes falling in the range of 0ms to 100ms (72% of model predictions) and only predicts a small proportion of steep positive slopes falling in the range of 100ms to 200ms (4% of model predictions). The model also predicts a modest proportion of shallow negative slopes falling within the range of -100ms to 0ms (23% of model predictions). Simulations of this model in Chapters 3 and 5 showed that it predicts a steep positive postponement slope. The outcomes of the current analysis reveal that this prediction is not a core feature of this model's behaviour and that the model does in fact exhibit an overall tendency to predict shallow positive slopes. The reason for this discrepancy is that the simulations of this model presented in previous chapters involved relatively low settings of the output interference parameter, whilst the simulations reported here have included a broad range of parameter values.

It is noteworthy that the predictions of the PM+OI+RS model correspond exceptionally well with the empirical distribution for six-item sequences shown in Figure 6-1A. However, despite this correspondence, the PM+OI+RS model is not a viable theory of the data. This is because applications of this model in Chapter 5 to empirical data clearly showed that the model is unable to accommodate the empirical pattern of the LDF when its parameters are estimated directly from the data. Basically, although the model predicts a majority of shallow postponement slopes, these shallow slopes are associated with unrealistic levels of recall accuracy. Qualified support for this claim is shown in Figure 6-2A, which shows that the PM+OI+RS model predicts on average levels of recall accuracy much lower than that observed in the experiments, in addition to an aggregate serial position curve devoid of recency.

Considering now the predictions of the PG+RS model (Figure 6-3C), it can be seen that this model predicts a majority of negative slopes. Indeed, 93% of the postponement slopes predicted by this model are negative slopes. Approximately half of these slopes are shallow negative slopes falling within the range of -100ms to 0ms (43% of model predictions) and approximately half are steep negative slopes falling within the range of -200ms to -100ms (50% of model predictions). The small proportion of remaining slopes predicted by this model are shallow positive slopes falling within the range of 0ms to 100ms (6% of model predictions). These simulation results confirm that the negative postponement slopes predicted by this model in previous chapters are a core feature of its overall behaviour.

Finally, the PG+PM+RS model (Figure 6-3D) predicts a large proportion of shallow postponement slopes falling within the range of -100ms to 100ms (60% of model predictions). Of these slopes the model predicts a greater proportion of shallow positive slopes falling within the range of 0ms to 100ms (38% of model predictions) than shallow negative slopes falling within the range of -100ms to 0ms (22% of model predictions). However, the model additionally predicts a significant proportion of steep positive slopes. Specifically, 33% of the slopes predicted by the model lie within the range of 100ms to 200ms, with a further 3% falling within the range of 200ms to 300ms. The results show that shallow negative and positive postponement slopes are a central prediction of the PG+PM+RS model. However, they are not a ubiquitous feature of its behaviour; the model additionally predicts an appreciable number of steep positive slopes.

Overall, the results of this first wave of analyses show that the PM+RS, PM+OI+RS, and PG+RS models exhibit a robust set of predictions. The PM+RS model consistently predicts steep positive postponement slopes, the PM+OI+RS model consistently predicts shallow positive postponement slopes, and the PG+RS model consistently predicts negative postponement slopes. Thus, there is no indication that either of these models is unduly flexible. However, the PG+PM+RS model exhibits more flexibility in its predictions than its rivals, predicting a majority of shallow slopes, but also a significant number of steep positive slopes. Nevertheless, since the model predicts shallow postponement slopes as its central prediction, this outcome suggests that the PG+PM+RS model is not unduly flexible.

Parameter space analysis 2

The second analysis replicated the first analysis in all respects, except that the sequence length was increased to nine-items – the sequence length employed in Experiment 9 of Chapter 5 examining serial reconstruction of spatial sequences. The main reason for conducting this second wave of analyses was to make sure that the core findings from the first wave of analyses are not specific to the sequence length studied.

Model predictions

Accuracy serial position curves

The accuracy serial position curves predicted by the models, averaged across simulations, are shown in Figure 6-4A. Consistent with the results of the first analysis, the PG+PM+RS model predicts on average a more realistic serial position curve than the PM+RS model, and both models in turn predict on average more realistic serial position curves than the PM+OI+RS and PG+RS

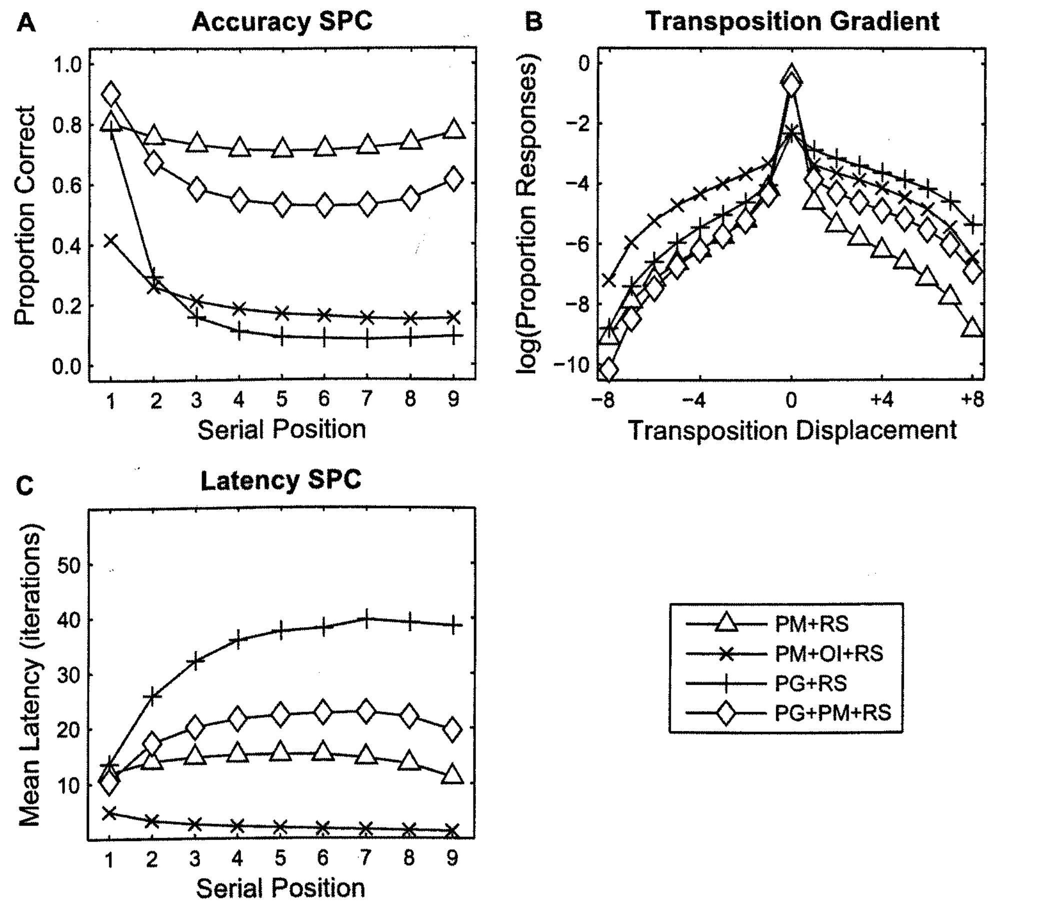

Figure 6-4 *Aggregate predictions of four models of serial order for the second parameter space analysis. Panels show accuracy serial position curves (A), transposition gradients (B), and latency serial positions curves (C).*

models. The latter models predict on average considerably lower levels of recall accuracy, and exhibit an overly extensive effect of primacy coupled with no effect of recency.

Transposition gradients

Figure 6-4B shows the transposition gradients predicted by the models, averaged across simulations. It can be seen that all four models predict a peak at transposition displacement zero, reflecting that the majority of responses are correct responses. The peaks are less pronounced on the transposition gradients for the PM+OI+RS and PG+RS models, reflecting the greater proportions of transpositions predicted by these models. As before, all models predict a locality constraint on the distribution of transpositions, in addition to approximately symmetrical error gradients for anticipations and postponements.

Latency serial position curves

The latency serial position curves associated with correct responses predicted by the models, averaged across simulations, are illustrated in Figure 6-4C. Consistent with the previous analysis, both the PM+RS and PG+PM+RS models predict on average inverted U shaped latency serial position curves; the PM+OI+RS model predicts on average a monotonically decreasing latency serial position curve; whilst the PG+RS model predicts on average a monotonically increasing latency serial position curve.

Latency-displacement function postponement slopes

The procedure adopted for converting the model latency predictions from measurment in model iterative cylces to milliseconds was exactly the same as employed in the first analysis, except that the individual LDFs predicted by the models were scaled with reference to the aggregate LDF of the ungrouped condition of Experiment 9, which like the model simulations is based upon sequences of nine-items. The distributions of LDF postponement slopes predicted by the models are shown in Figure 6-5 as density histograms.

Starting with the PM+RS model (Figure 6-5A), as before this model predicts a majority of positive postponement slopes (93% of model predictions). However, the model predicts fewer steep positive slopes falling within the range of 100ms to 200ms than in the first analysis (39% of model predictions, compared to 73% of model predictions, respectively). Additionally, the model predicts a larger proportion of shallow positive slopes lying within the range of 0ms to 100ms than in the first analysis (56% of model predictions, compared to 18% of model predictions, respectively). Thus, for longer sequences of nine-items the PM+RS model still predicts an abundance of positive slopes, but these slopes are less steep on average than those predicted for six-item sequences.

Turning to the PM+OI+RS model (Figure 6-5B), this model also once again predicts a majority of positive postponement slopes (80% of model predictions). All of these slopes are shallow slopes falling within the range of 0ms to 100ms. The model additionally predicts a modest number of shallow negative slopes falling within the range of -100ms to 0ms (20% of model predictions).

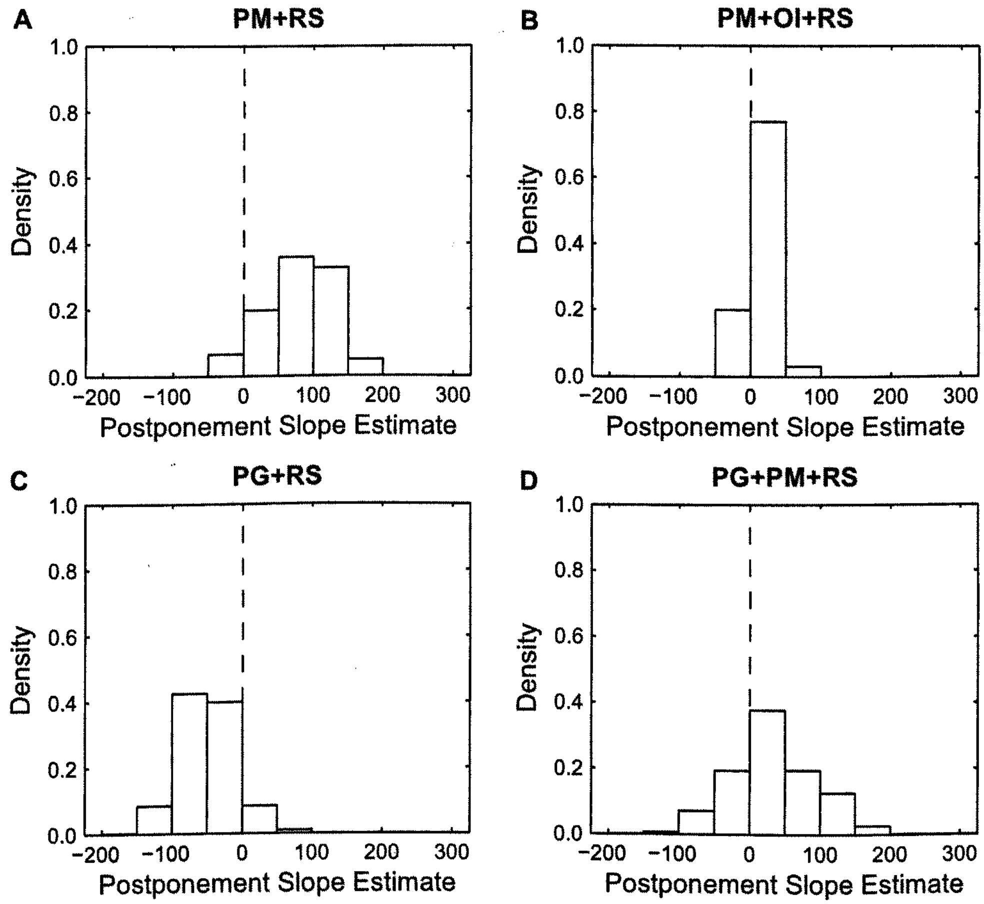

Figure 6-5 *Density histograms showing the distribution of latency-displacement function postponement slopes predicted by four models of serial order for the second parameter space analysis. Panels show predictions for the PM+RS model (A), the PM+OI+RS model (B), the PG+RS model (C), and the PG+PM+RS model (D).*

Thus, with longer sequences of nine-items the predictions of the PM+OI+RS model are essentially comparable to those generated with shorter sequences of six-items.

Considering now the PG+RS model (Figure 6-5C), consistent with the outcomes of the previous analysis this model once again predicted a majority of negative postponement slopes (89% of model predictions). However, relative to the results of the first analysis, the model predicts a larger proportion of shallow negative slopes falling within the range of -100ms to 0ms (81% of model predictions, compared to 43% of model predictions, respectively) and a smaller proportion of steep negative slopes falling within the range of -200ms to -100ms (8% of model predictions, compared to 50% of model predictions, respectively). As previous, the model predicts a small minority of shallow positive slopes falling within the range of 0ms to 100ms (9% of model predictions). These results show once again that the PG+RS model predicts a majority of negative postponement slopes, but differ from the results of the first analysis in showing that with longer sequences there is a greater tendency to produce shallow negative postponement slopes.

Finally, the PG+PM+RS model (Figure 6-5D) once again predicted a large proportion of shallow negative and positive postponement slopes falling within the range of -100ms to 100ms. Indeed, the proportion of simulations returning slopes within this range (84% of model predictions) was distinctly higher than that observed in the first analysis (60% of model predictions). Of these shallow slopes, 27% were negative slopes and 57% were positive slopes. The rest of the models predictions comprised a small number of steep negative slopes falling within the range of -200ms to -100ms (8% of model predictions) and a slightly larger number of steep positive postponement slopes falling within the range of 100ms to 200ms (15% of model predicitons). Thus, these results are even more reassuring than the results obtained from the first analysis. They clearly demonstrate that with longer sequences, the PG+PM+RS model predicts an overwhelming majority of shallow negative and positive postponement slopes.

Overall, the results of this second wave of analyses harmonize well with those of the first analysis. They confirm that with longer sequences, the PM+RS and PM+OI+RS models continue to predict a majority of positive postponement slopes, and that the PG+RS model continues to predict a majority of negative postponement slopes. The main change in the predictions of the PM+RS and PG+RS models was a tendency to produce shallower slopes than was the case for six-item sequences. The predictions of the PG+PM+RS model confirm that with longer sequences, shallow postponement slopes are a truly central feature of the model's behaviour.

General discussion

The preceding chapters have provided evidence for the role of a primacy gradient of activation, positional marking, and response suppression in verbal, visual, and spatial short-term order memory. The evidence for these three explanatory principles comes from applications of the different models to the verbal serial recall data of Farrell and Lewandowksy (2004), the visual serial reconstruction data of Experiment 2, and the spatial serial reconstruction data of Experiment 9. These model fitting exercises confirmed that only the PG+PM+RS model can accommodate the empirical pattern of the LDFs observed in the three domains when model parameters are estimated directly from the data.

However, as noted in the introduction, one shortcoming of this evidence is that it is based upon local predictions of the models derived from single parameter values. The purpose of the parameter space sensitivity analyses reported in this chapter was to examine the predictions of the models across a wide range of their parameter space, to verify whether or not the predictions of the models observed under their best fitting parameter values are representative of their more general behaviour. The analyses therefore sought to elucidate whether the predictions of the models, in particular the PG+PM+RS model, are sufficently robust to be attributed to their core representational principles.

The main outcomes of the parameter space sensitivity analyses are straightforward and can be summarised as follows. For the first wave of analyses involving short sequences of six-items the PM+RS and PM+OI+RS models both predicted an overwhelming majority of positive postponement slopes. The former model predicted a majority of steep positive slopes, consistent with the local predictions of the model observed in previous chapters. In contrast, the latter model predicted a majority of shallow positive slopes, which is contrary to the local predictions of this model observed in previous chapters in which it predicted a steep positive postponement slope. However, this outcome is readily explained by the fact that the settings of the output interference parameter were very low in the simulations of previous chapters, whereas in the current simulations the values of this parameter covered a broad range, which included moderate and high levels of output interference. The PG+RS model predicted an overwhelming majority of negative postponement slopes, consistent with the local predictions of this model witnessed in previous chapters. Of chief interest, the PG+PM+RS model predicted a large proportion of shallow negative and positive postponement slopes, consistent with the distribution of postponement slopes predicted by this model across fits to individual participant data. However, shallow slopes were not a unanimous prediction of this model, as it additionally predicted a moderate proportion of steep positive slopes.

For the second wave of analyses involving longer sequences of nine-items the outcomes were broadly the same, but with some noteworthy exceptions. The PM+RS and PM+OI+RS models once again predicted a majority of postive postponement slopes, however, relative to the first analysis the PM+RS model did not predict as many steep positive slopes, but predicted instead a greater proportion of shallow positive slopes. The PG+RS model once again predicted a majority of negative slopes, but relative to the first analysis predicted fewer steep negative slopes and a greater proportion of shallow negative slopes. Crticially, the PG+PM+RS model once again predicted a large proportion of shallow negative and positive postponements slopes. However, the proportion of shallow slopes was distinctly greater than that observed in the first analysis, with the implication that the model predicted only a very small proportion of steep slopes.

The outcomes of the current set of parameter space sensitivity analyses both replicate and extend those of Farrell and Lewandowsky (2004). They replicate the work of these authors in demonstrating that any involvement of postional marking in the absence of a primacy gradient predicts an overwhelming majority of positive postponement slopes, whereas a primacy gradient in conjunction with response suppression predicts a majority of negative postponement slopes. The analyses extend the work of these authors by bringing the PG+PM+RS model into the model comparisons, and in demonstrating that this model predicts a large proportion of shallow negative and postive postponement slopes. Taken overall, the modelling outcomes confirm that the four models exhibit a robust set of predictions indicating that they are not unduly flexible. With the exception of the PM+OI+RS model (see comments above), the modelling outcomes additionally confirm that the local predictions of the PM+RS, PG+RS, and PG+PM+RS models witnessed in previous chapters are central features of their general behaviour, and can thus be attributted to their core representational principles. Critically, these results suggest that the PG+PM+RS model is not an overly complex model.

Before summarising, one final issue deserves discussion. This concerns the relationship between the distributions of postponement slope estimates observed empirically across individual participants for the visual and spatial tasks, and the distribution of slope estimates predicted by the PG+PM+RS model. Looking back at the empirical distribution of postponement slope estimates for the visual (Figure 6-1A) and spatial (Figure 6-1B) data presented at the outset, it is apparent that in the case of the former data there is a distinctly greater proportion of shallow positive than negative slopes, whilst in the case of the latter data there is an approximately equal distribution of shallow negative and positive slopes. The distribution of slopes predicted by the PG+PM+RS model in the simulations reported here is most compatible with the visual data: the model exhibits a greater proportion of shallow positive than negative slopes. One might wonder whether this is an indication that the PG+PM+RS model is a less appropriate model of the spatial data than the visual data. However, this disparity can be explained by assuming that the attentional weight given to the primacy and positional dimensions of ordering is approximately equal in the case of the visual data (the weighting parameter was set to .5 for the parameter space sensitivity analyses reported here), but slightly skewed in the case of the spatial data so that more attention is allocated to the primacy than the positional dimension of ordering (this differential setting of the weighting parameter might arise because positional markers are less effective serial recall cues in the spatial domain, as was suggested in Chapter 5). If the analyses reported here were re-run, but with the attentional weight parameter set in this manner, then the distribution of slopes would more closely resemble the spatial data, with an approximately equal proportion of shallow positive and negative postponement slopes.

Chapter summary

In summary, the current modelling exercises have shown that the local predictions of the models observed in previous chapters are representative of their more general behaviour, and can therefore be attributed to their core representational principles. The only exception to this conclusion concerns the predictions of the PM+OI+RS model, which although robust, did not match those observed in previous chapters. Overall, the outcomes of the parameter space sensitivity analyses indicate that neither of the models examined can be considered overly flexible. Critically, the modelling outcomes indicate that the superior description of the empirical LDFs provided by the PG+PM+RS model in previous chapters does not appear to be attributable to this model being overly complex. This outcome notwithstanding, it is important to try to obtain converging evidence for the representational constructs underlying the PG+PM+RS model. This evidence is sought in Chapter 8 via an empirical and theoretical analysis of fill-in and infill errors in verbal, visual, and spatial serial memory.

7

Modelling grouping effects in verbal and spatial serial memory

Abstract

This chapter presents a single experiment, which directly compared the effects of temporal grouping on a verbal and a spatial serial memory task using a common recall method. A reliable increase in the probability of interpositions for temporally grouped relative to ungrouped sequences was observed for the verbal, but not the spatial task. This outcome suggests that the failure to observe an increase in interpositions for grouped spatial sequences in Experiments 7 and 9 is not due to the use of a recall method different to that employed in studies of grouping using verbal stimuli. Subsequent computational modelling work evaluates the hypothesis that this discrepancy is due to different representations of the positions of items in grouped verbal and spatial sequences, with verbal items being coded for their positions within groups, and spatial items being coded for their positions within the sequence overall. Applications of models to the data failed to support this hypothesis and instead buttress the notion that position within group representations are a common feature of both verbal and spatial short-term memory.

Introduction

Experiments 7 and 9 of Chapter 5 demonstrated effects of temporal grouping employing spatial stimuli similar to those documented using verbal materials. Relative to an ungrouped baseline, temporal grouping engendered an elevation in spatial serial memory performance, caused a scalloping of the accuracy and response latency serial position curves, as well as a reduction in the tendency for items to transpose between groups. Nevertheless, contrary to studies of temporal grouping using verbal material these experiments failed to detect an increase in the probability of interposition errors for temporally grouped relative to ungrouped spatial sequences. This discrepancy is noteworthy, because interpositions are a major hallmark of grouping in verbal short-term memory and constitute the evidence *par excellence* for positional marking (Henson, 1996, 1998a, 1999).

Two hypotheses were advanced to explain this discrepancy between the effects of grouping in the verbal and spatial domains. The first hypothesis is that it is an artefact of differences in the recall methodology employed in verbal studies and the experiments of Chapter 5 using spatial stimuli. To my knowledge, all verbal studies of grouping that have considered interposition errors (e.g., Farrell & Lelievre, 2009; Henson, 1996, 1999; Ng & Maybery 2002; Ryan, 1969a) have used serial recall, whereas the experiments employing spatial stimuli reported in Chapter 5 used serial reconstruction of order. The critical difference between these two recall methods is that the former method requires the retrieval of both item and order information, whilst the latter method requires the retrieval of order information only. It follows that if interpositions require the recall of item information for their expression then their absence in Experiments 7 and 9 can be traced to the use of a task that emphasises the retrieval of order, but not item information. The second hypothesis is that the discrepancy reflects a fundamental difference in the nature of the positional representations underlying temporally grouped verbal and spatial sequences. Specifically, it was suggested that in both domains order information is represented on at least two dimensions, one of which corresponds to a coarse representation of the positions of groups in sequence. However, it was argued that the increase in the probability of interpositions for grouped verbal sequences, implies that for these sequences the second dimension of ordering represents the positions of items within-groups, whereas the absence of an increase in the probability of interpositions for grouped spatial sequences, implies that for these sequences items are coded for their positions in the sequence overall.

The purpose of the current chapter is to evaluate these two competing hypotheses. In service of this goal, a single experiment is presented contrasting the effects of temporal grouping on a verbal and a spatial serial memory task when the same recall method – serial reconstruction – was employed. To anticipate the results of the experiment, common effects of grouping were observed across the verbal and spatial tasks, including an elevation in recall accuracy, a scalloping of the accuracy and response latency serial position curves, and a reduction in transpositions between groups. Critically however, an increase in the probability of interpositions for grouped relative to ungrouped sequences was observed for the verbal task, but not for the spatial task. Having ruled out the first hypothesis, three computational models that make different assumptions about the positional representations underlying grouped sequences are introduced and applied to the grouped data for the verbal and spatial tasks to test the viability of the second hypothesis.

Experiment 10

The purpose of this experiment was to directly compare the effects of temporal grouping on a verbal and a spatial serial memory task when a common recall method – serial reconstruction – was employed. Two separate groups of participants received either sequences of nine visually presented letters, or nine visually presented spatial locations, for immediate serial reconstruction. Half the sequences were ungrouped and half were organised into three groups of three items by inserting extended temporal pauses after every third item. Following the precedent set by Experiments 7 and 9, it was hypothesised that for the spatial task temporal grouping would not cause an increase in the probability of interpositions. If this is a by-product of the use of a serial reconstruction recall method, then an increase in such errors should also be absent for grouped verbal sequences. However, if a reliable increase in the probability of interpositions is observed for grouped verbal (but not spatial) sequences, then this would be indicative of a fundamental difference in the nature of the grouped representations underpinning verbal and spatial serial memory.

Method

Participants

Thirty-six members of the campus community at the University of York took part in the experiment in exchange for course credit or an honorarium of £3.

Stimuli & Apparatus

The stimuli for the verbal task were sequences containing random orderings of the letters *F, H, J, L, N, Q, R, S, Y*. Each letter was presented visually in the central screen position in black point 18 Arial upper case font on a white background.

The stimuli for the spatial task were sequences containing random orderings of nine visually presented spatial locations. The locations were nine grey icons (measuring 2.5cm x 2.5cm each) arranged haphazardly on a white background. The minimum and maximum distances between pairs of locations (measured from the centre of each icon) were 3cm and 9cm, respectively.

Stimulus presentation and response collection were controlled using software developed by the author running on a Dell Optiplex (Intel Core 2 Duo, 2.13 GHz processor) PC equipped with a 19" monitor and a Razer Copperhead high precision mouse.

Design & procedure

The experiment manipulated two independent variables in a 2 X 2 mixed design. Task (Verbal / Spatial) was a between participants factor, whilst Sequence-Type (Ungrouped / Grouped) was a within-participants factor. Half the participants were assigned to the verbal task, the remaining half to the spatial task. Participants always received ungrouped sequences first, because otherwise participants receiving grouped sequences initially were expected to subjectively group the ungrouped sequences, despite the absence of any objective grouping cues.

Participants initiated each trial by clicking on a 'begin trial' icon located in the central screen position using the computer mouse. A 2000ms delay then ensued before presentation of the sequence during which a central fixation cross was displayed for the verbal task and all locations were simultaneously visible for the spatial task. For ungrouped verbal sequences, each letter was presented singly for 500ms, separated by a 250ms inter-stimulus interval. For ungrouped spatial sequences, each icon illuminated yellow temporarily for 500ms, followed by a 250ms inter-stimulus interval during which all icons remained grey. For grouped verbal and spatial sequences, the inter-stimulus intervals following the third and sixth items were increased to 1000ms to create the impression of three groups of three items.

Following the final item there was a 1000ms blank interval, after which in the verbal task, the letters were simultaneously presented in random positions within a 3 x 3 matrix, and in the spatial task, the icons reappeared in their previous spatial coordinates. Participants were required to click on the letters or icons in their presentation order using the mouse-driven pointer. Once an item was selected its colour changed transitorily to green for 50ms to acknowledge that the computer had registered the response, after which the item could be selected again, meaning that repetition errors were possible. Once nine items had been selected there was a 3000ms inter-trial interval, which was followed by presentation of the 'begin trial' icon for the next trial.

Participants attempted 20 trials for each sequence-type, which were preceded by 2 practice trials. In the spatial task, participants were instructed to encode sequences without deploying supplementary verbal or gestural encoding strategies, and all reported compliance with these instructions. The experiment lasted approximately 30 minutes.

Results

The data were scored using a strict serial reconstruction scoring procedure: an item was only scored as correct if it was reported in the serial position it was presented. The results are organised into three sections: (1) accuracy serial position curves, (2) latency serial position curves, and (3) transposition errors.

Accuracy serial position curves

Serial reconstruction accuracy was higher for the verbal than for the spatial task, the mean proportion of correct responses (averaged across serial positions and sequence-type) being .60 and .53, respectively. Nevertheless, the impact of the grouping manipulation on recall accuracy was similar: the mean proportions of correct responses for ungrouped and grouped sequences were .55 and .66, respectively for the verbal task, and .48 and .58, respectively for the spatial task. Thus, grouping led to an approximately 17% increase in serial reconstruction accuracy in both tasks.

Statistical confirmation of these observations was sought by means of a 2 (Task) X 2 (Sequence-Type) ANOVA performed on the mean proportion of correct responses. This revealed a reliable main effect of Sequence-Type, $F(1, 34) = 30.184$, $MSE = .201$, $p < .001$, confirming that

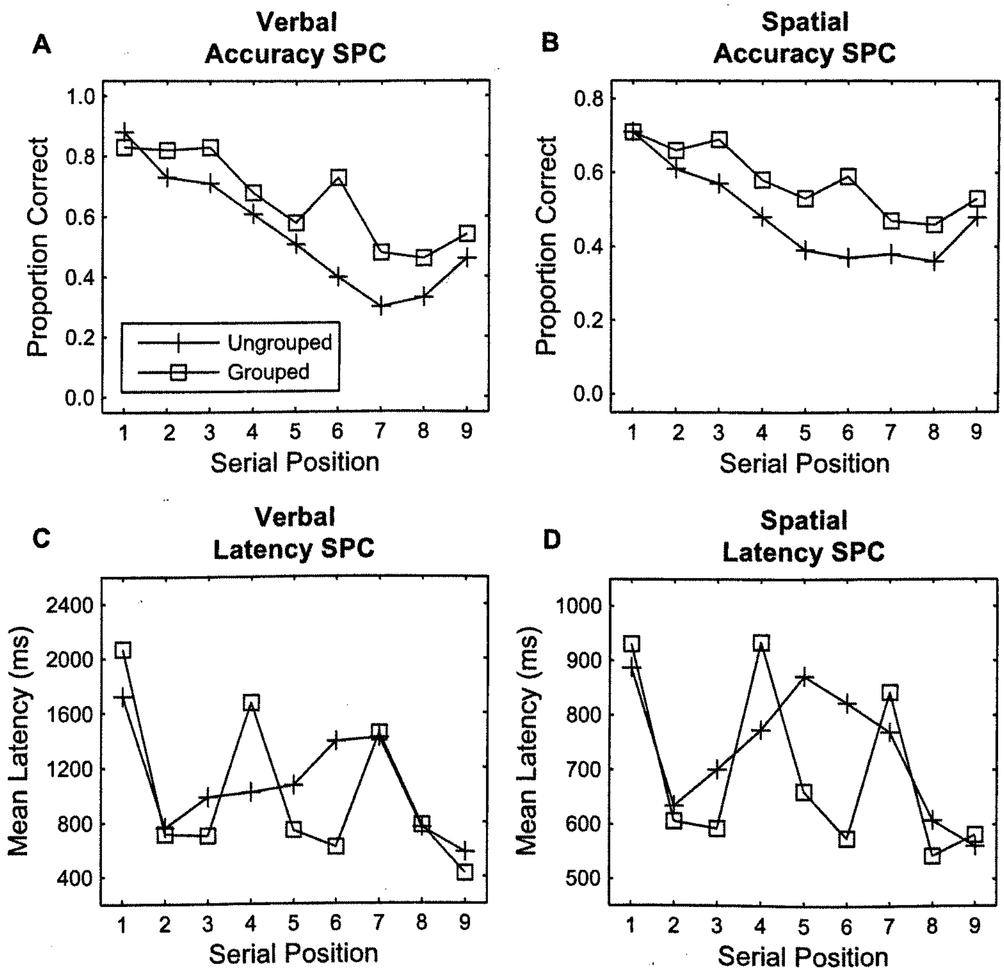

Figure 7-1 *Serial position data for Experiment 10. Panels show the accuracy serial positions curves for the verbal (A) and spatial (B) tasks and the response latency serial position curves for the verbal (C) and spatial (D) tasks. Note—the y-axis scales on the panels showing the data for the verbal and spatial tasks are different.*

grouping enhanced performance, however the main effect of Task did not quite reach conventional significance levels, $F(1, 34) = 3.056$, $MSE = .093$, $p = .089$, indicating that performance in the verbal task did not differ reliably from the spatial task. Finally, the Task X Sequence-Type interaction, $F(1, 34) = .142$, $MSE = .001$, $p = .709$, was non-significant.

The accuracy serial position curves for the verbal task can be inspected in Figure 7-1A, whilst the corresponding data for the spatial task can be seen in Figure 7-1B. Immediately apparent is that in both tasks grouping enhanced performance at all serial positions, except the first position, in addition to causing a scalloping of the serial position curves, as reflected by the emergence of mini within group primacy and recency effects. The serial position curves were analysed separately for the verbal and spatial tasks using 2 (Sequence-Type) X 9 (Serial Position) ANOVAs. For the verbal-task, reliable main effects of Sequence-Type, $F(1, 17) = 202.710$, $MSE = 9.399E7$, $p < .001$, and Serial Position, $F(8, 136) = 6.899$, $MSE = 2628305.028$, $p < .001$, were obtained, as well as a significant interaction between the two factors, $F(8, 136) = 6.895$, $MSE = 2627025.662$, $p < .001$. Similarly, for the spatial task, there were also reliable main effects of Sequence-Type, $F(1, 17) = 345.841$, $MSE = 4.386E7$, $p < .001$, and Serial Position, $F(8, 136) = 4.888$, $MSE = 217180.254$, $p < .001$, and the Sequence-Type X Serial Position interaction also reached significance, $F(8, 136) = 4.886$, $MSE = 217134.801$, $p < .001$.

Latency serial position curves

The latency serial position curves associated with correct responses are shown in Figure 7-1C for the verbal task, and Figure 7-1D for the spatial task. As can be seen from inspection of these figures, all the serial position curves exhibit a long initial latency, but their overall shape varies as a function of sequence-type. Specifically, when the latency for the first position is ignored the serial position curves for ungrouped verbal and spatial sequences exhibit inverted U shaped trends, although the saddle point occurs later for the verbal (serial position 7), than the spatial (serial position 5) task. In contrast, the serial position curves for grouped sequences exhibit pronounced peaks at the first position of each sub-group (positions 4 and 7) with latencies becoming accelerated for subsequent positions within-groups (with the exception of position 9 for spatial sequences).

The response latency serial position curves for the verbal and spatial tasks were analysed separately using 2 (Sequence-Type) X 9 (Serial Position) ANOVAs. For the verbal-task, there were reliable main effects of Sequence-Type, $F(1, 17) = 202.710$, $MSE = 9.399E7$, $p < .001$, and Serial Position, $F(8, 136) = 6.899$, $MSE = 2628305.028$, $p < .001$, as well as a significant interaction between the two factors, $F(8, 136) = 6.895$, $MSE = 2627025.662$, $p < .001$. Likewise, for the spatial task, there were also reliable main effects of Sequence-Type, $F(1, 17) = 345.841$, $MSE = 4.386E7$, $p < .001$, and Serial Position, $F(8, 136) = 4.888$, $MSE = 217180.254$, $p = .001$, and the Sequence-

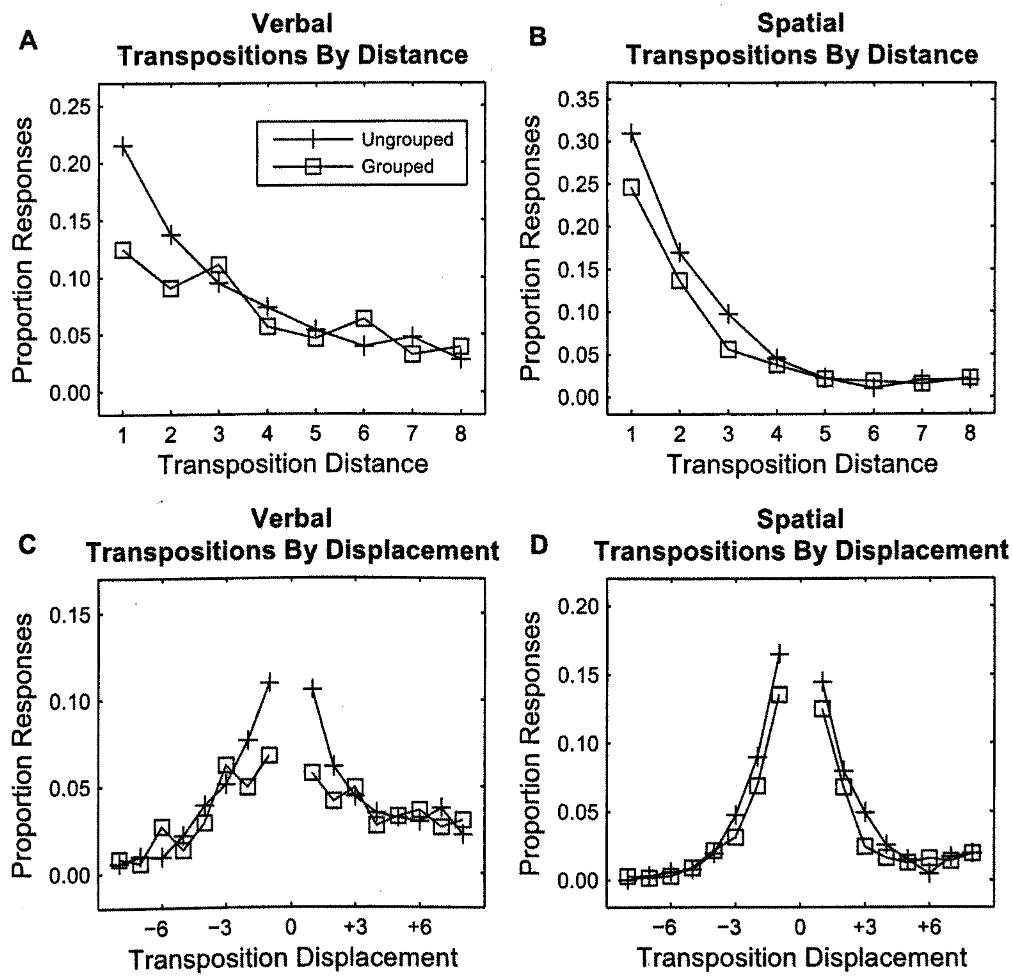

Figure 7-2 *Transposition gradients for Experiment 10. Panels show transpositions as a function of temporal distance for the verbal (A) and spatial (B) tasks and transpositions as a function of transposition displacement for the verbal (C) and spatial (D) tasks. Note—the y-axis scales on the panels showing the data for the verbal and spatial tasks are different.*

Type X Serial Position interaction also reached significance, $F(8, 136) = 4.886$, $MSE = 217134.801$, $p = .001$.

Transposition errors

Turning now to the data of central interest, the proportions of transpositions as a function of temporal distance, which ignores transposition direction, can be inspected in Figure 7-2A for the verbal task, and Figure 7-2B for the spatial task. For the verbal task, the proportion of transpositions for ungrouped sequences decreased with increasing transposition distance, whereas for grouped sequences there are noticeable peaks for transposition distance values three and six, which necessarily represent interposition errors. Also apparent is that the benefit of grouping was

Sequence-Type	Within Groups	Between Groups	
		Interpositions	Other
Verbal-Task			
Ungrouped	0.43 (.07)	0.17 (.05)	0.40 (.08)
Grouped	0.45 (.14)	0.29 (.10)	0.27 (.10)
Spatial-Task			
Ungrouped	0.55 (.07)	0.13 (.05)	0.32 (.05)
Grouped	0.71 (.09)	0.11 (.05)	0.18 (.05)

Table 7-1 *Proportion of transpositions within and between groups for Experiment 10. Standard deviations are shown in parentheses.*

due to a reduction in transpositions at distance values one, two, and four. For the spatial task, the proportions of transpositions decreased with increasing transposition distance, for both ungrouped and grouped sequences. There was no peak in the transposition gradient for grouped sequences at distance value three, but the proportion of transpositions at distance value six is slightly higher than for ungrouped sequences. It is evident that the benefit of grouping was due to a decrease in transpositions at distance values one to three. To take transposition direction into account, Figure 7-2C shows the proportions of transpositions this time as a function of transposition displacement for the verbal task, whilst the corresponding data for the spatial task is shown in Figure 7-2D. For the verbal task, an increase in interpositions for grouped compared to ungrouped sequences is apparent for -6, -3, +3, and +6 transpositions, whereas for the spatial task the only evidence for an increase in interpositions is an elevation for +6 transpositions.

To scrutinize the errors in further detail, Table 7-1 shows the proportions of transpositions within and between groups for the different conditions, with transpositions between groups being further sub-divided into interpositions and other (position non-preserving) errors. It is apparent that for the verbal task, grouping had a negligible impact on the incidence of transpositions within groups, but simultaneously increased the incidence of interpositions and decreased the incidence of other transpositions between groups. In contrast, for the spatial task, grouping caused an increase in the incidence of transpositions within groups, a minor reduction in the incidence of interpositions, and an increase in the incidence of other transpositions between groups. To obtain statistical confirmation of these trends the incidence of each error type was compared between ungrouped and grouped sequences, using t-tests performed on the log-odds transformed error proportions. To compensate for inflated type-1 error rates a more conservative alpha level (α = .01) was employed. For the verbal task, grouping had no effect on the incidence of transpositions within groups, $t(17) = .510$, $p = .616$, but significantly increased the incidence of interpositions, $t(17) = 5.093$, $p < .001$, and significantly decreased the incidence of other transpositions between groups, $t(17) = 5.575$, $p < .001$. For the spatial task, grouping significantly increased the incidence of transpositions within groups, $t(17) = 6.234$, $p < .001$, yet did not modify the incidence of interpositions, $t(17) = 12.567$, $p = .520$. However, grouping did significantly decrease the incidence of other transpositions between groups, $t(17) = 6.681$, $p < .001$.

Discussion

In brief, temporal grouping exerted both similar and dissimilar effects on verbal and spatial serial memory performance. The similarities included an elevation in recall accuracy, a scalloping of the accuracy and response latency serial position curves, as well as a reduction in transpositions between groups. The main dissimilarity, and the chief empirical outcome, was that grouping fostered a reliable increase in the probability of interposition errors for verbal sequences, in accordance with previous studies (Farrell & Lewandowsky, 2004; Farrell & Lelievre, 2009; Henson, 1999; Maybery et al., 2002; Parmentier & Maybery, 2009; Ryan, 1969a, b), but not for spatial sequences, consistent with Experiments 7 and 9, as well as Experiment 3 of Parmentier et al. (2006). A further dissimilarity was that grouping did not increase the proportions of errors within groups as much for verbal as for spatial sequences, although this discrepancy is likely due to grouping having increased the proportion of interpositions for verbal sequences, which implies a corresponding reduction in opportunities for errors within groups.

That a reliable increase in the probability of interpositions was observed for grouped verbal sequences indicates that the failure to detect an increase in the probability of such errors for grouped spatial sequences, here and in Experiments 7 and 9, is not a consequence of the recall method employed. On the contrary, this outcome lends credibility to the hypothesis that there is a subtle, yet fundamental difference in the positional representations underlying grouped verbal and spatial sequences. In the next section, computational modelling work is presented that seeks to elucidate the nature of this difference.

Computational modelling

The aim of this section is to use quantitative model comparisons to determine how the positional representations underlying grouped verbal and spatial sequences vary. As noted earlier, the working hypothesis is that in both domains one dimension of ordering represents the positions of groups in sequence and this information is supplemented by a second dimension of ordering, which in the case of verbal sequences represents the positions of items within groups, and in the case of spatial sequences represents the positions of items in the sequence overall. To evaluate this hypothesis a modelling framework was employed that enabled the simple derivation of predictions for serial recall of grouped sequences: the Start-End Model (SEM; Henson, 1998). The SEM constitutes a specific instantiation of a position marking model. The benefit of using this model is that it can be easily elaborated to incorporate multidimensional positional representations. For example, items can be associated with markers that encode the positions items within groups (or in the sequence overall), as well as markers that encode the positions of groups in sequence[1].

The Start-End Model (SEM), is a model of verbal serial recall whose namesake originates from its proposition that order information is encoded by associating items to two sets of positional markers, a start marker that is strongest upon presentation of the first item and decreases gradually in strength with each subsequent item, and an end marker that is weakest for the first item and increases gradually in strength with each subsequent item. In SEM items are stored in memory as episodic tokens, which contain information about the identity of items and their positions relative

[1] Technically, the above hypotheses could have been investigated by elaborating the PG+PM+RS model employed in previous chapters. I chose to employ the SEM here largely as a matter of convenience, since at the time of writing this chapter I was using the model to address grouping data from Hurlstone (2006).

to the start and end of the sequence. Each token is specified as a two-element vector, with one element representing the value of the start marker the item was associated with, and the second element representing the value of the end marker the item was associated with. Recall involves probing the episodic tokens in parallel with a sequence of position markers, each of which is a two element vector combining the start and end marker values for a specific serial position. For each position, the overlap between the corresponding position marker and the positional information stored in each of the episodic tokens is calculated in parallel, and the item with the strongest overlap is selected for recall. Thus, like the generic architecture presented in Chapter 2, recall in the SEM involves a competitive queuing selection process.

In the modelling reported here the SEM is employed as a generic framework within which the predictions of three models that make different assumptions about the positional representations underlying grouped sequences can be compared. The models are evaluated by applying them to the grouped verbal and spatial sequence conditions of Experiment 10. In one of the models, dubbed GP-WP (Group Position – Within Group Position), items are associated with markers that encode the positions of groups in sequence, as well as markers that encode the positions of items within groups. In the second model, dubbed GP-SP (Group Position – Sequence Position), items are also associated with markers that encode the positions of groups, but the second set of markers encode the positions of items in the sequence overall. In a third model, dubbed SP (Sequence Position), items are associated with a single group marker that encodes the sequence as a single group, combined with markers that encode the positions of items in the sequence overall. This model assumes that positional information is organised along a single dimension that is effectively insensitive to grouping, and is included as a baseline model to ascertain whether the use of multidimensional positional representations confers an improvement in descriptive accuracy over a one-dimensional positional representation.

In the modelling that follows a stripped down implementation of the SEM was employed, in which positional information was coded with respect to the start, but not the end of the sequence. The reason for the omission of end positional marking is that its inclusion in the SEM was motivated principally to explain the propensity for interpositions in irregularly grouped verbal sequences to maintain relative, as opposed to absolute within group position. However, for the target data modelled here absolute and relative interpositions cannot readily be distinguished. Moreover, interest centres here on contrasting models principally in terms of whether or not they predict interpositions. That is, the modelling is agnostic concerning whether the errors in question stem from absolute or relative representations, or a combination of the two. Furthermore, recent model comparison work using the SEM (Farrell and Lelievre, 2009), suggests that at least for serial recall of grouped verbal sequences, there is little empirical support for a continuous end positional marker (though there is evidence for a restricted end marker that is activated following the final item of each sequence or sub-sequence). On the basis of these considerations, the inclusion of end positional marking here is arguably an encumbrance unnecessary and its omission reduces the number of free model parameters required.

Model implementation

It is assumed that presentation of a sequence of items for recall results in the creation of a set of unordered tokens in memory. Each token contains information about the item presented, as well as its position in the sequence. The position of an item is given by two exponentially decreasing positional markers, an item marker and a group marker, that respectively represent the positions of items within groups (or within the sequence as a whole), and the positions of groups in sequence. The values of the item marker I for each position j were determined by:

$$I(j) = \alpha^{j-1} \tag{7-1}$$

Where α is a free parameter ($0 < \alpha < 1$) that determines the distinctiveness of the item marker values. For the GP-WP model, the item marker encoded the positions of items within each group (j = 1..3, for each of the three groups), whereas in the GP-SP and SP models it encoded the positions of items in the sequence overall (j = 1...9). The values of the corresponding group marker G representing the position of each group l within the sequence were given by:

$$G(l) = \beta^{l-1} \tag{7-2}$$

Where β is a free parameter ($0 < \beta < 1$) in the GP-SP and GP-WP models, which governed the distinctiveness of the group marker values. In the SP model β was set to a constant value (equal to 1), which resulted in the sequence being encoded as a single group that did not differentiate sequence items.

Recalling a sequence involved reinstating the position markers for each position to probe for a response. For each position, the overlap between the position marker being used to cue recall and the positional information stored in each of the episodic tokens was computed in parallel and used to determine the strength with which items competed for recall in response to the cue. For the item markers, the overlap between the item marker being used to cue recall j and the corresponding positional information stored in each token i in memory $O_I(j,i)$ was determined by:

$$O_I(j,i) = \sqrt{I(j)I(i)}\exp(-|I(j)-I(i)|) \qquad (7\text{-}3)^2$$

Which is a simplified version of equation (2) presented in Henson (1998a). In a similar fashion, the overlap between the group marker being used to cue recall and the corresponding positional information stored in each token $O_G(j,i)$ was given by Equation (7-3), but substituting values of $I(j)$ and $I(i)$ with values of $G(j)$ and $G(i)$, respectively:

$$O_G(j,i) = \sqrt{G(j)G(i)}\exp(-|G(j)-G(i)|)$$

Finally, the overall strength with which each item competed for selection $C(i)$ in response to the item and group position markers for a given position j was determined by:

$$C(i) = O_I(j,i)O_G(j,i)(1-r(i)) + n(0,G_c) \qquad (7\text{-}4)$$

[2] Note that the second term in equation 7-3 is the Euclidean metric of similarity between two vectors, which is sharpened by the exponent. The square root pre-multiplier is incorporated to lower and widen the positional confusability functions at medial serial positions making errors more likely at these positions than terminal positions.

Where $r(i)$ represents the suppression of item i once recalled, and n represents random noise drawn from a Gaussian distribution with a mean of 0 and standard deviation given by the free parameter G_c ($0 < G_c < 1$). The amount of response suppression r was set to a constant value (equal to .95) in order to minimize the number of free model parameters during fitting.

To summarize, the free parameters for the GP-SP and GP-WP models were the distinctiveness of the item (α) and group (β) markers, and the standard deviation of noise (G_c). The SP model took as its free parameters the distinctiveness of the item marker (α), and the standard deviation of noise (G_c). Thus, the GP-SP and GP-WP models incorporated three free parameters, whereas the SP model incorporated two.

Fitting procedure

Each model was fitted to the transposition matrix of individual participants for the grouped verbal and spatial sequence conditions of Experiment 10. In a transposition matrix, the row vectors represent output serial positions, whilst the column vectors represent input serial positions. The values within the matrix represent the frequencies with which items from different input positions were recalled at each output position.

The models were fit to the data using Maximum Likelihood Estimation; that is parameter values were sought that maximized the likelihood that a model produced the data if it were correct. The likelihood of the observed response frequencies for each output position (row vector) given the corresponding vector of response probabilities predicted by the model being fit under a specific combination of parameter values was calculated using the multinomial log-likelihood function (equation 2-10 in Chapter 2). A likelihood estimate for the entire matrix of response frequencies was then obtained by summing the multinomial log-likelihood estimates for each output position, in order to produce a joint multinomial log-likelihood estimate.

Parameter estimates were obtained using the simplex algorithm (Nelder & Mead, 1965) minimizing the negative joint multinomial log-likelihood estimate (note that minimizing the negative log-likelihood is equivalent to maximizing likelihood). Each parameter vector explored by the search algorithm involved 50,000 model simulation trials. To help prevent the algorithm from converging on local maxima the search process was conducted multiple times employing different starting parameter values. These were chosen by defining three values for each free model parameter before factorially crossing these to create a grid of starting points.

To help facilitate model comparisons, in addition to reporting the maximum log-likelihood estimate $\ln L$ of each model, two generalizability measures are also incorporated: the Akaike Information Criterion (AIC; Akaike, 1973) and the Bayesian Information Criterion (BIC; Schwarz, 1978). These information criteria (described in Chapter 2) supplement a model's $\ln L$ with a complexity term that penalizes models in terms of their number of free parameters (in the case of AIC and BIC), as well as the sampling variability of estimated parameters (in the case of BIC). The use of these information criteria is relevant in the context of comparing the descriptive accuracy of the GP-WP and GP-SP models, with reference to the SP baseline model, since the former two models incorporate an additional free parameter. Recall from Chapter 2 that a problem when comparing AIC and BIC scores is that it is difficult to determine whether differences in these values are meaningful, because they are not organised along a continuous measure of evidence. Accordingly, the AIC and BIC scores were converted into AIC weights (ωAIC) and BIC weights (ωBIC), which can be interpreted as the conditional probabilities that each model should be chosen, given the cohort of competitor models and the data (see Chapter 2 for further details).

Model predictions

The goodness-of-fits of the models to the transposition matrices of individual participants for the verbal and spatial data can be scrutinized in Appendix 1. These are summarised in Table 7-2, which provides the goodness-of-fits averaged across individual participants. Considering first the fits to the verbal data, it can be seen from inspection of this table that as anticipated the GP-WP model obtained the highest mean $\ln L$ estimate, followed by the GP-SP and SP models. Since the former two models incorporate an extra free parameter, it is important to contrast the fits of these models, relative to the SP model, with reference to the model AIC and BIC scores (*note*—when comparing models using these information criteria the model with the smallest value is preferred).

Model	k	ln$L$	AIC	ωAIC	BIC	ωBIC
Verbal						
SP	2	-270.28	544.57	0.00	549.36	0.00
GP-SP	3	-259.48	524.96	0.00	532.14	0.00
GP-WP	3	-230.58	467.15	1.00	474.33	1.00
Spatial						
SP	2	-256.17	516.33	0.00	521.12	0.00
GP-SP	3	-249.66	505.32	0.06	512.50	0.06
GP-WP	3	-239.17	484.35	0.94	491.53	0.94

Table 7-2 *Average goodness-of-fits of the models to the transposition matrices for the grouped verbal and spatial sequence conditions of Experiment 10. Note—k = number of free model parameters; lnL = maximum log-likelihood; AIC = Akaike Information Criterion; ωAIC = AIC weight; BIC = Bayesian Information Criterion; ωBIC = BIC weight.*

It can be seen that the AIC and BIC scores are lower for the GP-WP and GP-SP models than the SP model, indicating that the additional free parameter in these models (β) does not simply increase their ability to fit noise. The mean AIC and BIC weights for the GP-WP model were both equal to 1 (and necessarily equal to 0 for the GP-SP and SP models), which amounts to very strong evidence in favour of this model. Considering now the goodness-of-fits for the spatial data, unexpectedly, the GP-WP model once again obtained the highest mean ln$L$ followed by the GP-SP and SP models. The better fits of the former two models, relative to the SP model, were preserved after the model ln$L$ values were transformed to AIC and BIC scores. The mean AIC and BIC weights for the GP-WP model were both equal to .94 (equal to .06 for the GP-SP model and necessarily 0 for the SP model), which again amounts to very strong evidence in favour of this model.

In summary, contrary to the hypothesis advanced at the outset, the results of the model comparisons indicate that the GP-WP model is the preferred model – of the competitor models contrasted – for both the verbal and the spatial data. This is further underscored by the fact that this model provided the best fit to the data of every single participant. I now consider the predictions of the models under their best fitting parameter settings. Although the models were fit to the

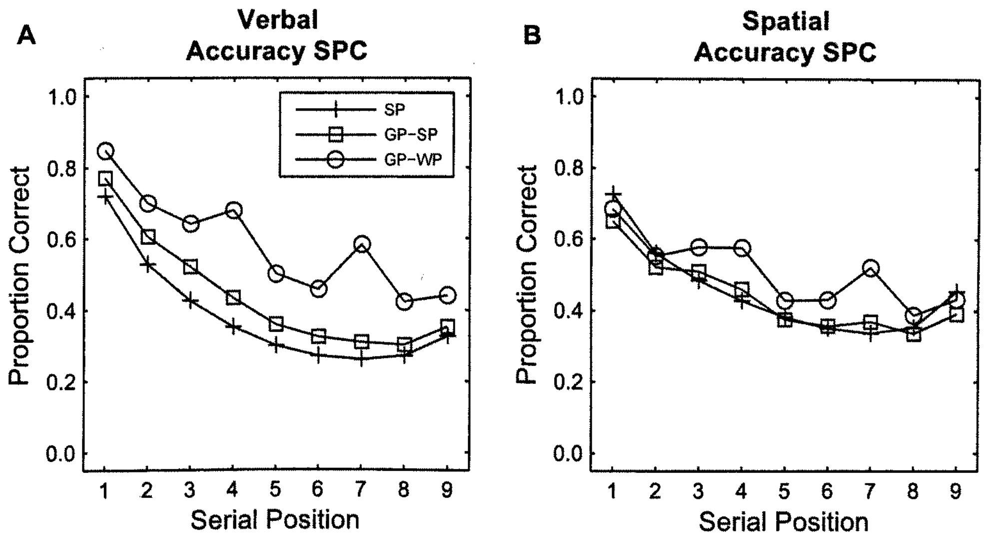

Figure 7-3 *Fits of the models to the accuracy serial position curves for the grouped sequence conditions of Experiment 10. Panels show fits to the verbal (A) and spatial (B) data.*

transposition matrices the predictions of the models for these data are not presented, because it is difficult to visualize discrepancies between transposition matrices, given the large number of data points involved (equal to 81). Accordingly, the model predictions are summarized more parsimoniously in terms of serial position curves and transposition gradients, as for the data of Experiment 10.

Figure 7-3 shows the fits of the models, averaged across participants, to the accuracy serial position curves for the verbal and spatial data. Considering first the fits of the models to the verbal data (Figure 7-3A), it is apparent that the GP-WP model predicts some scalloping of the serial position curve, consistent with the data (Figure 7-1A), but it under-predicts performance at the second and third positions and fails to predict a within group recency effect at position six. The GP-SP model on the other hand did not generate any scalloping what so ever, predicting instead a monotonically decreasing serial position curve with a one-item recency effect, as predicted by the SP model. It is apparent however, that the serial position curve for the former model sits slightly higher than that of the latter model, yielding a smaller discrepancy between observed and predicted values. Turning to the fits to the spatial data (Figure 7-3B), it can be seen that the GP-WP model predicts a scalloped serial position curve, consistent with the data (Figure 7-1B), although it under predicts the extent of within group recency at position six. The GP-SP model predicts some

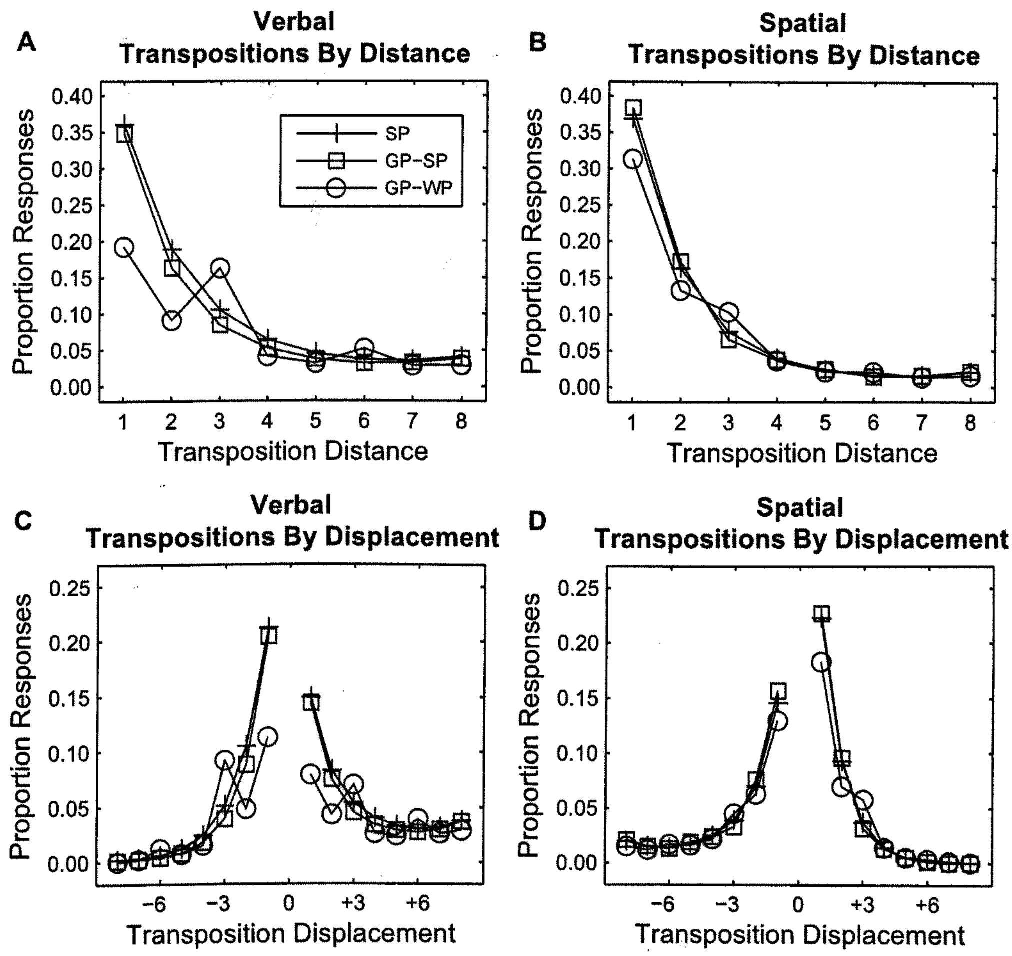

Figure 7-4 *Fits of the models to the transposition gradients of the grouped sequence conditions of Experiment 10. Panels show transpositions as a function of temporal distance after fitting to the verbal (A) and spatial (B) data, as well as transpositions as a function of transposition displacement after fitting to the verbal (C) and spatial (D) data.*

scalloping, but these effects are not as prominent as observed empirically, whilst the SP model predicts a monotonically decreasing serial position curve with an upturn at the final position.

Figure 7-4 shows the transposition distance and displacement gradients predicted by the models, averaged across participants, for the verbal and spatial data. Considering first the predictions for transpositions as a function of temporal distance, it can be seen that for the fits to the verbal data (Figure 7-4A) the GP-WP model predicts a pronounced peak at transposition distance three and a less marked peak at transposition distance six, consistent with the data (Figure 7-2A), whereas the GP-SP and SP models both predict a monotonic decrease in transpositions with increasing transposition distance. For the fits to the spatial data (Figure 7-4B), the GP-SP and SP

Model	Within Groups	Between Groups	
		Interpositions	Other
Verbal			
SP	0.51 (.07)	0.13 (.04)	0.37 (.04)
GP-SP	0.56 (.11)	0.11 (.05)	0.34 (.06)
GP-WP	0.45 (.12)	0.31 (.10)	0.23 (.09)
Spatial			
SP	0.58 (.06)	0.10 (.04)	0.33 (.03)
GP-SP	0.70 (.08)	0.08 (.03)	0.23 (.05)
GP-WP	0.69 (.08)	0.15 (.04)	0.16 (.05)

Table 7-3 *Proportion of transpositions within and between groups predicted by the models after fitting to grouped sequence conditions of Experiment 10. Standard deviations are shown in parentheses.*

models both correctly predict a monotonic decrease in transpositions, whereas the GP-WP model predicts a noticeable increase in interpositions at transposition distance three relative to the data (Figure 7-2B). Turning to the predictions for transpositions when plotted as a function of transposition displacement, it is apparent that for the fits to the verbal data (Figure 7-4C) the GP-WP model predicts noticeable peaks for -6, -3, +3 and +6 transposition displacement values, as observed empirically (Figure 7-2C). For the fits to the spatial data (Figure 7-4D), the GP-WP model predicts a small elevation for +3 displacements and an even smaller elevation for -3 displacements, which are not seen empirically (Figure 7-2D).

The proportions of transpositions within and between groups predicted by the models, averaged across participants, are given in Table 7-3 and can be compared with the data shown in Table 7-1. For the verbal data, the GP-WP model provided excellent fits to the distributions of errors, whereas the GP-SP and SP models over predicted the incidence of transpositions within groups and other transpositions between groups, whilst at the same time grossly under predicting the incidence of interpositions. For the spatial data, the GP-WP and GP-SP models both provided excellent fits to the distributions of errors. However, the GP-WP model predicted a 4% increase in the incidence of interpositions relative to the data (15% predicted, compared to 11% observed). Unsurprisingly, the

SP model provided a poor fit to the distributions of transpositions. This model under predicted the incidence of transpositions within groups and over predicted the incidence of transpositions between groups.

The results of the model fitting exercise present a conundrum. As anticipated, the GP-WP model provided the best account of the verbal data, but contrary to expectations the same model also provided the best account of the spatial data. Nevertheless, this model predicted an increase in interpositions relative to the data. Thus, although this model provided a better quantitative fit to the data than its rivals, it also predicts a data pattern not seen empirically. An additional surprise finding was the poor fit of the GP-LP model to the accuracy serial position curves, which were devoid of the within group primacy and recency effects seen in the data. The theoretical implications of these modelling outcomes are considered in the following section.

General discussion

The main empirical contribution of this chapter has been to show that the failure to observe an increase in the probability of interpositions for temporally grouped relative to ungrouped spatial sequences in Experiments 7 and 9 was not the consequence of the use of a recall method different to that employed in verbal studies of temporal grouping. A direct comparison of the effects of temporal grouping on a verbal and a spatial serial memory task employing a common recall method – serial reconstruction – revealed an increase in the incidence of interpositions for grouped sequences for the verbal task, but echoing the results of Chapter 5 no such increase was observed for the spatial task. The aim of the subsequent model comparison work was to test the hypothesis that this discrepancy between the effects of grouping in the two domains reflects a difference in the positional representations underlying grouped verbal and spatial sequences. Specifically, it was hypothesised that in both domains one dimension of ordering represents the positions of groups in sequence, and that this information is supplemented by a second dimension of ordering, which in the case of verbal sequences represents the positions of items within groups, and in the case of spatial sequences represents the positions of items in the sequence overall.

The results of the quantitative model fitting exercise did not lend support for this hypothesis. On the contrary, it was found that the GP-WP model provided the best fit to the grouped sequence data for both the verbal and the spatial task. The improvement in fit of this model relative to the GP-SP model was non-trivial, as reflected by the finding that it provided the best fit to the data of every single participant. For grouped verbal sequences, this model predicted the scalloped form of the serial position curve, as well as the local peaks in the transposition gradient at displacements representing interpositions. In contrast, the GP-SP model failed to capture either the scalloping of the serial position curve or the non-monotonicity of the transposition gradient. For grouped spatial sequences, the GP-WP model again provided a good account of the scalloped form of the serial position curve. Nevertheless, it predicted a 4% increase in the incidence of interpositions relative to the data, which manifested as a bump in the transposition gradients for transpositions spanning three positions. Although this model only predicted a relatively small number of interpositions, it is important to note that this outcome constitutes a departure from its more general behaviour. Explorations of the model's parameter space indicates that when the value of the item marker parameter (α) is set to a value of .95 or greater the model predicts a smooth transposition gradient that is not characterised by any discontinuities at transposition distances representing interpositions. However, as the setting of this parameter falls increasingly further below this value the model generates increasingly non-monotonic transposition gradients due to the presence of punctuated peaks at transposition distances reflecting interpositions. It is noteworthy that the average value of this parameter for the GP-WP models fit to the spatial data is equal to .93, which falls just below this cut-off point. Finally, the GP-SP model correctly predicted the negative monotonicity of the transposition gradient, but it failed to produce a sufficiently scalloped serial position curve. It was this shortcoming of the GP-SP model that was the source of its poorer fit to the spatial data.

The absence of marked within group primacy and recency effects in the serial position curve for the GP-SP model is noteworthy given that it is capable of producing such effects provided the value assigned to its group marker parameter (β) is sufficiently high. The problem for this model is that settings of this parameter that foster pronounced within group primacy and recency effects simultaneously engender too many transpositions within groups, and too few transpositions between groups. Under its best fitting parameter settings the GP-SP model gave an excellent fit to the distribution of transpositions for the spatial data predicting that 70% of errors occurred within groups and 30% of transpositions occurred between groups, compared to observed values of 71% and 29%, respectively. The mean setting of the parameter β for these simulations was equal to .93. Although a higher setting of this parameter would have improved the models predicted serial position curve it would also have jeopardised its capacity to fit the distribution of errors within and between groups, which comprise a larger number of data points in the transposition matrix. Interestingly, the average values of the group marker parameter for the GP-SP and GP-WP models were identical, suggesting that the greater scalloping of the serial position curve predicted by the latter model actually arose from its position within group, rather than group position representations.

That the GP-SP model can only produce within group primacy and recency effects at the expense of inflated errors within groups is supported by another model fitting exercise in which this model was fit to the accuracy serial position curves and transposition displacement gradients for the spatial data. The number of data points fitted for each participant in these simulations was 26 (9 points for the serial position curve and 17 points for the transposition gradient), compared with 81 data points in the simulations above. Under these conditions, the GP-SP model predicted a serial position curve characterised by marked primacy and recency effects within each sub-group, whilst also reproducing the general form of the transposition gradient. However, when the proportions of transpositions within and between groups were calculated the model values were 96% and 4% respectively – values grossly at odds with the empirical data. Thus, the cost of emphasising the fit of the model to the serial position curve and a coarse description of the distribution of errors (the transposition gradient) was to compromise the models ability to capture the delicate balance of errors within and between groups.

The above observations compromise the viability of the hypothesis that the difference between grouped representations in verbal and spatial serial memory is the reliance of the latter on position in sequence, rather than position within group representations. Instead, the results of the quantitative model comparisons favour the view that position within group representations are a feature of both verbal and spatial short-term memory. However, it must be borne in mind that the poor account of the data provided by the GP-SP model might not be representative of this general class of model. Instead, it might be specific to the implementation employed here based on a one-dimensional context representation of group and sequence position[3]. It is possible that a version of the GP-SP model using a two-dimensional context signal, in which each group and item is coded for its position relative to both the start and end of a sequence – by incorporating start and end positional markers – might fare better. In particular, the coding of the positions of groups relative to the end of a sequence might render the model better able to predict the within group recency effects that the current instantiation of the model has struggled to accommodate. Future modelling work will have to explore this possibility to decisively rule out the GP-SP model.

Nevertheless, on the basis of the current modelling at least, the GP-WP model is clearly the preferred model of the spatial data. Of course, the drawback to this conclusion is that the GP-WP model predicted an increase in interpositions that was not observed empirically after it was fit to the spatial data. One interpretation of this shortcoming is that the GP-WP model is incomplete and needs to be supplemented with further ancillary assumptions that have a bearing on the incidence of interpositions. Below I advance two related explanations, both of which invoke the notion of an interaction between the temporal and spatial organisation of items to explain the absence of an increase in interpositions for grouped spatial sequences.

The first account appeals to the idea that in the spatial domain errors of temporal order are modulated by the spatial similarity of items. Hitch (1974) demonstrated that in spatial recall items that are spatially proximal to one another are more vulnerable to confusion than items that are spatially distinct – a spatial analogue of the phonological similarity effect in verbal serial recall. It follows that if items are proximal to one another on the temporal dimension (rendering them vulnerable to confusion), but distant on the spatial dimension then this may protect those items

[3] I thank Tom Hartley for raising this point.

from erroneously exchanging positions with one another. The implications for grouped sequences are as follows: all things being equal two items occupying different groups, but sharing the same positions within groups should be prone to exchanging positions, but if those items are spatially distinct from one another then this may diminish the probability of such interpositions materializing. Thus, the failure to detect an increase in the incidence of interpositions for grouped spatial sequences might be because those errors were washed out due to interactions between the temporal and spatial similarity of items.

The second account invokes the idea of an interaction between the temporal grouping pattern imposed at encoding and any spatial grouping of items that may be afforded by the spatial structure of the sequence to-be-remembered. As noted by De Lillo (2004), the encoding of spatial sequences in serial memory tasks such as the Corsi-Blocks Task may be heavily influenced by spatial constraints such as the relative proximity of locations and their alignment along spatial vectors. Such spatial constraints, which are always present to some degree, may be detected and used spontaneously by participants to divide sequences into spatial groups in order to reduce the demands on serial order processing. It follows that if participants do spontaneously group sequences on the basis of spatial constraints then this is likely to interact with any temporally defined grouping of items. For example, imagine a nine-item spatial sequence that is temporally grouped into three groups of three locations. Imagine further that the first five locations in this sequence are all clustered around one another permitting their organisation into a spatial group and that the following four locations are also clustered around one another permitting their organisation into a second spatial group. In this example, the spatial grouping afforded by the sequence is at odds with the experimenter imposed temporal grouping of the sequence. One way to envisage the impact of this is that the temporal and spatial parsings of the sequence can cooperate to help disambiguate the positions of items. Alternatively, the two parsings may compete with one another to some degree, with one parsing receiving a greater weighting than the other. In the current experiment, it is clearly the case that the temporal parsing of the sequence was the dominant form of organisation (e.g., the scalloping of the accuracy and latency serial position curves points to a 3-3-3 grouping pattern), but this does not preclude the possibility that the spatial organisation of items also exerted a more subtle background effect on performance. Either way, the presence of two different parsings of the sequence could conceivably have affected interposition rates.

If the explanations offered above were shown empirically to be valid then this would indicate that the GP-WP model needs to incorporate the effects of such spatial constraints to accurately model the data reported here. Clearly these hypotheses are speculative and need to be appraised via further empirical and modelling efforts. However, this issue is not pursued further within the context of this thesis, since a more important objective is to obtain converging evidence for the combination of a primacy gradient and positional marking in verbal, visual, and spatial serial memory.

It is worth considering briefly how the fits of the GP-WP model to the serial position curves could have been improved. One shortcoming of this model in general was its failure to predict stronger within group recency effects (indeed these effects were absent for the second sub-group for both the verbal and spatial data). This problem could be remedied by including an end positional marker, which codes the position of each item relative to the end of its group, as in the full instantiation of Henson's (1998) model. However, as demonstrated by Farrell and Lelievre (2009), the benefits of end positional marking can be obtained without positing a fully continuous end marker that codes the position of every item relative to the end of its group. They can be more parsimoniously accomplished by positing a restricted end marker that is activated only for the final item of each group. Such a restricted end marker, which has been shown to contribute to verbal serial recall (Farrell & Lelievre, 2009), was not included here, because interest centred on comparing the different models under the most elementary assumptions necessary to produce grouping behaviour.

Before closing, it merits comment that the outcomes of the current modelling exercise provide a useful illustration of some of the pitfalls associated with a purely verbal approach to theorising. Although the theory that spatial serial memory depends upon group position and position within sequence representations of order is intuitively plausible at the verbal level, implementation of that theory as a formal model has revealed a fundamental deficiency that is not apparent from consideration of the verbal theory alone. Specifically, although the GP-SP model can capture the appropriate level of scalloping of the accuracy serial positions curve, as well as capture the underlying distribution of errors within and between groups, it cannot capture these two things at the same time. Thus, a reliance solely on the verbal theory would have led incorrectly to the acceptance of that theory. The value of the modelling reported here has been to reveal the inadequacies of that theory allowing it to be discounted so that alternative theories can be pursued. The outcomes of the current modelling exercise therefore serve as a further demonstration of how formal models can be used to overcome fundamental limitations of human reasoning (see Hintzman, 1991 and Lewandowsky, 1993 for further illustrative examples).

Chapter summary

This chapter demonstrated that the failure to observe an increase in interpositions for temporally grouped spatial sequences in Experiments 7 and 9 was not the consequence of the use of a recall method different to that employed in studies of temporal grouping using verbal stimuli. A single experiment, which directly compared the impact of temporal grouping on a verbal and a spatial serial memory task using a common recall method, revealed a reliable increase in the probability of interpositions for grouped relative to ungrouped sequences for the verbal task, but not for the spatial task. The outcomes of subsequent model comparison work revealed that this discrepancy does not appear to be the result of different representations in the two domains of the positions of items in grouped sequences. On the contrary, a common model assuming representations of the positions of groups and the positions of items within groups provided the best description of the verbal and spatial data. The results tentatively support the hypothesis that position within group representations of items are a feature of grouped representations in both the verbal and spatial domains.

8

Modelling fill-in and infill errors across domains

Abstract

This chapter reports three experiments that examined the distribution of fill-in and infill errors underlying serial reconstruction of sequences of verbal, visual, and spatial stimuli. The experiments consistently revealed more fill-in than infill errors across all locations at which these errors are possible under both a conservative and a liberal scoring procedure. These results were invariant with respect to the sequence length manipulation employed across experiments. Fits of four models of serial order to the data for each experiment revealed that neither a model implementing a primacy gradient in conjunction with response suppression, nor a model implementing positional marking in conjunction with response suppression could accommodate the data. A model instantiating all three explanatory principles provided a satisfactory account of the empirical pattern, and this account was further enhanced by the incorporation of a fourth explanatory principle: namely a restricted end positional marker.

Introduction

The preceding chapters have provided evidence for the role of a primacy gradient, positional marking, and response suppression in verbal, visual, and spatial short-term order memory. However, the evidence for the combined action of these principles has been gleaned from a single paradigm involving analyses of error latency patterns across the three domains. The aim of the current chapter is to try to obtain converging evidence for this combination of principles.

In addition to the dynamics of recall errors, a further means by which adjudication between alternative mechanisms for representing serial order might be accomplished is on the basis of their fill-in and infill error predictions. Recall from Chapter 1 that a fill-in error is a conditional recall error in which following the anticipation of an item *n* immediately ahead of its correct position, item n-1 is recalled at the subsequent output position. For example, in response to the input sequence ABC the second item B is anticipated at the first output position, followed by the postponement of A at the second output position. Fill-in errors can be contrasted with infill errors in which following the initial anticipatory error, item $n+1$ is recalled at the subsequent output position. For example, after anticipating B at the first output position, C is anticipated at the second output position.

Henson (1996) conducted a meta-analysis of a number of verbal serial recall experiments and found that fill-in errors occurred approximately twice as frequently as infill errors. Page and Norris (1998) reported similar findings from a re-analysis of a series of experiments presented in Henson et al. (1996). However, Page and Norris only examined the incidence of fill-in and infill errors at the first and second serial positions, whilst Henson examined the incidence of fill-in and infill errors without reference to their position of occurrence. The analyses of these authors therefore ignore a potentially valuable and diagnostic pattern of errors. Surprenant et al. (2005) remedied this omission by examining the relative incidence of fill-in and infill errors for seven-item sequences across all locations at which both errors are possible. In a seven-item sequence there are five such locations represented by the pairs of output positions: 12, 23, 34, 45, and 56. For example, recalling AB*DC*EFG in response to the sequence ABCDEFG involves a fill-in error at the third location (positions 34), whilst recalling ABCD*FG*E involves an infill error at the fifth location (positions 56). The authors used a conservative and a liberal scoring procedure to record the incidence of fill-in and infill errors. In the former, an error was recorded only if all responses preceding it were correct, whereas in the latter an error was recorded irrespective of whether responses preceding it were correct or not.

Across two experiments, manipulating the modality of sequence presentation (auditory / visual) employing an open (Experiment 1) or closed (Experiment 2) verbal stimulus ensemble, Surprenant et al. (2005) observed more fill-in than infill errors, not only overall but at each possible error location throughout the sequence, irrespective of the modality and ensemble-size manipulations. The overall ratio of fill-in to infill errors was approximately 2:1. However, the relationship between error location and the magnitude of the ratios of fill-in to infill errors varied as a function of the two scoring procedures. For example, in the visual presentation condition of Experiment 1 the ratios were approximately 2:1 for each error location under conservative scoring, whereas under liberal scoring the ratio started at 2:1, but then decreased gradually across locations[1].

The results of the above studies are problematic for various classes of models of short-term order memory. As noted by Henson et al. (1996), the greater incidence of fill-in than infill errors is antithetic to the predictions of chaining models (e.g., Lewandowsky & Murdock, 1989; Murdock, 1993, 1995). Such models necessarily predict more infill than fill-in errors, because following the erroneous recall of an item *n* ahead of its correct position the direct associative link between this item and item $n+1$ will render the latter item the strongest competitor at the next output position. The results are also problematic for models in which serial order is encoded via associations between items and symmetrical cue positional markers (e.g., Brown et al., 2000; Burgess & Hitch, 1992), which predict fill-in and infill errors to be approximately equally likely (unless augmented with ancillary assumptions e.g., a primacy gradient of activations over items as in Burgess & Hitch, 1999, 2006). This is because the positional marker vectors associated with items in these models change in a constant manner, meaning that the amount of overlap between any pair of neighbouring position markers will always be the same. Thus, having erroneously recalled item *n* a position too soon, the position marker vector for the next position will cue items $n-1$ and $n+1$ to the same extent. Models that represent serial order on the basis of a primacy gradient in conjunction with response suppression (Farrell & Lewandowsky, 2004; Page & Norris, 1998), on the other hand, correctly predict more fill-in than infill errors. This is because if item *n* is recalled a position too soon, owing to the decreasing nature of the primacy gradient item $n-1$ will be a stronger competitor at the next output position than item $n+1$. Nevertheless, as noted by Henson et al. (1996), one problem for primacy gradient models is that they dramatically over predict the extent of fill-in.

[1] Surprenant et al. (2005) only present the data for conservative scoring. The data for liberal scoring can be obtained from the following URL: http://memory.psych.mun.ca/pubs/reprints/tr2005-03.pdf.

However, this shortcoming does not rule out the basic primacy gradient and response suppression mechanism, but rather points to the need for auxiliary representational assumptions.

It is apparent from the foregoing discussion that fill-in and infill are an informative pattern of errors that can distinguish the predictions of different mechanisms for representing serial order in short-term memory. Accordingly, given their apparent diagnosticity, one of the objectives of the current chapter is to obtain more data on fill-in and infill errors by examining their relative incidence across verbal, visual, and spatial serial memory tasks. A second objective is to fit a number of computational models built from different combinations of representational principles to the data from these experiments, in order to contrast their fill-in and infill error predictions and thereby identify a preferred combination of principles for representing serial order in the three domains. Of particular interest, is whether the combination of a primacy gradient, positional marking, and response suppression, which has proved so successful in explaining the dynamics of transpositions across domains in previous chapters can also accommodate the pattern of fill-in and infill errors.

The organisation of the rest of this chapter is as follows. First, the generic modelling architecture used for the simulations and the competitor models to be contrasted are introduced, followed by details of the model evaluation and comparison procedure. Subsequently, three experiments are presented that examined the distribution of fill-in and infill errors across verbal, visual, and spatial serial memory tasks, each experiment being complemented by quantitative fits of the rival models. Following Surprenant et al. (2005), fill-in and infill errors were examined across all output positions and the pattern of errors evaluated under both conservative and liberal scoring procedures. Additionally, in the experiments that follow, a sequence length manipulation was incorporated to ascertain whether the distribution of fill-in and infill errors is sensitive to differences in the number of to-be-remembered items and the accuracy of recall. This is particularly pertinent given that previous examinations of these errors (Henson, 1996; Page & Norris, 1998; Surprenant et al., 2005) have been restricted to a single sequence length. These data will provide a rich set of constraints for evaluating the predictions of the different models.

Computational modelling

In this section, I describe the modelling architecture employed for the simulations, introduce the competitor models to-be-contrasted, and delineate the procedure for fitting the models to the empirical data.

Modelling architecture and competitor models

The generic modelling architecture employed to obtain fill-in and infill error predictions for different combinations of representational principles was that described in Chapter 2, but with two critical changes relative to the version employed in Chapters 3, 4, 5, and 6 to model transposition latencies. Specifically, the decision threshold (*T*) for a response was omitted and the recurrent competitive field was switched off by setting the excitatory (w^+) and inhibitory (w^-) weights to values of zero. This alteration renders the network non-dynamic, meaning that decisions about a response at each output position are determined based upon the initial starting activations (modified by random noise to induce transposition errors). Thus, there was no response accumulation process and no need for item nodes to exceed a decision threshold. These changes were incorporated because the dependent measure of interest was response probabilities as opposed to latencies and because these changes significantly speed-up the process of estimating model parameters during fitting.

The fill-in and infill error predictions of four models and associated mechanisms for representing serial order were compared. These included the position marking and response suppression (PM+RS) model and the primacy gradient and response suppression (PG+RS) model, both examined in previous chapters. As noted in the introduction to this chapter, it is already known that fill-in and infill errors pose a problem for symmetrical cue position marking models (of which the PM+RS model is an example) and primacy gradient models. However, to my knowledge there have been no archived demonstrations of the predictions of these models for fill-in and infill errors, rendering it unclear the extent to which they deviate from the empirically observed pattern. For example, it is not known to what degree primacy gradient models over-predict fill-in errors and how this varies as a function of error location and scoring procedure. The PM+RS and PG+RS models constitute the baseline models, whilst the models described below examine the extent to which their explanatory power can be enhanced by combining their principles and incorporating further auxiliary assumptions.

The third model was the primacy gradient, position marking, and response suppression (PG+PM+RS) model, also examined in earlier chapters. This combination of model principles was naturally motivated on the basis of the results of the modelling and empirical work examining error latency patterns reported previously. Since the effect of augmenting a primacy gradient with positional markers will be to reduce the relative difference in activation between neighbouring items, relative to the activations specified by a primacy gradient alone, it was anticipated that the PG+PM+RS model would be better able to accommodate the magnitude of the observed ratios of fill-in to infill errors than the PG+RS model.

The fourth model, dubbed the PG+PM+RE+RS model, built on the above model by incorporating an ancillary representational assumption, namely that of a restricted end positional marker (RE), which was described in Chapter 7. The restricted end marker implements the idea that the final item in a sequence is tagged for its position relative to the end of that sequence. The incorporation of this representational assumption is motivated by the data of Surprenant et al. (2005), in which a recency effect was generally observed for the ratio of fill-in to infill errors at the final error location. To foreshadow the data, such recency effects are observed in all three of the experiments reported in this chapter. The restricted end marker, which is buttressed by recent empirical and modelling work reported by Farrell and Lelievre (2009), provides a basis for capturing this recency effect, since it increases the distinctiveness of the item at the final serial position, thereby protecting it from being involved in an infill error. For the sake of brevity, the implementation of the restricted end marker is described in Appendix 4.

Model fitting procedure

In the current chapter, models were fit to the data for all the conditions of each experiment and the results of the quantitative modelling exercises are presented alongside the analyses of the corresponding empirical data. The models were fit to the transposition matrix of response frequencies, summed across participants, for each sequence length condition. Thus, unlike previous modelling efforts the models were fit to the aggregate data, as opposed to individual participant data. This approach was adopted because in the experiments that follow several participants failed to contribute fill-in and infill errors at certain error locations. Because the models tend to produce errors at all locations, this can result in the models overestimating the magnitude of the ratios at locations where errors are sometimes missing empirically, and so fitting to the aggregate data was considered a preferred option.

The models were fit to the data using Maximum Likelihood Estimation; that is parameter values were sought that maximized the likelihood that a model produced the data if it were correct. The likelihood of the observed response frequencies for each output position (row vector) given the corresponding vector of response probabilities predicted by the model being fit under a specific combination of model parameter values was calculated using the multinomial log-likelihood function (equation 2-10). A likelihood estimate for the entire matrix of response frequencies was then obtained by summing the multinomial log-likelihood estimates for each output position, in order to produce a joint multinomial log-likelihood estimate. It is important to emphasise that the transposition matrix does not contain any information about the distribution of fill-in and infill errors.

Parameter estimates were obtained using the simplex algorithm (Nelder & Mead, 1965) minimizing the negative joint multinomial log-likelihood estimate. Each parameter vector explored by the search algorithm involved 50,000 model simulation trials. To help prevent the search algorithm from converging on local minima the search process was conducted multiple times employing different starting conditions. Below I describe the parameters that were varied and frozen during the fitting of the different models. Whilst reading this section the reader may find it useful to refer back to Chapter 2 for a refresher on the functions of the different parameters.

The parameters minimized in the fitting process for the PM+RS model were the distinctiveness of the position markers (ϕ), and the standard deviation of noise (σ); for the PG+RS model these were the steepness of the primacy gradient (γ), and the standard deviation of noise (σ). For the

PG+PM+RS model these were the steepness of the primacy gradient (γ), the distinctiveness of the position markers (ϕ), and the standard deviation of noise (σ); for the PG+PM+RE+RS model these were the steepness of the primacy gradient (γ), the distinctiveness of the position markers (ϕ), the strength of the restricted end marker (ε), and the standard deviation of noise (σ).

The constrained parameters for the PM+RS model were the weighting parameter for the strength of the position markers (λ = 1), and the amount of response suppression (α = 1); for the PG+RS model these were the weighting parameter for the strength of the first item (a_1 = 1), and the amount of response suppression (α = 1); for the PG+PM+RS and PG+PM+RE+RS these were the weighting parameter for the strength of the first item (a_1 = 1), the weighting parameter for the strength of the position markers (λ = 1), the attentional weighting parameter for the primacy and positional dimensions of ordering (ω = .5), and the amount of response suppression (α = 1).

In summary, the number of free model parameters was two for the PM+RS and PG+RS models, three for the PG+PM+RS model, and four for the PG+PM+RE+RS model.

Model comparison procedure

The models were initially fit to the transposition matrices according to the procedure described above, which yielded for each model a set of best fitting parameter values and an associated maximum log-likelihood estimate for the data for each sequence length condition. The models were then compared in two stages. In the first stage, the maximum log-likelihood estimates for the fits to the transposition matrices were converted into Bayesian Information Criterion (BIC; Schwarz, 1979) scores[2]. This enabled a comparison of the models when both their descriptive accuracy and complexity were taken into consideration. To facilitate model comparisons, the discrete raw BIC scores were converted into BIC weights, as in the model comparisons reported in Chapter 7. In the second stage, the best fitting model parameter values obtained from the fits to the transposition matrices were used to generate fill-in and infill error predictions for the rival models, and their

[2] The BIC was chosen over the AIC because it offers a more stringent correction for model complexity.

goodness-of-fits to the ratios of fill-in to infill errors observed empirically were calculated. Note that given the non-availability of an appropriate likelihood function, it was not possible to evaluate the goodness-of-fits of the models to these data using a likelihood objective. Instead, goodness-of-fits were evaluated by calculating the root mean square deviations (RMSDs) between the observed ratios of fill-in to infill errors and those predicted by the models. The reliance on RMSD meant that it was not possible to compare the fits of the models to the ratios of fill-in to infill errors using BIC scores to trade-off goodness-of-fit with model complexity.

Experiment 11

The purpose of this experiment was to examine the distribution of fill-in and infill errors underlying serial reconstruction of verbal sequences varying in length from five to seven items. The sequences were composed of words drawn from an open stimulus ensemble, in order to facilitate comparison with the data from the subsequent two experiments using visual and spatial stimuli, in which an open stimulus set is employed for the purpose of minimizing the likelihood of verbal encoding.

Method

Participants

Eighteen individuals recruited from the campus community at the University of York took part in the experiment in exchange for course credit or an honorarium of £10.

Stimuli & apparatus

The stimuli were sequences of five to seven words, which were created by randomly sampling without replacement from a stimulus ensemble containing 976 words. The words were obtained from the MRC Psycholinguistic Database, and were 6-8 letters in length and had a Kucera and Francis (1968) written word frequency of between 10 and 50. The words were presented singly in the central screen position in Arial point 18 lower case font. Stimulus presentation and data collection were controlled using software developed by the author running on a Dell Optiplex (Intel

Core 2 Duo, 2.13 GHz processor) PC equipped with a 19" monitor and a Razer Copperhead high precision laser mouse.

Design & procedure

The experiment manipulated a single independent variable: Sequence length (sequences of five, six, and seven items), which was a within-participants factor. The six possible permutations of the three conditions were fully counterbalanced across participants.

Participants initiated each trial by clicking on a 'begin trial' icon located in the centre of the computer screen using the computer mouse. Following a 500ms delay, a central fixation cross was presented for 500ms, which was followed by an additional 500ms delay before presentation of the first item. Each item was presented for 1000ms, separated by a 200ms inter-item delay. Following the final item there was a 200ms delay, after which the items reappeared simultaneously within a noisy circular array. Participants were required to click on the items in their presentation order using the mouse-driven pointer. Once an item was selected it cleared from the reconstruction array so that it could not be selected again. Participants were encouraged to guess whenever they were unsure of the correct item for a given position, otherwise they could select a question mark located in the centre of the reconstruction array to record a 'don't know response'. Once a response had been registered for each output position there was a 3000ms delay, following which the 'begin trial' icon for the next trial appeared.

Participants attempted 100 trials for each sequence length, which were preceded by two practice trials. Because of the large number of trials, participants completed the experiment across two approximately 60 minute testing sessions (spaced at least 24 hours apart) each containing 50 trials for each sequence length. The counterbalancing of conditions was identical for the two testing sessions.

Results

Accuracy serial position curves

The accuracy serial position curves for the data are shown in the top left panel of Figure 8-1. Immediately apparent is that the sequence length manipulation fostered a reliable reduction in serial reconstruction accuracy, which was statistically confirmed by a one-way ANOVA performed on the proportion of correct responses, averaging across serial positions, for each condition, $F(2, 34) = 85.738$, $MSE = .278$, $p < .001$. Follow up tests revealed that performance was better for five-item than for six-item sequences, $t(17) = 7.172$, $p < .001$, and that in turn performance was better for six-item than for seven-item sequences, $t(17) = 6.219$, $p < .001$. There were significant effects of serial position for five-item sequences, $F(4, 68) = 28.886$, $MSE = .197$, $p < .001$, six-item sequences, $F(5, 85) = 26.191$, $MSE = .351$, $p < .001$, and seven-item sequences, $F(6, 102) = 23.629$, $MSE = .471$, $p < .001$.

Transposition gradients

The transposition gradients underlying the serial position curves are shown in the top left panel of Figure 8-2 and exhibit the same hallmarks as documented in previous chapters. Specifically, for each sequence length transpositions are distributed approximately symmetrically around displacement zero, reflecting the proportion of transpositions for each absolute displacement is approximately the same for anticipations and postponements. The greatest proportions of errors were transpositions with an absolute displacement value of one, with the proportions of transpositions decreasing monotonically as the absolute value of the displacement increased. Consistent with the serial position analysis, the incidence of anticipations and postponements increased as a function of sequence length.

The latency-displacement functions associated with these data are characterised by steep negative slopes for anticipations and flat slopes for postponements (Appendix 2).

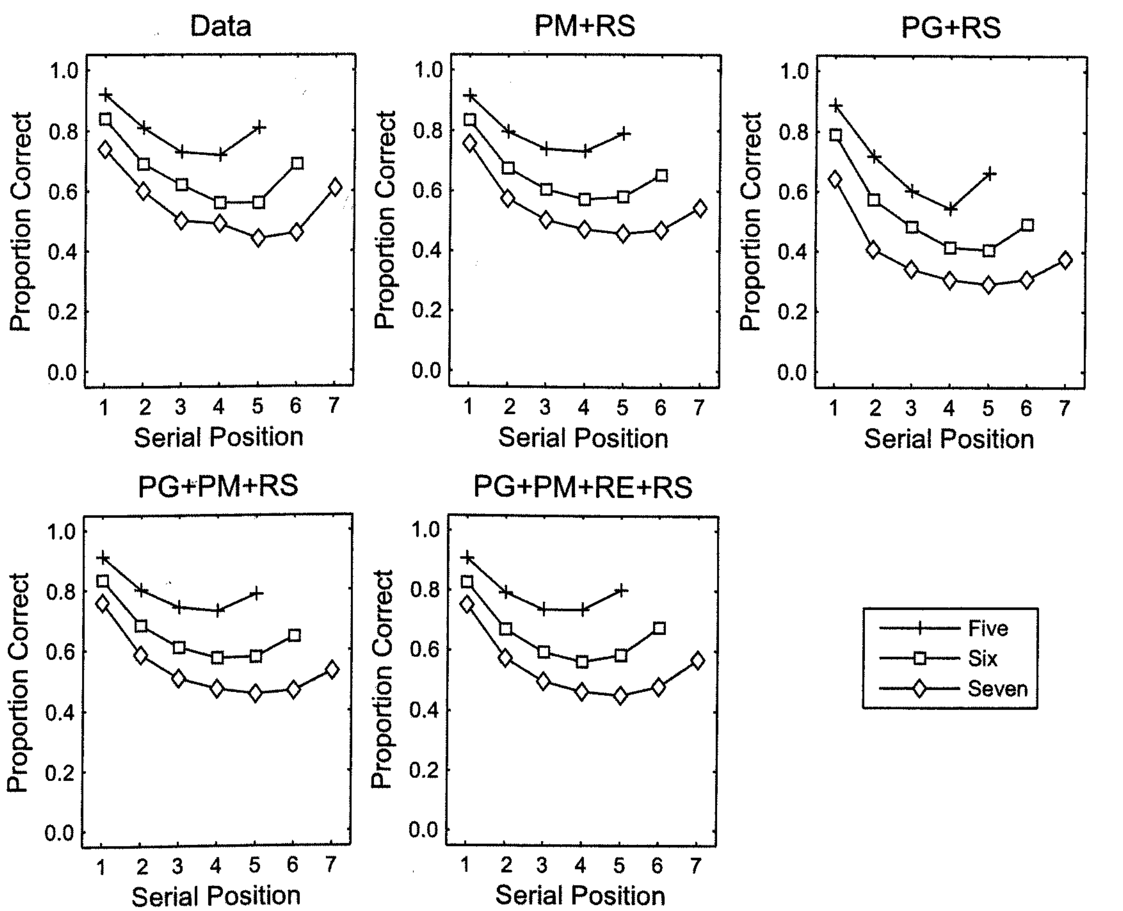

Figure 8-1 *Serial position curves for Experiment 11 and fits of four models of serial order.*

Fill-in and infill errors

Turning now to the data of central interest, Table 8-1 shows the mean number of fill-in and infill errors as a function of error location and sequence length under conservative and liberal scoring. To help facilitate interpretation, a brief description of these data is in order. The first column indicates the sequence length condition, whilst each subsequent pair of columns gives the mean number of fill-in and infill errors for a specific error location. For example, columns two and three give respectively, the mean number of fill-in and infill errors at the first error location, which comprises serial positions 1 and 2. A fill-in error at this location involves reporting item 2 at position 1, before reporting item 1 at position 2, whilst an infill error involves the same initial mistake, after which item 3 is reported at position 2. As a further example, columns four and five give respectively, the mean number of fill-in and infill errors at the second error location, which comprises serial positions 2 and 3. A fill-in error at this location involves reporting item 3 at

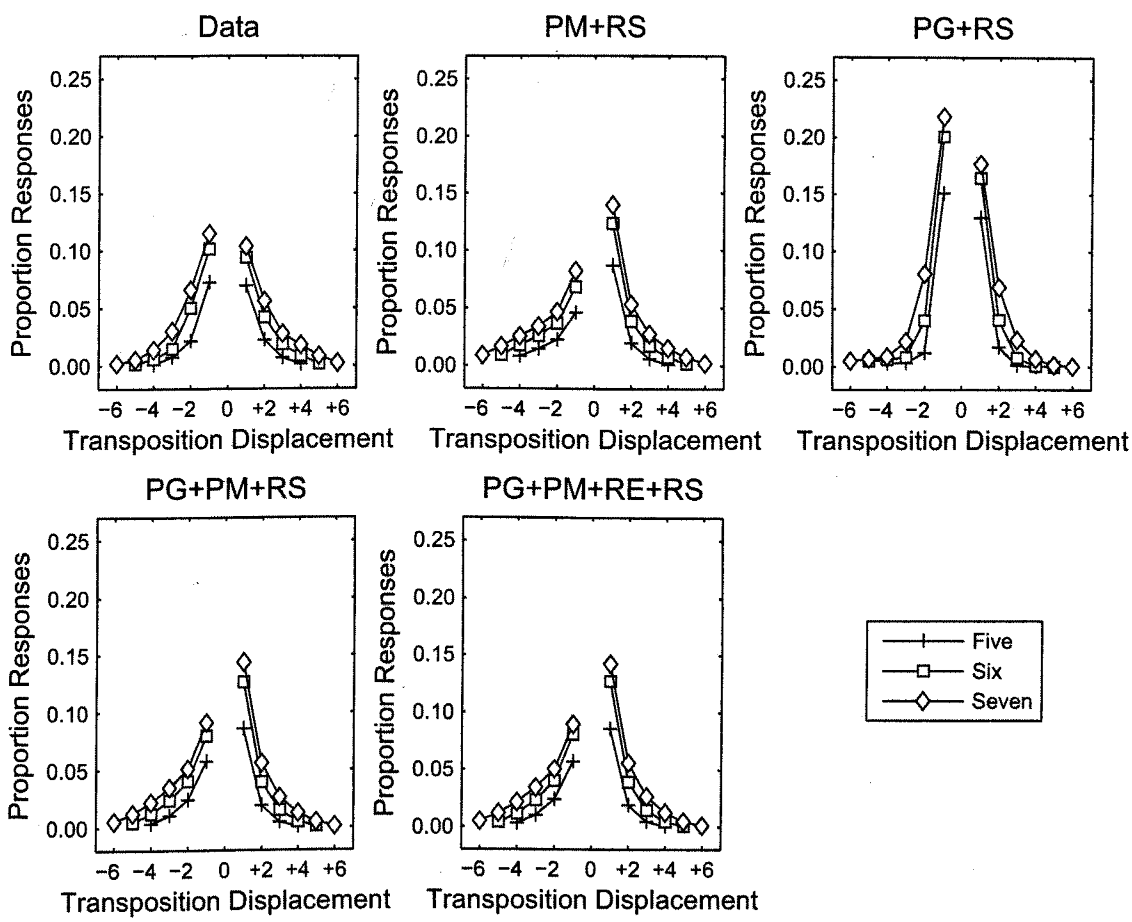

Figure 8-2 *Transposition gradients for Experiment 11 and fits of four models of serial order.*

position 2, before reporting item 2 at position 3, whilst an infill error involves the same initial mistake, after which item 4 is reported at position 3. There are three possible error locations for five-item sequences, four for six-item sequences, and five for seven-item sequences. The upper half of Table 8-1 gives the mean number of fill-in and infill errors for conservative scoring, under which errors are recorded only if the responses at all serial positions preceding the location at which the error occurred were correct. For example, an error at location 4 (positions 4-5) would only be recorded if responses at serial positions 1, 2, and 3 were correct. In contrast, the lower half of Table 8-1 gives the mean number of fill-in and infill errors for liberal scoring, under which errors are recorded irrespective of whether responses at locations preceding the error were correct or not. Note that for the first error location, the mean number of fill-in and infill errors is necessarily the same for conservative and liberal scoring.

	Positions 1-2		*Positions 2-3*		*Positions 3-4*		*Positions 4-5*		*Positions 5-6*	
	21	*23*	*32*	*34*	*43*	*45*	*54*	*56*	*65*	*67*
					Conservative Scoring					
Five	2.72	0.89	4.94	2.11	6.67	1.00				
Six	3.89	1.39	4.17	2.28	4.17	1.78	4.50	1.89		
Seven	3.44	1.78	4.39	2.22	3.33	1.33	2.56	1.56	4.17	1.17
					Liberal Scoring					
Five	2.72	0.89	5.00	2.50	7.39	1.72				
Six	3.89	1.39	4.56	2.67	5.00	3.39	6.17	3.78		
Seven	3.44	1.78	4.83	2.78	4.33	2.61	4.22	3.67	7.50	4.22

Table 8-1 *Mean number of fill-in and infill errors for Experiment 11.*

In brief, there were more fill-in than infill errors for all sequence length conditions under both scoring procedures. This is illustrated more clearly in the top left panel of Figure 8-2, which plots the ratios of fill-in to infill errors for the different conditions. There are several features of these data that merit comment. First, the ratios are uniformly greater than one, indicating a ubiquitous tendency for fill-in errors to outweigh infill errors. Second, the ratios of fill-in to infill errors are greater under conservative scoring than under liberal scoring. Third, under conservative scoring the ratios are relatively constant across locations, except for recency effects at the final error location for lengths five and seven. In contrast, under liberal scoring the ratios decrease gradually across locations, but also exhibit recency effects at the final error location, albeit smaller in magnitude than for conservative scoring. Fourth, in general the ratios decrease as a function of sequence length: the mean ratios for five-item, six-item, and seven-item sequences, averaged across error locations and scoring procedure, were 3.57:1, 2.12:1 and 1.99:1, respectively.

Statistical confirmation of the greater incidence of fill-in than infill errors was sought by means of separate 2 Scoring (Conservative / Liberal) x 2 Error-Type (Fill-in / infill) x 3, 4 and 5 Error Locations ANOVAs performed on the frequency of errors for five-item, six-item, and seven-item

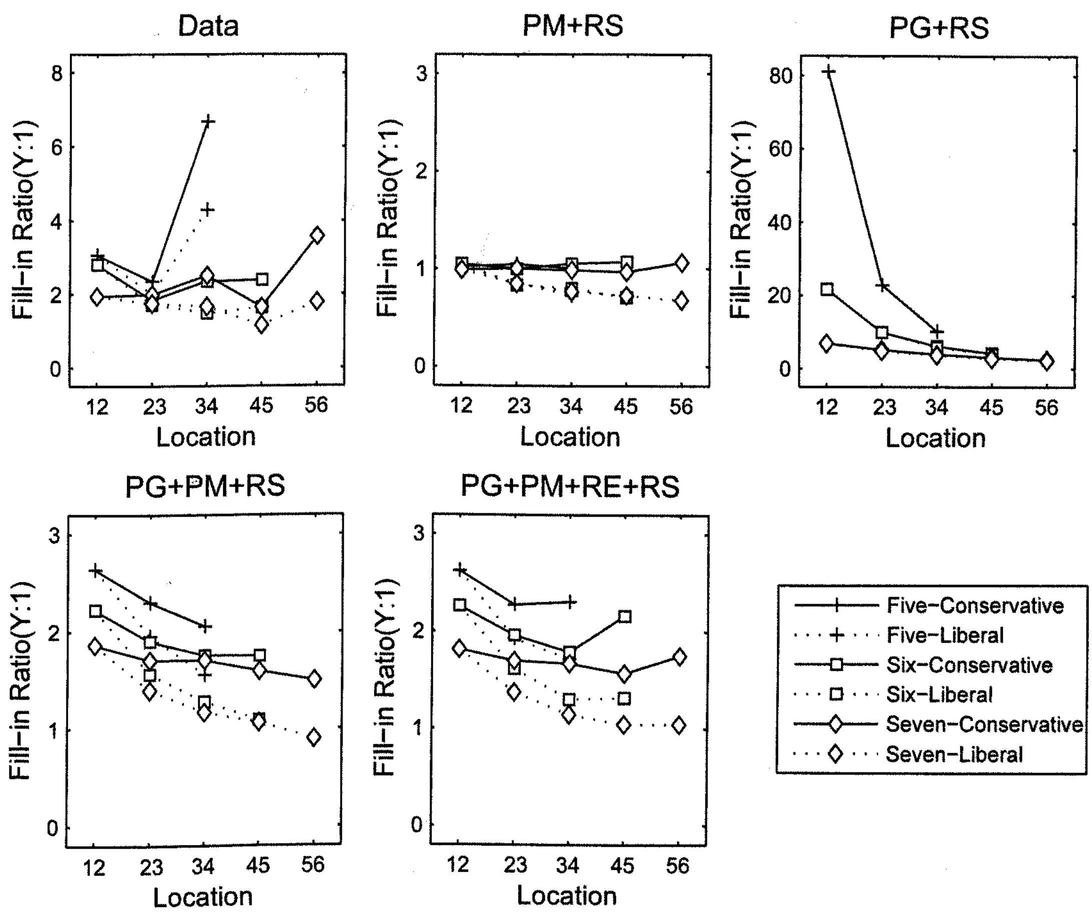

Figure 8-3 *Ratios of fill-in to infill errors for Experiment 11 and fits of four models of serial order. Note—the y-axis scales vary across panels.*

sequences[3]. For five-item sequences, there was a reliable main effect of Scoring, $F(1, 17) = 8.161$, $MSE = 5.352$, $p = .01$, due to more errors for liberal than conservative scoring, a reliable main effect of Error-Type, $F(1, 17) = 47.633$, $MSE = 620.167$, $p < .001$, owing to more fill-in than infill errors, and a reliable main effect of Location, $F(2, 34) = 6.758$, $MSE = 112.519$, $p < .01$, which materialized due to an increase in errors across locations. The only significant interactions were the Scoring x Location interaction, $F(2, 34) = 6.758$, $MSE = 112.519$, $p < .01$, which manifested because the incidence of errors did not differ between the two scoring procedures at the first error

[3] For economy of exposition only significant effects are reported for all analyses of fill-in and infill errors throughout this chapter.

location, but there were more errors at locations two and three under liberal scoring[4], and the Error-Type x Location interaction, $F(2, 34) = 9.904$, $MSE = 73.167$, $p < .001$, which arose because fill-in errors increased monotonically across locations, whereas infill errors dropped off at the final location.

Turning to the analysis for six-item sequences, there was a reliable main effect of Scoring $F(1, 17) = 34.872$, $MSE = 51.681$, $p < .001$, again reflecting more errors for liberal than conservative scoring, as well as a reliable main effect of Error-Type, $F(1, 17) = 36.673$, $MSE = 355.556$, $p < .001$, again due to a greater incidence of fill-in than infill errors. The only significant interaction was the Scoring x Location interaction, $F(3, 51) = 20.199$, $MSE = 15.602$, $p < .001$, which arose due to a greater increase in errors across locations for liberal than conservative scoring.

Finally, for seven-item sequences, there were reliable main effects of Scoring, $F(1, 17) = 105.064$, $MSE = 162.678$, $p < .001$, and Error-Type, $F(1, 17) = 48.054$, $MSE = 328.711$, $p < .001$, consistent with the above analyses, in addition to a reliable main effect of Location, $F(4, 68) = 2.768$, $MSE = 30.801$, $p < .05$, which transpired due to an elevation in errors across locations. There was also a reliable Scoring x Location interaction, $F(4, 68) = 27.101$, $MSE = 47.236$, $p < .001$, which was again due to a greater increase in errors across locations for liberal than conservative scoring.

Modelling

The goodness-of-fits of the models to the transposition matrices and the ratios of fill-in to infill errors for the different sequence length conditions are shown in Table A1-11 of Appendix 1. To facilitate interpretation, I focus here on the goodness-of-fits of the models averaged across sequence lengths. Averaging across conditions is generally more informative than considering fits

---

[4] Note that in all of the analyses of fill-in and infill errors reported in this chapter the Scoring x Location interaction will always be significant. This is because the incidence of fill-in and infill errors at the first error location will necessarily always be the same under the two scoring procedures, whereas at subsequent error locations there will always be more errors recorded under liberal than conservative scoring.

Model	k	Transposition Matrices ln$L$	BIC	ωBIC	Fill-in Ratios RMSD
PM+RS	2	-12265	24536	0.00	1.84
PG+RS	2	-13240	26487	0.00	20.02
PG+PM+RS	3	-12178	24367	0.08	1.15
PG+PM+RE+RS	4	-12170	24355	0.92	1.05

Table 8-2 *Goodness-of-fits of four models of serial order, averaged across sequences length conditions, to the data of Experiment 11. Note—k = number of free model parameters; lnL = maximum log-likelihood; BIC = Bayesian Information Criterion; ωBIC = BIC weight; RMSD = root mean square deviation.*

to each condition separately, since the purpose of fitting the models to multiple conditions is to obtain a more reliable 'overall' picture of their descriptive accuracy.

Table 8-2 presents the goodness-of-fits averaged across the three sequence length conditions. The first column indicates the model, whilst columns two to five provide information relevant to the fits of the models to the transposition matrices. Specifically, column two gives the number of free model parameters; column three gives the maximum log-likelihood estimate; column four gives the BIC; whilst column five gives the ωBIC (BIC weight), which is the conditional probability that a model is the best model, given the field of competitor models and the data. Note that for the goodness-of-fit indices in columns three and four the model with the smallest value is preferred, whereas for ωBIC the model with the largest value is preferred. The final column gives the root mean square deviation (RMSD) for the model fits to the ratios of fill-in to infill errors – smaller RMSDs are preferred over larger RMSDs.

Considering first the fits to the transposition matrices, it can be seen from inspection of the mean BIC scores that the PG+RS model provided the worst fit. The PG+PM+RE+RS model provided a better fit than the PG+PM+RS model, and both of these models in turn provided better fits than the PM+RS model. It can be seen that the mean BIC weight for the PG+PM+RE+RS model is distinctly larger than that of its competitor models, indicating that it provided the best fit to the transposition matrices by some margin. Turning to the fits of the models to the ratios of fill-in to fill-in errors, the PG+RS model again provided the worst fit. The PG+PM+RE+RS model again provided a better fit than the PG+PM+RS model, and both of these models in turn performed better than the PM+RS model. In summary, the PG+PM+RE+RS model provided the best fit to both the transposition matrices and the ratios of fill-in to infill errors.

Figure 8-1 shows the accuracy serial position curves generated by the models. It can be seen that the PM+RS, PG+PM+RS, and PG+PM+RE+RS models all provide similarly good fits to the accuracy serial position curves. Indeed, the predictions of the above models are virtually identical. The PG+RS model generated less convincing serial position curves characterised by steeper effects of primacy than observed empirically. The associated transposition gradients predicted by the models are shown in Figure 8-2, from which it can be seen that the PM+RS, PG+PM+RS, and PG+PM+RE+RS models, once again provided similarly good accounts of the data. However, all three models predicted a greater probability of +1 than -1 transpositions, which is contrary to the data. The PG+RS model gave the poorest fit to the transposition gradients, as it considerably over predicted the incidence of transpositions with an absolute displacement value of one.

Turning to the predictions of chief interest, Figure 8-3 shows the ratios of fill-in to infill errors predicted by the models. Starting with the PM+RS model, it can be seen that this model correctly predicts larger ratios of fill-in to infill errors under conservative than liberal scoring. However, under conservative scoring, the model predicts fill-in ratios of approximately 1:1 across error locations. That is, as anticipated the model erroneously predicts an equivalent number of fill-in and infill errors. Under liberal scoring, the ratio starts at 1:1, but then decreases gradually across error locations. Although this decrease in ratios across locations is qualitatively consistent with the data, because the ratios fall below a value of 1:1, the model erroneously predicts that infill errors outweigh fill-in errors by gradually increasing amounts across error locations. The ratios of fill-in to infill errors predicted by this model, averaged across error locations and scoring procedure, were 0.95:1, 0.94:1, and 0.90:1, respectively, for five-item, six-item, and seven-item sequences. These mean ratios indicate that overall, the PM+RS model predicts slightly more infill errors than fill-in errors, which is antithetic to the data. A further shortcoming of this model is that it does not predict the recency effects observed at the final error locations.

Considering now the PG+RS model, as expected this model over-predicted the magnitude of the ratios of fill-in to infill errors. The mean ratios predicted by this model, averaged across error locations and scoring procedure, were 38:1, 10:1, and 4:1, for five-item, six-item, and seven-item sequences, respectively. Although the mean ratio for seven-item sequences does not deviate markedly from the observed mean ratio, the predicted mean ratios for five-item and six-item sequences are distinctly larger than observed empirically. It is also apparent that the model predicts comparable ratios of fill-in to infill errors under conservative and liberal scoring, which is at odds with the data. Under conservative scoring, the model predicts a steep decrease in the fill-in ratios across error locations, which is at odds with the approximate consistency of the ratios observed empirically at pre-recency error locations. The model generates essentially the same predictions under liberal scoring, and although the decrease in the ratios across locations is qualitatively consistent with the data, the predicted drop in the ratios is too steep. Like the PM+RS model, the PG+RS model also failed to predict the recency effects observed at final error locations.

As hoped, the PG+PM+RS model predicted ratios of fill-in to infill errors of a magnitude more similar to those seen empirically than the abovementioned models. The mean ratios predicted by this model were 2.2:1, 1.7:1, and 1.5:1 for five-item, six-item, and seven-item sequences, respectively, which are close to the empirically observed mean ratios. This model correctly predicted larger ratios of fill-in to infill errors under conservative than liberal scoring. It also predicted a small decrease in the size of the error ratios across error locations under conservative scoring, combined with a steeper decrease in the size of the error ratios across locations under liberal scoring, which is broadly compatible with the empirical data. Nevertheless, as for the PM+RS and PG+RS models, this model failed to capture the recency effects observed at final error locations.

The PG+PM+RE+RS model enjoyed the same benefits as its restricted model counterpart, whilst additionally predicting 'some' recency effects under conservative scoring, in accordance with the data. The mean error ratios predicted by this model, averaged across error locations and scoring procedure, were 2.3:1, 1.8:1, and 1.5:1 for five-item, six-item, and seven-item sequences, respectively. Nevertheless, the recency effects predicted by this model are smaller in magnitude than those seen empirically. Moreover, the model did not predict any recency effects whatsoever under liberal scoring, which is contrary to the data. Consequently, the results of the current model fitting exercise provide only tentative support for the contribution of a restricted end marker.

Discussion

The outcomes of the current experiment demonstrate that for serial reconstruction of verbal stimuli there are more fill-in than infill errors at each error location throughout the sequence, under both conservative and liberal scoring procedures, and that this empirical pattern holds across the different sequence lengths examined. The ratios of fill-in to infill errors are larger under conservative scoring than under liberal scoring, and whilst the error ratios are relatively constant across locations under conservative scoring, under liberal scoring the error ratios decrease gradually across locations. The current experiment also demonstrated effects of recency on the ratios of fill-in to infill errors: for five-item and seven-item sequences (but not six-item sequences) the ratio of fill-in to infill errors increased at the final error location under conservative and liberal scoring.

The modelling outcomes confirm that neither the combination of positional marking with response suppression, nor the combination of a primacy gradient with response suppression can accommodate the distribution of fill-in and infill errors underlying verbal serial reconstruction. The former combination of principles either predicts an equivalent incidence of fill-in and infill errors (under conservative scoring) or predicts more infill than fill-in errors (under liberal scoring), whilst the latter combination of principles predicts ratios of fill-in to infill errors of a magnitude distinctly greater than that seen empirically. The same modelling outcomes demonstrate that when all three representational principles are combined a more accurate description of the magnitude and pattern of the ratios of fill-in to infill errors is obtained, and that this description is further enhanced through the addition of a restricted end marker to model effects of recency.

Reanalysis of Experiment 1

This section comprises a re-analysis of Experiment 1 reported in Chapter 4, in order to examine the distribution of fill-in and infill errors underlying serial reconstruction of sequences of visual stimuli of varying length. For brevity, the details of this experiment are only re-iterated in their simplest form; for more specific details the reader should re-consult Chapter 4. Participants were presented with sequences of four to six unfamiliar faces for immediate serial reconstruction. Sequences were created by randomly sampling from a large ensemble of images of faces subject to the constraint that no face was presented on more than two occasions throughout the entire experiment (there were not quite sufficient images to generate sequences of entirely unique items). Like Experiment 11, participants took part in two testing sessions, each containing 50 trials for each sequence length.

Results

Accuracy serial position curves

The accuracy serial position curves are shown in the top left panel of Figure 8-4. Serial position analyses of these data were reported in Chapter 4, and so for brevity these are not repeated here. In brief, performance was significantly better for four-item than five-item sequences, and performance was better in turn for five-item than six-item sequences. The usual effects of serial position were significant for all three sequence lengths.

Transposition gradients

The transposition gradients underlying the serial position curves are shown in the top left panel of Figure 8-5 and exhibit the same empirical regularities as documented in the preceding experiment.

Fill-in and infill errors

The mean number of fill-in and infill errors as a function of error location and sequence length for conservative and liberal scoring, are shown in Table 8-3. In short, there was once again more

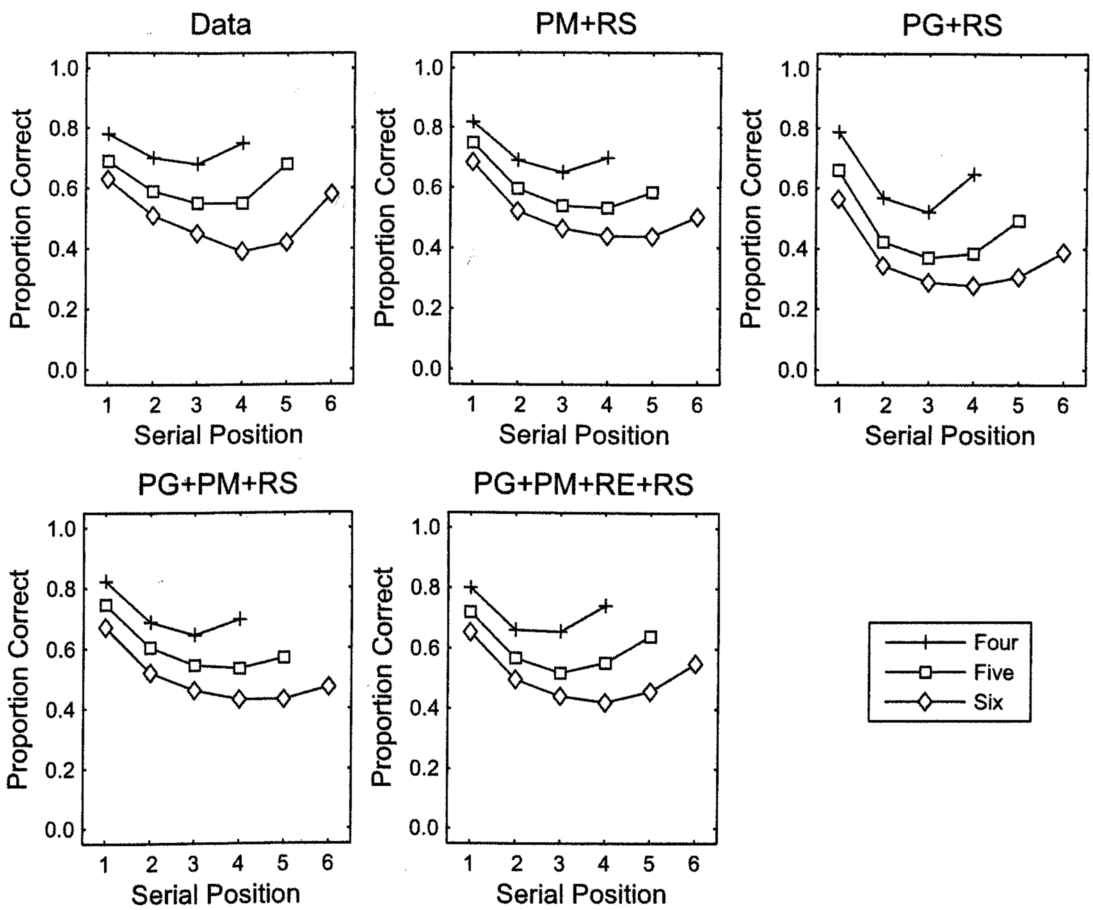

Figure 8-4 *Serial position curves for Experiment 1 and fits of four models of serial order.*

fill-in than infill errors for all conditions under both scoring procedures. This can be determined more readily by inspecting the top left panel of Figure 8-6, which plots the ratios of fill-in to infill errors as a function of sequence length and error location. The first thing to note about these data is that the ratios are consistently greater than one, revealing a universal tendency for more fill-in than infill errors, with larger ratios under conservative than liberal scoring. Second, for conservative scoring, the ratios increase gradually across error locations for four-item and five-item sequences, whereas the ratios are approximately constant across locations for six-item sequences, but with a recency effect at the final location. For liberal scoring, the ratios decrease across locations for four-item and six-item sequences, but with a recency effect at the final location for six-item sequences. For five-item sequences, the ratios are constant across the first two locations, but with an upturn at the final error location. Finally, the mean ratios, averaged across error locations and scoring procedure, decreased as a function of sequence length: the mean ratios of fill-in to infill errors,

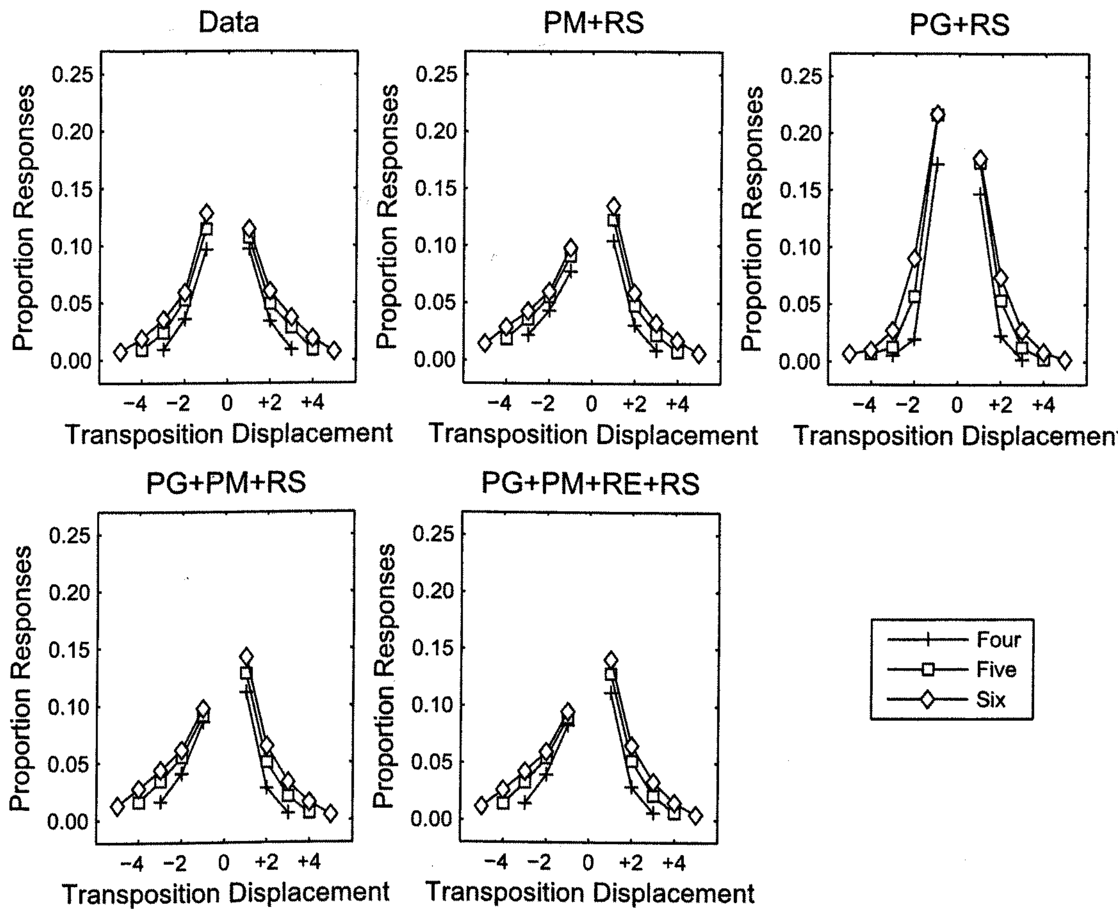

Figure 8-5 *Transposition gradients for Experiment 1 and fits of four models of serial order.*

averaged across error locations and scoring procedure, were 3:1, 2.4:1, and 1.8:1, for four-item, five-item, and six-item sequences, respectively.

Statistical confirmation of the greater incidence of fill-in than infill errors was sought by means of separate 2 Error-Type (Fill-in / infill) x 2 Scoring (Conservative / Liberal) x 2, 3, and 4 Error Locations ANOVAs performed on the frequency of errors for four-item, five-item, and six-item sequences. For the analysis of four-item sequences, there was a reliable main effect of Error-Type, $F(1, 17) = 46.289$, $MSE = 826.563$, $p < .001$, due to more fill-in than infill errors, as well as a reliable main effect of Scoring, $F(1, 17) = 8.673$, $MSE = 1.562$, $p < .01$, owing to more errors under liberal than conservative scoring. The only significant interaction was the Scoring x Error Location interaction, $F(1, 17) = 8.673$, $MSE = 1.562$, $p < .01$, which arose because the incidence of errors was identical under the two scoring procedures at the first location, but higher at the second location for liberal scoring.

	Positions 1-2		*Positions 2-3*		*Positions 3-4*		*Positions 4-5*	
	2-1	*2-3*	*3-2*	*3-4*	*4-3*	*4-5*	*5-4*	*5-6*
			Conservative Scoring					
Four	7.39	2.39	6.83	1.94				
Five	5.78	3.06	6.17	2.39	5.39	1.56		
Six	3.83	2.50	4.11	2.44	3.44	2.11	4.56	1.39
			Liberal Scoring					
Four	7.39	2.39	6.94	2.67				
Five	5.78	3.06	6.94	3.56	7.17	2.72		
Six	3.83	2.50	5.39	3.94	5.39	4.56	7.22	3.89

Table 8-3 *Mean number of fill-in and infill errors for Experiment 1.*

For the analysis of five-item sequences, there was once again a reliable main effect of Error-Type, $F(1, 17) = 37.947$, $MSE = 234.722$, $p < .001$, reflecting more fill-in than infill errors, as well as a reliable main effect of Scoring, $F(1, 17) = 77.063$, $MSE = 171.125$, $p < .001$, due to more errors under liberal than conservative scoring. The only significant interaction was the Scoring x Error Location interaction, $F(3, 51) = 27.374$, $MSE = 23.468$, $p < .001$, which emerged because the incidence of errors at the first location did not diverge between conservative and liberal scoring, whereas for locations two and three there were more errors under liberal scoring.

Finally, for the analysis of six-item sequences, there was once more a reliable main effect of Error-Type, $F(1, 17) = 37.947$, $MSE = 234.722$, $p < .001$, owing to more fill-in than infill errors, in addition to a reliable main effect of Scoring, $F(1, 17) = 77.063$, $MSE = 171.125$, $p < .001$, which again manifested due to more errors under liberal than conservative scoring. Once more, the only significant interaction was the Scoring x Error Location interaction, $F(3, 51) = 27.374$, $MSE = 23.468$, $p < .001$, which materialized because the number of errors was identical for conservative and liberal scoring at the first location, but was greater at all other locations for liberal than conservative scoring.

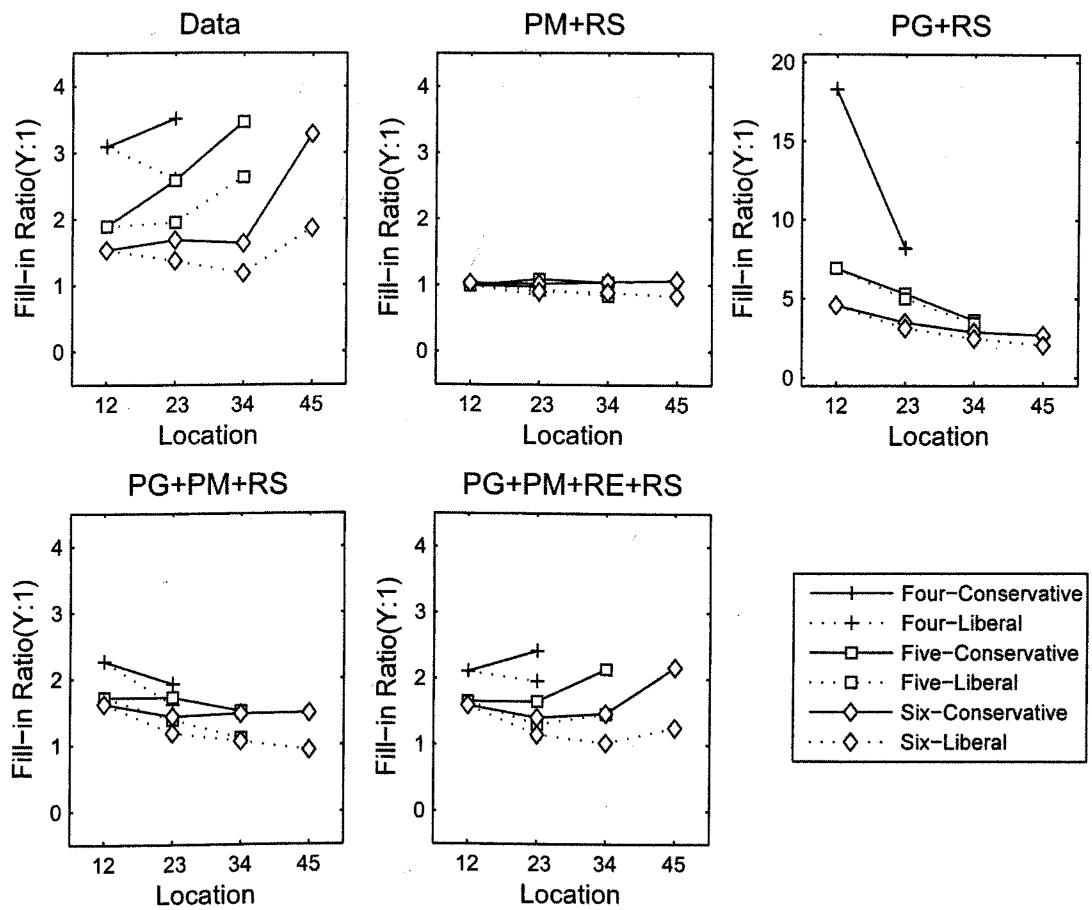

Figure 8-6 *Ratios of fill-in to infill errors for Experiment 1 and fits of four models of serial order. Note—the y-axis scales vary across panels.*

Modelling

The goodness-of-fits of the models to the transposition matrices and the ratios of fill-in to infill errors for the different sequence length conditions are shown in Table A1-12 of Appendix 2. As previous, I compare the descriptive accuracy of the models by considering their mean goodness-of-fits (averaged across fits to the three sequence length conditions), which are given in Table 8-4. Considering first the fits of the models to the transposition matrices, it is apparent from inspection of the mean BIC scores that the PG+RS model provided the worst fit. The PG+PM+RE+RS model provided a better fit than the PG+PM+RS model, and both of these models in turn provided better fits than the PM+RS model. It can be seen that the mean BIC weight for the PG+PM+RE+RS model assumed the maximum possible value, indicating that it provided the best fit to the transposition matrices by a considerable margin. Turning to the fits of the models to the ratios of

	Transposition Matrix				Fill-in Ratios
Model	k	ln$L$	BIC	ωBIC	RMSD
PM+RS	2	-10642	21294	0.00	1.55
PG+RS	2	-11339	22684	0.00	5.53
PG+PM+RS	3	-10632	21271	0.00	0.98
PG+PM+RE+RS	4	-10604	21220	1.00	0.76

Table 8-4 *Goodness-of-fits of four models of serial order, averaged across sequence length conditions, to the data of Experiment 1. Note—k = number of free model parameters; lnL = maximum log-likelihood; BIC = Bayesian Information Criterion; ωBIC = BIC weight; RMSD = root mean square deviation.*

fill-in to infill errors, it is visible from inspection of the mean RMSDs that the PG+RS model again performed the poorest. The PG+PM+RS model provided a better fit than the PM+RS model, but the PG+PM+RE+RS model again provided the best fit. Thus, as for Experiment 11, the PG+PM+RE+RS model provided the best fits to both the transposition matrices and the ratios of fill-in to infill errors.

From inspection of the serial position curves shown in Figure 8-4, it can be seen that the PG+RS model produced unrealistic serial position curves, which exhibited a sharp drop in performance from the first to the second serial position combined with insufficient recency. The PM+RS and PG+PM+RS models both captured the effects of primacy, but under predicted the effects of recency. The PG+PM+RE+RS model matched the explanatory power of the latter two models, whilst also predicting stronger effects of recency, although still not as strong as seen empirically. The associated transposition gradients predicted by the models are shown in Figure 8-5, from which it can be seen that the PM+RS, PG+PM+RS, and PG+PM+RE+RS models all provided similarly good accounts of the data. The only apparent discrepancy is that all models predicted a greater incidence of +1 than −1 transpositions, as was the case for the fits of the same models to the data for Experiment 11. The PG+RS model generated less realistic transposition gradients, because once again it considerably over predicted the incidence of transpositions with an absolute displacement value of one.

The model predictions for the ratios of fill-in to infill errors are illustrated in Figure 8-6. It is apparent from inspection of this figure that the PM+RS model predicted only slightly larger mean ratios under conservative than liberal scoring. Under conservative scoring, the model once again failed to predict more fill-in than infill errors, the error ratios being approximately 1:1 across all error locations. Under liberal scoring, the ratio starts at 1:1, but then decreases slightly across error locations, with infill errors becoming more frequent than fill-in errors, contrary to the data. It is evident that the model additionally failed to capture the observed recency effects. The ratios of fill-in to infill errors predicted by this model, averaged across error locations and scoring procedure, were 0.95:1, 0.98:1, and 0.97:1 for four-item, five-item, and six-item sequences, respectively. Taken as a whole then, the PM+RS model predicted an approximately equal number of fill-in and infill errors.

The PG+RS model correctly predicted more fill-in than infill errors, but it over-predicted the extent of fill-in for four-item and five-item sequences, although less so for six-item sequences: the ratios of fill-in to infill errors predicted by this model, averaged across error locations and scoring procedure, were 13:1, 5:1, and 3.22:1, for four-item, five-item, and six-item sequences, respectively. It merits comment that the PG+RS model did not over-predict the extent of fill-in as much as when it was applied to the data of Experiment 11. Nevertheless, a further shortcoming of this model was that it failed to predict sufficiently larger ratios of fill-in to infill errors under conservative than liberal scoring. The model additionally predicted a decrease in the size of the error ratios across error locations under conservative scoring, which is once again contrary to the data. Like the PM+RS model, the PG+RS model additionally failed to capture the observed recency effects.

Turning to the PG+PM+RS model, as before this model fared much better than the two abovementioned models, predicting ratios of fill-in to infill errors of a magnitude and pattern more similar to those seen empirically. This model correctly predicted larger ratios of fill-in to infill errors under conservative than liberal scoring. The ratios of fill-in to infill errors predicted by this model, averaged across error locations and scoring procedure, were 2:1, 1.5:1, and 1.3:1, for four-item, five-item, and six-item sequences, respectively. These mean ratios are smaller than, but nevertheless close to the empirically observed mean ratios. The main shortcoming of this model was that like the PM+RS and PG+RS models, it failed to accommodate the observed recency effects. The PG+PM+RE+RS model, predicted ratios of fill-in to infill errors that were a slightly better approximation to those seen empirically than the ratios predicted by the PG+PM+RS model. The ratios of fill-in to infill errors predicted by this model, averaged across error locations and scoring procedure, were 2.2:1, 1.7:1, and 1.5:1, for four-item, five-item, and six-item sequences, respectively. This model additionally predicted recency effects, although as for the fits of the same model to the data of Experiment 11 these effects were smaller in magnitude than seen empirically. Nevertheless, the correspondence between the model and the data is impressive given that it was not fitted directly to the ratios of fill-in to infill errors.

Discussion

In brief, the pattern of ratios of fill-in to infill errors observed empirically for serial reconstruction of visual stimuli is generally comparable to that observed in Experiment 11 using verbal stimuli. There are once again more fill-in than infill errors at each error location throughout the sequence, under both conservative and liberal scoring procedures, and this empirical pattern holds across the different sequence lengths examined. As before, the ratios of fill-in to infill errors are larger under conservative than under liberal scoring. Additionally, the current experiment once again demonstrated effects of recency on the ratios of fill-in to infill errors: for all sequence lengths examined the ratio of fill-in to infill errors increased at the final error location under both conservative and liberal scoring. However, the pattern of error ratios across locations under the two scoring procedures is slightly different to that witnessed in the previous experiment. Specifically, for sequences of four-items and five-items the ratios of fill-in to infill errors increased across error locations under conservative scoring (this is non-surprising in the case of the data for four-item sequences given that there are only two possible error locations), and whilst the error ratios decreased across locations under liberal scoring for four-item sequences, they did not for five-item sequences. However, the pattern for six-item sequences is comparable to that observed in the previous experiment. Specifically, under conservative scoring the ratios of fill-in to infill errors are relatively stable across error locations, whereas under liberal scoring the ratios decrease gradually across error locations (notwithstanding the apparent recency effects at the final error location under both scoring procedures).

The outcomes of the model fitting exercise are also similar to that of Experiment 11. As before, the combination of position marking and response suppression failed to predict a fill-in advantage, whilst the combination of a primacy gradient with response suppression generally over-predicted the magnitude of the ratios of fill-in to infill errors, albeit to a lesser extent than was the case for the fits of the same model to the data of Experiment 11. As previous, the confluence of a primacy gradient, positional marking, and response suppression provided a more accurate description of the magnitude and pattern of the ratios of fill-in to infill errors, whilst the addition of a restricted end marker to this combination of representational principles yielded a further improvement in descriptive accuracy by permitting modelling of the observed effects of recency.

Experiment 12

The final experiment examined the distribution of fill-in and infill errors underlying spatial sequences comprising seven and nine items[5]. Unlike the spatial tasks employed in previous chapters, which used a fixed set of spatial coordinates, in the current experiment the spatial coordinates of locations varied randomly across trials, thereby engendering greater spatial uncertainty in their positioning. This procedure was adopted to minimize the possibility of any verbal recoding of the spatial sequences.

Method

Participants

Eighteen individuals from the campus community at the University of York took part in the experiment in exchange for course credit or an honorarium of £6.

[5] Time constraints prevented the inclusion of a third sequence length condition.

Stimuli & apparatus

The stimuli were sequences of seven or nine visually presented spatial locations. The locations were grey icons (measuring 2.5cm x 2.5cm), which were assigned to random spatial coordinates on each trial subject to the constraint that the minimum and maximum distances between pairs of icons (measured from the centre of each icon) were 3cm and 12cm, respectively. The apparatus used to present stimuli and record responses were the same as for Experiment 11.

Design & procedure

The experiment manipulated a single independent variable: Sequence length (sequences of seven and nine items), which was a within-participants factor. Half of the participants received sequences of seven-items followed by sequences of nine-items, whilst the remaining half of participants completed the two sequence length conditions in the converse order.

Participants initiated each trial by clicking on a 'begin' trial icon located in the centre of the computer screen using the computer mouse. After a 1000ms blank delay the icons were displayed individually on-screen according to a random order determined by the computer programme controlling stimulus presentation. Each icon was presented for 500ms, separated by a 500ms inter-item interval. Following the final item there was a 1000ms delay, after which the icons simultaneously reappeared in their presentation coordinates. Participants were required to click on the icons in their presentation order using the mouse-driven pointer. Once an icon was selected its colour changed transitorily to green for 50ms to acknowledge that the computer had registered the response. Unlike the two preceding experiments, items could be selected on multiple occasions, meaning that repetition responses were possible. Additionally, to maximize fill-in and infill errors the 'don't know response' option of Experiments 1 and 11 was omitted. Once a response had been registered for all output positions there was a 3000ms delay, following which the 'begin trial' icon was presented for the next trial.

Participants attempted 80 trials for each sequence length preceded by 2 practice trials. Unlike Experiments 1 and 11, data were collected within a single testing session lasting approximately 70 minutes. Participants were instructed to encode sequences without deploying supplementary verbal or gestural encoding strategies and all reported adherence to these instructions.

Results

Accuracy serial position curves

The serial position curves for accuracy can be inspected in Figure 8-7. Immediately apparent is that serial memory performance was considerably better for seven-item sequences than for nine-item sequences, which was statistically confirmed via a t-test performed on the mean proportion of correct responses for the two sequence length conditions, $t(17) = 12.1218$, $p < .001$. There were also reliable effects of serial position for seven-item sequences, $F(7, 119) = 42.154$, $MSE = .698$, $p < .001$, and for nine-item sequences, $F(8, 136) = 18.525$, $MSE = .551$, $p < .001$.

Transposition gradients

The transposition gradients underlying the serial position curves are shown in Figure 8-8 and exhibit the same empirical regularities as documented in the two preceding experiments. The latency-displacement functions associated with these data are characterised by steep negative slopes for anticipations and shallow positive slopes for postponements (Appendix 2).

Fill-in and infill errors

The mean number of fill-in and infill errors as a function of error location and sequence length for conservative and liberal scoring, are shown in Table 8-5. In short, there was once again more fill-in than infill errors for all of the conditions. This can be determined more readily by consulting the top left panel of Figure 8-9, which shows the same data, but expressed as ratios of fill-in to infill errors. It is apparent from inspection of this figure that the error ratios are uniformly greater than one, indicating that fill-in errors consistently outnumbered infill errors. Also apparent is that the ratios of fill-in to infill errors are greater under conservative than liberal scoring. Under conservative scoring, the ratios of fill-in to infill errors are relatively consistent across error locations except for recency effects at the final error locations. In contrast, under liberal scoring,

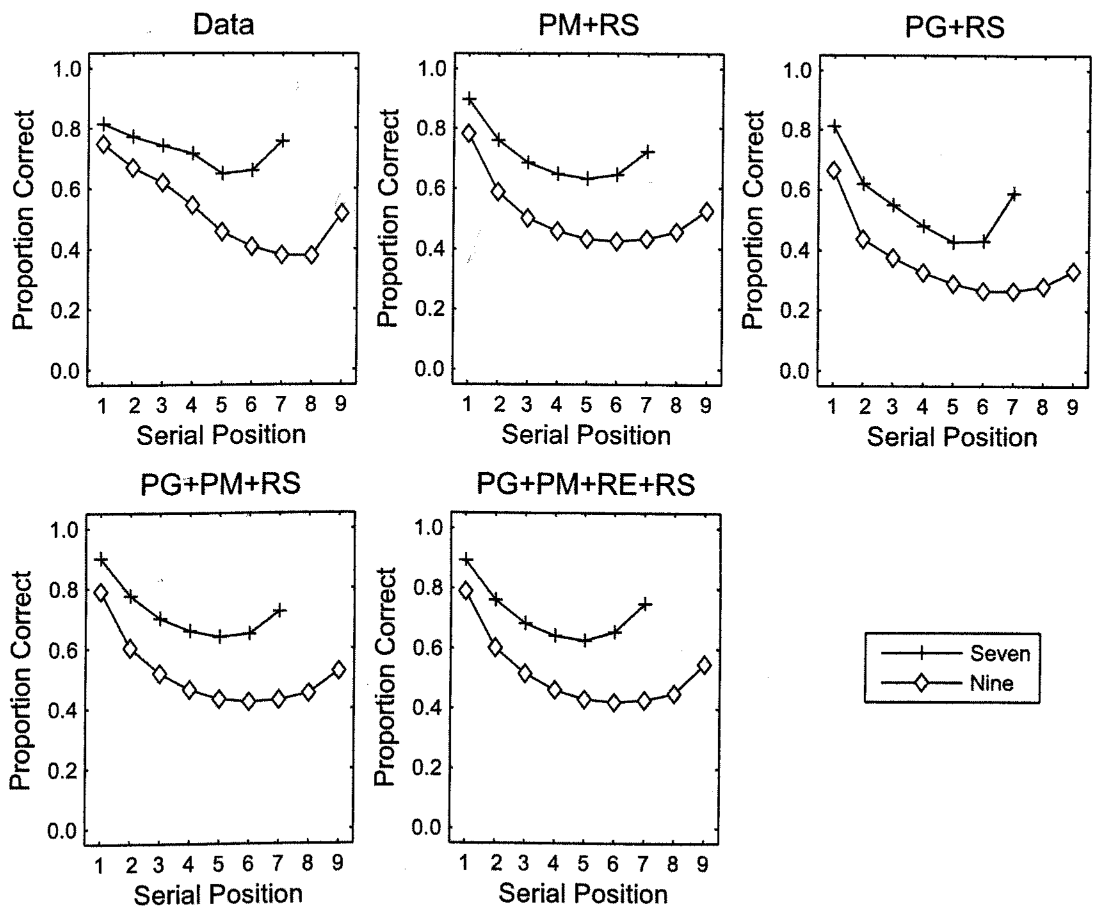

Figure 8-7 *Serial position curves for Experiment 12 and fits of four models of serial order.*

the ratios of fill-in to infill errors decrease gradually across locations, but with recency effects once again at the final error locations. It can be seen that the error ratio curves are noisier for sequences of nine-items: the curves are punctuated by alternating peaks and troughs. As in the previous experiments the mean error ratios, averaged across error locations, decreased as sequence length increased. The mean ratios of fill-in to infill errors, averaged across error locations and scoring procedure, were 3:1 and 2:1 for seven-item and nine-item sequences, respectively.

Statistical confirmation of the greater incidence of fill-in than infill errors was sought by means of separate 2 Error-Type (Fill-in / infill) x 2 Scoring (Conservative / Liberal) x 5 and 7 Error Locations ANOVAs performed on the frequency of errors for seven-item and nine-item sequences. For the analysis of seven-item sequences, there was a reliable main effect of Error-Type, $F(1, 17) = 52.551$, $MSE = 466.944$, $p < .001$, due to more fill-in than infill errors, there was also a reliable main effect of Scoring, $F(1, 17) = 28.998$, $MSE = 71.111$, $p < .001$, reflecting more errors under

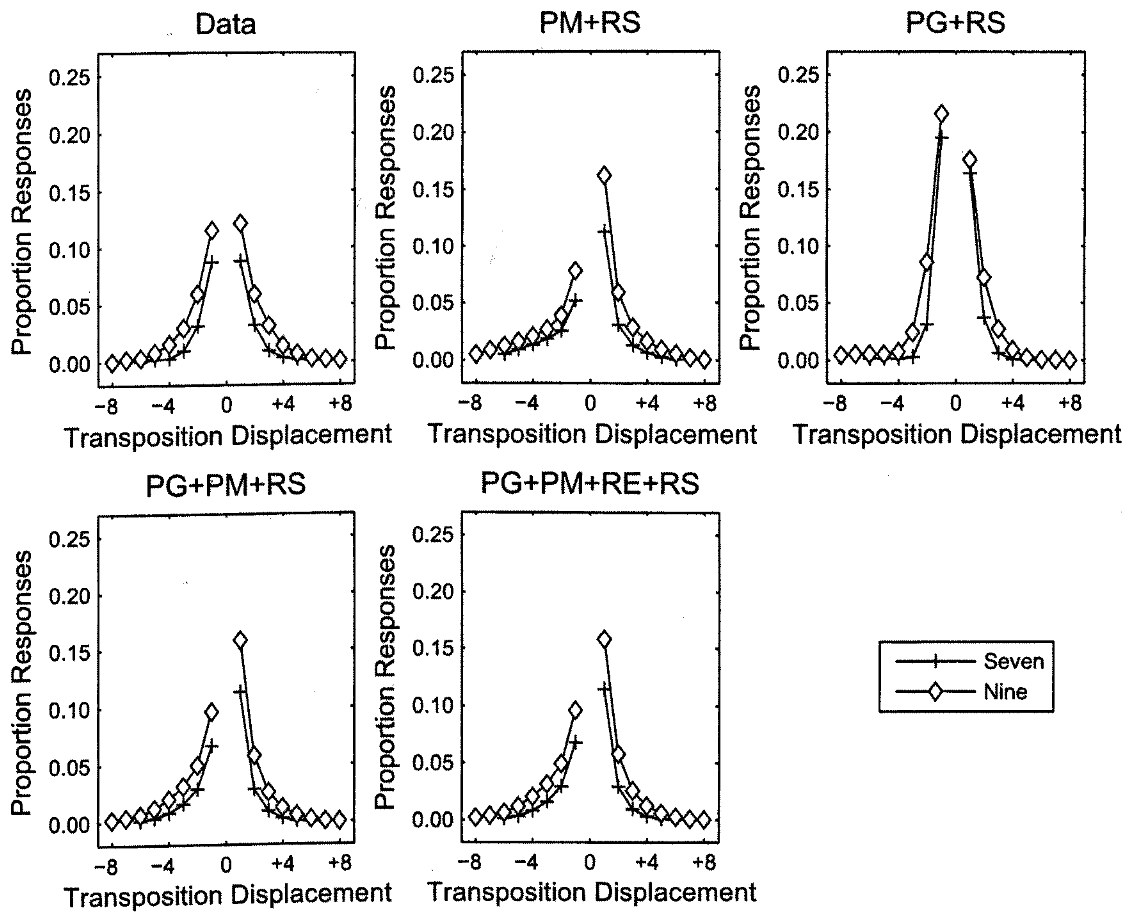

Figure 8-8 *Transposition gradient for Experiment 12 and fits of four models of serial order.*

liberal than conservative scoring, as well as a reliable main effect of Error-Location, $F(4, 68) = 9.754$, $MSE = 61.933$, $p < .001$, owing to an increase in errors across locations. The only reliable interactions were the Scoring x Error Location interaction, $F(4, 68) = 13.053$, $MSE = 17.368$, $p < .001$, and the Error-Type x Location interaction, $F(4, 68) = 2.468$, $MSE = 10.361$, $p < .05$. The former interaction arose because the frequency of errors was identical at the first location under liberal and conservative scoring, whereas for all other locations there were more errors under liberal scoring. The latter interaction arose because fill-in errors outweighed infill errors to a greater extent at the first and final error locations than at medial error locations.

For the analysis of nine-item sequences, there was again a reliable main effect of Error-Type, $F(1, 17) = 25.683$, $MSE = 322.240$, $p < .001$, due to more fill-in than infill errors, a reliable main effect of Scoring, $F(1, 17) = 136.169$, $MSE = 244.446$, $p < .001$, due to more errors under liberal than conservative scoring, in addition to a reliable main effect of Error Location, $F(6, 102) = 3.912$,

	Positions 1-2		Positions 2-3		Positions 3-4		Positions 4-5		Positions 5-6		Positions 6-7		Positions 7-8	
	21	*23*	*32*	*34*	*43*	*45*	*54*	*56*	*65*	*67*	*76*	*78*	*87*	*89*
							Conservative Scoring							
Seven	4.67	1.83	2.17	0.44	2.00	0.61	2.94	1.06	4.28	1.22				
Eight	4.28	1.72	3.22	0.83	2.78	0.67	2.61	1.11	3.50	1.72	3.67	1.28		
Nine	4.44	2.00	2.78	1.11	2.44	0.61	1.17	0.56	2.50	0.89	1.50	0.94	2.56	0.89
							Liberal Scoring							
Seven	4.67	1.83	2.50	1.00	3.11	1.39	4.33	2.00	6.39	2.89				
Eight	4.28	1.72	3.56	1.94	3.67	1.83	3.94	2.56	5.89	4.06	6.94	4.17		
Nine	4.44	2.00	3.28	1.94	3.28	1.28	2.50	1.89	4.44	2.44	3.22	2.83	6.78	3.56

Table 8-5 *Mean number of fill-in and infill errors for Experiment 12.*

MSE = 64.381, $p < .05$, reflecting that errors increased across locations. There were reliable two-way interactions between Scoring and Error Location, $F(6, 102) = 32.196$, *MSE* = 22.173, $p < .001$, and Error-Type and Error Location, $F(6, 102) = 2.104$, *MSE* = 11.485, $p = .05$, which were subsumed under a reliable Error-Type x Scoring x Error Location three-way interaction, $F(6, 102) = 2.865$, *MSE* = 2.903, $p < .05$. The interpretation of this three-way interaction is complicated and so for brevity is not considered here.

Modelling

The goodness-of-fits of the models to the transposition matrices and the ratios of fill-in to infill errors for the different sequence length conditions are shown in Table A1-13 of Appendix 1. As previous, I compare the descriptive accuracy of the models by considering their mean goodness-of-fits (averaged across fits to the two sequence length conditions), which are given in Table 8-6. Considering first the fits of the models to the transposition matrices, it can be seen from inspection of the mean BIC scores that the PG+RS model provided the worst fit. The PG+PM+RE+RS model provided a better fit than the PG+PM+RS model, which in turn provided a better fit than the PM+RS model. It is apparent from inspection of the mean BIC weights that the PG+PM+RE+RS model obtained a notably larger mean weight than the PG+PM+RS model, which in turn obtained a

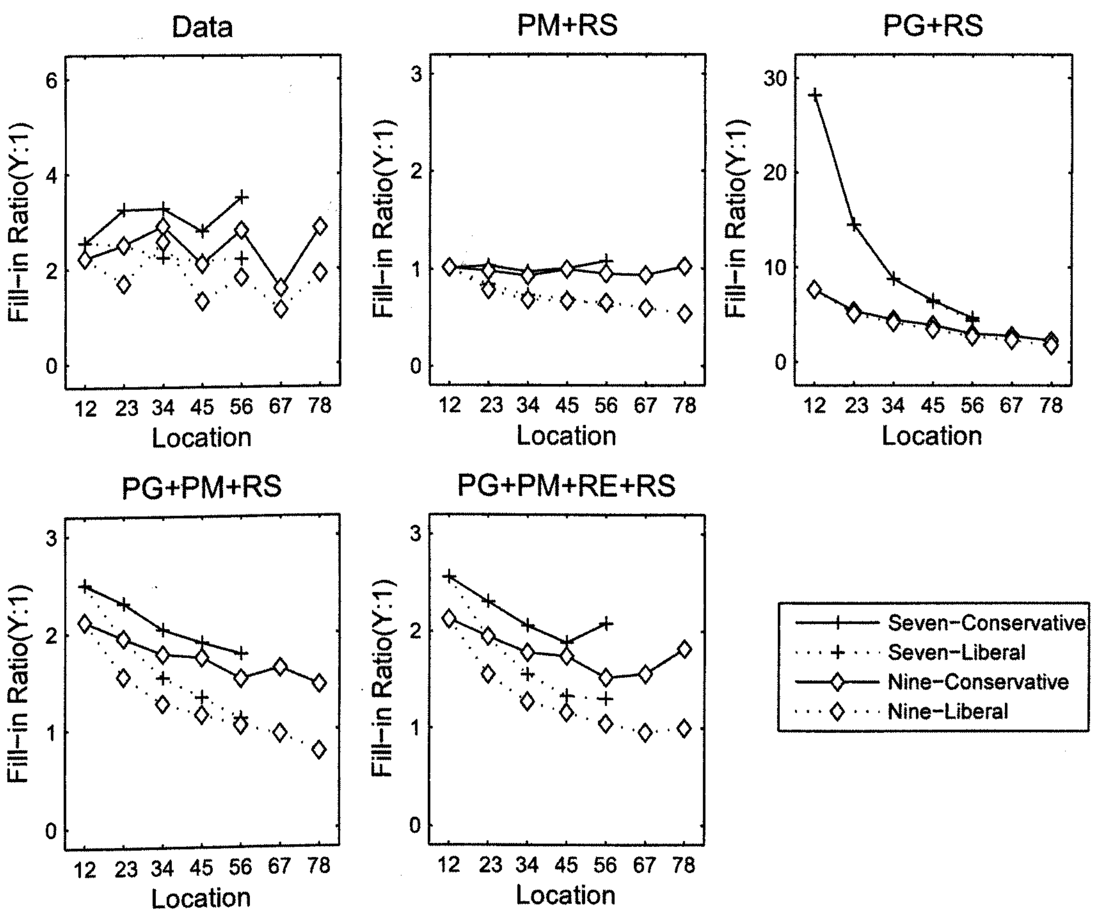

Figure 8-9 *Ratios of fill-in to infill errors for Experiment 12 and fits of four models of serial order. Note—the y-axis scales vary across panels.*

notably larger mean weight than the remaining two models. These results indicate that the PG+PM+RS model provided a much better fit than the PM+RS and PG+RS models, and that the descriptive accuracy of this model was moderately enhanced by the incorporation of a restricted end marker. Turning to the fits to the ratios of fill-in to infill errors, inspection of the mean RMSDs reveals that the worst fit was once again provided by the PG+RS model. The PG+PM+RE+RS model again provided the best fit, followed by the PG+PM+RS model, which in turn provided a better fit than the PM+RS model. Thus, the rankings of the models based upon their goodness-of-fits to the transposition matrices and ratios of fill-in to infill errors was identical to that for the fits of the models to the data of Experiments 1 and 11.

The model predictions for accuracy serial position curves and transposition gradients are shown in Figure 8-7 and Figure 8-8, respectively. The PM+RS, PG+PM+RS, and PG+PM+RE+RS

Model	k	lnL	BIC	ωBIC	RMSD
		Transposition Matrix			Fill-in Ratios
PM+RS	2	-17512	35033	0.00	1.60
PG+RS	2	-18420	36849	0.00	4.65
PG+PM+RS	3	-17234	34480	0.33	0.99
PG+PM+RE+RS	4	-17230	34478	0.67	0.80

Table 8-6 *Goodness-of-fits of four models of serial order, averaged across sequence lengths, to the data of Experiment 12. Note—k = number of free model parameters; lnL = maximum log-likelihood; BIC = Bayesian Information Criterion; ωBIC = BIC weight; RMSD = root mean square deviation.*

models all generated similar serial position curves and transposition gradients. Although all three models captured the extensive primacy of the serial position curves they also predicted effects of recency extending beyond a single item, which is contrary to the one-item recency observed empirically. The same cohort of models provided good fits to the transposition gradients, the only discrepancy being that they predicted more +1 than –1 transpositions, as in the model fits reported previously. The PG+RS model generated less realistic serial position curves and transposition gradients: the model under predicted performance across serial positions and considerably over predicted the incidence of transpositions with absolute displacement values of one and two.

Considering now the predictions of main concern, Figure 8-9 shows the ratios of fill-in to infill errors predicted by the models. Starting with the PM+RS model, this model correctly predicted larger ratios of fill-in to infill errors under conservative than liberal scoring. However, consistent with the previous model fits, the PM+RS model erroneously predicted an equivalent number of fill-in and infill errors under conservative scoring, as reflected by ratios of fill-in to infill errors of approximately 1:1 across error locations. Under liberal scoring, the ratios fell increasingly further below 1:1 across error locations, indicating a greater incidence of infill than fill-in errors. The mean ratios of fill-in to infill errors, averaged across error locations and scoring procedure, were 0.9:1

and 0.84:1 for seven-item and nine-item sequences, respectively. These mean ratios indicate that overall, the PM+RS model predicted more infill than fill-in errors, which is antithetic to the data.

The PG+RS model considerably over-predicted the incidence of fill-in relative to infill errors for seven-item sequences: the mean ratio of fill-in to infill errors averaged across error locations and scoring procedure was 12:1. As for the previous model fits, this problem was most pronounced at the first error location and then became gradually less marked across error locations. Of note, however, is that the PG+RS model did not over-predict by a significant margin the incidence fill-in relative to infill errors for nine-item sequences: the mean ratio of fill-in to infill errors averaged across error locations and scoring procedure was 4:1, which is close in value to the empirically observed mean ratio. Nevertheless, there are further shortcomings of this model. As before, it failed to predict sufficiently larger ratios of fill-in to infill errors for conservative relative to liberal scoring. It also predicted a decrease in the size of the ratios of fill-in to infill errors across error locations under conservative scoring, which is contrary to the empirical pattern. Additionally, the PG+RS model, like the PM+RS model, failed to capture the observed recency effects.

Turning to the PG+PM+RS model, this model predicted ratios of fill-in to infill errors of a magnitude and pattern more akin to that seen empirically. The mean ratios of fill-in to infill errors predicted by this model, averaged across error locations and scoring procedure, were 1.9:1 and 1.5:1, for seven-item and nine-item sequences, respectively. These mean error ratios are smaller than, but nevertheless close to the empirically observed mean ratios. This model correctly predicted larger ratios under conservative than under liberal scoring, as well as a gradual decrease in the error ratios across locations under liberal scoring. However, contrary to the data, the model additionally predicted a decrease in the error ratios across locations under conservative scoring. Like the PM+RS and PG+RS models this model failed to predict the recency effects observed at the final error positions. The PG+PM+RE+RS model exhibited the same virtues and vices as its restricted model counterpart, except that this model predicted recency effects at the final error positions, albeit smaller in magnitude than those seen in the data, consistent with the fits of this model to the data of previous experiments. The mean ratios of fill-in to infill errors predicted by this model, averaged across error locations and scoring procedure, were 2.1:1 and 1.6:1, for seven-item and nine-item sequences, respectively.

Discussion

In short, the distribution of fill-in and infill errors underlying serial reconstruction of spatial stimuli is generally comparable to that observed for verbal stimuli in Experiment 11 and visual stimuli in Experiment 1. The data once again confirm that there are more fill-in than infill errors at each error location throughout the sequence, under both conservative and liberal scoring procedures, and that this empirical outcome is invariant with respect to the sequence lengths examined. The ratios of fill-in to infill errors are larger under conservative scoring than under liberal scoring, and whilst the error ratios are relatively constant across locations under conservative scoring, in comparison the error ratios decrease gradually across locations under liberal scoring. Consistent with the previous experiments there are recency effects on the ratios of fill-in to infill errors at the final error location under both scoring procedures.

The outcomes of the modelling are generally comparable to those reported previously. As before, the combination of positional marking and response suppression failed to predict a fill-in advantage, whilst the combination of a primacy gradient with response suppression generally over-predicted the magnitude of the ratios of fill-in to infill errors, albeit to a lesser extent than was the case for the fits of the same model to the data of Experiment 11. As previous, the confluence of a primacy gradient, positional marking, and response suppression provided a more accurate description of the magnitude and pattern of the ratios of fill-in to infill errors, and the addition of a restricted end marker to this combination of representational principles yielded a further improvement in descriptive accuracy. However, the PG+PM+RE+RS model did not provide as good a description of the data as in the applications of the same model to the verbal and visual data. Specifically, in the current modelling, this model predicted a steep decrease in the ratios of fill-in to infill errors across locations under conservative scoring, which is contrary to the data showing that at pre-recency error locations the ratios of fill-in to infill errors are relatively stable across error locations (notwithstanding some peaks and troughs in the ratios for nine-item sequences).

Nevertheless, this model clearly once again provided the most accurate description of the ratios of fill-in to infill errors of the field of competitor models considered.

General discussion

I begin general discussion by summarising the main empirical findings of the three experiments, before describing the outcomes of the model fitting exercises.

The main pattern of results was strikingly similar across the verbal, visual, and spatial domains. All three experiments consistently revealed more fill-in than infill errors at each possible error location under both conservative and liberal scoring procedures. This empirical outcome was invariant with respect to the sequence length manipulation employed across the three experiments. The ratios of fill-in to infill errors were larger under conservative scoring than under liberal scoring. Under conservative scoring, the ratios of fill-in to infill errors were generally relatively stable across error locations, but exhibited recency effects at the final error locations. Under liberal scoring, the ratios of fill-in to infill errors decreased gradually across error locations, but also exhibited recency effects at the final error locations. These functional similarities in the pattern of fill-in and infill errors across the three short-term memory domains are suggestive of some common underlying principles for representing serial order, a claim which is supported by the outcomes of the quantitative model fitting exercises, which are summarised below.

The results of Experiment 11 using verbal material essentially replicate the outcomes of previous verbal studies in revealing a greater incidence of fill-in than infill errors (e.g., Henson, 1996; Page & Norris, 1998; Surprenant et al., 2005). The current data are most comparable with those of Surprenant et al. (2005), who examined the incidence of fill-in and infill errors across error locations under conservative and liberal scoring procedures for sequences composed of seven-items. They found that fill-in errors outweighed infill errors at all error locations under both scoring procedures. Experiment 11 replicated this result, whilst additionally showing that the preponderance of fill-in errors across error locations holds across different sequence length conditions. The data from Experiments 1 and 12 are more novel. There are no existing data examining the incidence of fill-in and infill errors in the visual domain, and although Guerard and

Tremblay (2008) have examined fill-in and infill errors in a spatial serial recall task they only reported overall levels of errors, noting that fill-in errors were more frequent than infill errors. However, unlike in the current experiments they did not examine the distribution of these errors across error locations.

Turning to the results of the model fitting exercises, although interest centred chiefly on the fits of the models to the ratios of fill-in to infill errors the fits of the models to the transposition matrices were also informative. Across all three experiments, the PG+RS model consistently provided the worst fits to the transposition matrices. Basically this model considerably over-predicted the incidence of adjacent-neighbour transpositions and this further compromised the model's ability to capture the accuracy serial position curves. The PG+PM+RE+RS model consistently provided the best fits, followed by the PG+PM+RS model, which in turn provided better fits than the PM+RS model. The better fits of the former two models were not due to these models being overly complex, because the BIC scores used for the model comparisons trade off the goodness-of-fits of the models with their number of free parameters. Notwithstanding the consistent superior fits of the PG+PM+RE+RS model to the transposition matrices, there was generally little heterogeneity between the predictions of the models for accuracy serial position curves and transpositions gradients. Indeed, the predictions of the PG+PM+RE+RS, PG+PM+RS, and PM+RS models, were for the best part visually indistinguishable. The absence of clear qualitative differences in the predictions of the models for these data prevents confident adjudication between the models.

Considering now the fits of the models to the ratios of fill-in to infill errors, the PG+RS model again consistently provided the worst fits. The PG+PM+RE+RS model once again consistently provided the best fits, followed by the PG+PM+RS model, which in turn provided better fits than then the PM+RS model. Thus, reassuringly the rankings of the models in terms of their goodness-of-fits to the ratios of fill-in to infill errors were the same as for the fits to the transposition matrices. This correspondence notwithstanding, there was greater heterogeneity between the models in terms of their fill-in and infill error predictions. As anticipated, the PM+RS model predicted an approximately equal number of fill-in an infill errors across error locations under conservative scoring. This is because if an item *n* is recalled immediately ahead of its correct position at the next output position the position marker will cue item *n-1* and item *n+1* equally due to the symmetrical manner in which positional uncertainty is distributed in the positional markers. These results confirm the intuitions of previous authors (e.g. Henson, 1996; Page & Norris, 1998) who have noted this shortcoming of symmetrical cue positional models, without however providing a formal demonstration of it. An interesting and novel finding is that under liberal scoring the same model actually predicts more infill errors than fill-in errors (an outcome reminiscent of the predictions of chaining models), a tendency that increases across error locations. This is because sometimes an item will be recalled prematurely in the sequence, before being suppressed, thereby reducing the likelihood that item will be involved in a fill-in error later in the sequence and thus increasing the likelihood of an infill error. To explain, imagine that a participant recalls DBCExxx in response to the sequence ABCDEFG (where x represents a position in which a response has not yet been made). The erroneous recall of the letter E at the fourth output position sets up the possibility of a fill-in or infill error at the fifth output position. The question of interest is whether recall of the letter D (a fill-in error) or recall of the letter F (an infill error) is more probable. In this example, an infill error involving recall of the letter F is more likely, because D has already been recalled at the beginning of the sequence and because it will have been suppressed once emitted this reduces its ability to compete with F at position five following the erroneous recall of E at position four[6]. The probability of an infill error increases across error locations, because as output position increases, it becomes increasingly more likely that an item will have been recalled prematurely earlier in the sequence.

The PG+RS model correctly predicted more fill-in than infill errors, but as anticipated it over-predicted the extent of fill-in errors. This problem was most pronounced in the case of the fits of this model to the verbal data. For those data, the ratio of fill-in to infill errors, averaged across error

[6] For clarity, under conservative scoring this could not happen, because fill-in and infill errors are only recorded if all responses prior to the location of the error were correct.

locations and scoring procedure, for five-item, six-item, and seven-item sequences, were 38:1, 10:1 and 4:1, respectively, compared to empirically observed values of 3.57:1, 2.12:1, and 1.99:1. It is apparent that although the mean ratio for seven-item sequences is not far off the empirically observed mean value, the mean ratios for five-item and for six-item sequences are distinctly larger than seen empirically. Non-surprisingly, given the exponential nature of the primacy gradient the ratios of fill-in to infill errors decreased exponentially across error locations. Thus, the ratios were typically considerably larger for the first couple of error locations than for subsequent locations due to the much greater disparity in activations of the first few items in the sequence. For instance, for the verbal data, the mean ratios of fill-in to infill errors for error locations one and two for five-item sequences, were 81:1 and 22:1, compared to empirical values of 3:1 and 2:1. Note that the decrease in the error ratios across locations predicted by this model under conservative scoring is contrary to the data showing that the ratios of fill-in to infill errors are relatively constant across error locations. The PG+RS model additionally failed to predict much difference between the ratios of fill-in to infill errors for the two scoring procedures. That is, the predicted error ratio curves for conservative and liberal scoring sit virtually on top of one another, which is contrary to the larger error ratios for conservative scoring witnessed empirically.

The simulations of the PG+RS model reported here, and in the previous chapters, are based upon an exponential primacy gradient. Accordingly, one question is whether a version of the PG+RS model based upon a linear primacy gradient might be better able to accommodate the distribution of fill-in and infill errors. The predictions of a linear version of the PG+RS model, implemented according to equation 2-5 of Chapter 2, are presented in Appendix 3. Figures A3-1, A3-2, and A3-3 show the accuracy serial position curves, transposition gradients, and ratios of fill-in to infill errors predicted by this model after fitting to the transposition matrices of Experiment 11, 1, and 12 using the same fitting procedure described above. It is apparent from inspection of these figures that this version of the PG+RS model predicted even less realistic serial position curves and transposition gradients than the version instantiating an exponential primacy gradient. Crucially, the linear version of the PG+RS model also over-predicted the extent of fill-in errors, albeit to a lesser extent than the exponential version. For example, the mean ratios of fill-in to infill errors predicted by the linear version of the model, applied to the verbal data of Experiment 11, were 27:1, 7:1, and 4:1, for sequences of five-items, six-items, and seven-items, respectively. The corresponding mean ratios predicted by the exponential version of the model were 38:1, 10:1, and 4:1. It can also be seen that the linear version of the PG+RS model generates qualitatively different error ratio curves than the exponential version. Specifically, the linear version predicts that under conservative scoring the ratios of fill-in to infill errors are constant across error locations. Notwithstanding these differences between the two different versions of the PG+RS models, it is clear that both over-predict the extent of fill-in errors.

The inability of primacy gradient and symmetrical cue positional models to predict the correct level of fill-in has already been documented by previous authors (e.g., Henson, 1996; Page & Norris, 1998). However, as mentioned in the introduction, to my knowledge these claims have never been substantiated by formal simulations of the models to highlight the extent of their shortcomings. The current modelling exercises provide formal confirmation that symmetrical cue position marking models (e.g., Brown et al., 2000; Burgess & Hitch, 1992) fail to predict more fill-in than infill errors. Indeed a novel result of the simulations was to show that when these errors are measured using a liberal scoring procedure, such models actually predict more infill than fill-in error. The results additionally verify that primacy gradient models do indeed considerably over-predict the extent of fill-in errors. Furthermore, the modelling outcomes demonstrate that these models, as well as being unable to accommodate the pattern of fill-in and infill errors in the verbal domain, are also unable to accommodate the pattern of these errors in the visual and spatial domains, which generally follow a similar profile.

The central objective of the current chapter was to establish whether the combination of a primacy gradient, positional marking, and response suppression, which has proved so successful in modelling the dynamics of transpositions in the previous chapters could additionally accommodate the pattern of fill-in and infill errors across the verbal, visual, and spatial domains. The results of the modelling exercises confirm that this combination of representational principles provides a significantly better account of the ratios of fill-in to infill errors than the combination of a primacy gradient and response suppression or position marking and response suppression. The PG+PM+RS model predicted ratios of fill-in to infill errors of an overall magnitude similar to those seen empirically in all three domains and correctly predicted the larger ratios observed under conservative than under liberal scoring. In the case of the verbal and visual domains, this model also provided a satisfactory account of the pattern of error ratios seen across error locations under the two scoring procedures, notwithstanding its failure to predict the observed recency effects. However, in the case of the spatial data, the model was less successful in capturing the pattern of error ratios seen across error locations. The model predicted a steep decrease in the error ratios across error locations under conservative scoring when in fact these ratios were relatively constant, notwithstanding the observed recency effects, which the model again failed to accommodate. The model did, however, correctly predict the decrease in the ratios observed across error locations under liberal scoring.

The better descriptive accuracy of the PG+PM+RS model arose because the impact of combining the primacy gradient with a set of positional markers is to reduce the disparity in activation between neighbouring items, whilst still retaining the asymmetric bias towards earlier items in the sequence that is necessary to accommodate fill-in. This means that if item *n* is recalled a position ahead of its correct position then item *n-1* will only be a slightly stronger competitor at the next output position than item *n+1*. Thus, a greater incidence of fill-in errors is still predicted, but the ratio of fill-in to infill errors is reduced considerably relative to that predicted by the combination of a primacy gradient and response suppression alone.

The PG+PM+RE+RS model built on the above model by incorporating a restricted end marker in an attempt to model the anticipated effects of recency on the ratios of fill-in to infill errors – motivated by the observation of such effects in the data of Surprenant et al. (2005). This model enjoyed the same benefits as its restricted model counterpart, the PG+PM+RS model, whilst additionally predicting effects of recency on the ratios of fill-in to infill errors. The action of the restricted end marker was to increase the distinctiveness of the final item in the sequence, thereby protecting it from being anticipated and becoming involved in an infill error. Notwithstanding the added benefits of the restricted end marker, the effects of recency predicted via its incorporation were generally smaller in magnitude than those observed empirically, and with the exception of the fits to the visual data of Experiment 1, these effects were restricted to the predictions for conservative, but not liberal scoring, which is contrary to the data showing effects of recency under both scoring procedures. It is important to bear in mind, however, that these predictions were obtained by fitting the models to data that do not contain any information about the distribution of fill-in and infill errors. Better fits would no doubt be possible if model parameters were estimated by incorporating information about the distributions of these errors. This strategy was not adopted here due to the danger of overfitting the data.

The decision to incorporate the restricted end marker as opposed to some other process for modelling effects of recency merits comment. The empirical justification for incorporating this assumption originates from the modelling work of Farrell and Lelievre (2009) mentioned previously in Chapter 7 showing that a restricted end marker contributes to verbal serial recall. The evidence for this comes from the pattern of terminal between group transpositions when sequences are organized into temporal groups of unequal sizes. It is possible that the recency effects seen in the fill-in ratios are attributable to a recency-based process other than a restricted end marker, since any process that renders the final list-item more salient would predict a recency effect on the ratios of fill-in to infill errors. However, only the restricted end marker would appear to predict the error patterns in grouped sequences alluded to above. It seems more likely therefore that in the case of the verbal data the observed recency effects are indeed attributable to the action of a restricted end marker. The case for a restricted end marker is weaker in the visual and spatial domains, because direct evidence in the form of the errors mentioned above has not been demonstrated with these types of material. Nevertheless, on grounds of parsimony, it would seem prudent to assume that the same process contributed to recency across the three domains. In any case, the clear message provided by the outcomes of the modelling exercises reported here is that some additional process is necessary to accurately model the observed recency effects on the ratio of fill-in to infill errors across the three domains.

One criticism that might be levelled at the current model fitting exercises is that unlike the model fitting of previous chapters, in which models were applied to individual participant data, in the current chapter the models were applied to aggregate data. It must be emphasised at this juncture that the fitting of models to group level data is in fact the standard approach employed by cognitive modellers (Estes, 2002). Nevertheless, the convenience of fitting models to averaged data can sometimes come at a cost, because the empirical trend seen in the averaged data will not always be representative of the trends observed across the individual participants whose data was averaged (see e.g., Estes, 2002; Estes & Maddox, 2005). Ultimately, the only way to verify whether this problem applies to the current model fitting exercises will be to replicate them, but this time fitting to individual participant data. However, I consider it unlikely that fitting the models in this way will alter the conclusions reached in this chapter. This is because in separate work to this thesis I applied the current batch of models to aggregate and individual participant data taken from the study of Surprenant et al. (2005; Experiment 1) and found that the core predictions of the models and their rankings based upon their goodness-of-fits to the transposition matrices and ratios of fill-in to infill errors did not diverge under the two fitting procedures. Similarly, before fitting models to the individual participant data of Farrell and Lewandowsky (2004) in Chapter 3, and the data of Experiments 2 and 9 of Chapters 4 and 5, respectively, I first applied those models to the aggregate data. These simulations also returned model predictions that did not qualitatively diverge from the fits of the models to the individual participant data. I conclude therefore that the application of the models to the aggregate data is not a major cause for concern.

In summary, the main contribution of the current chapter has been to provide converging evidence for the role of a primacy gradient, positional marking, and response suppression in verbal, visual, and spatial serial memory. Combined with the transposition latency data and associated modelling of Chapters 3, 4, 5, and 6 the current findings suggest that all three representational principles must coexist in any adequate model of serial order in these three domains. At the same time, the findings once again confirm that the accurate modelling of serial order in these domains cannot be accomplished via a model in which serial order is represented solely on the basis of a primacy gradient or positional marking. Added to this, the current results lend tentative support for the role of a fourth representational principle: namely a restricted end marker.

Chapter summary

Three experiments examined the distribution of fill-in and infill errors underlying serial reconstruction of sequences of verbal (Experiment 11), visual (Experiment 1), and spatial (Experiment 12) stimuli. All three experiments consistently revealed more fill-in than infill errors across all possible locations at which these errors can occur under both a conservative and a liberal scoring procedure. These effects were insensitive to the sequence length manipulation employed across experiments. Application of four models of serial order to the data from the three experiments consistently revealed that neither the combination of a primacy gradient with response suppression, nor the combination of positional marking with response suppression provided an adequate account of the ratios of fill-in to infill errors. The union of all three representational principles provided a satisfactory description of the magnitude and pattern of the ratios of fill-in to infill errors across the verbal, visual, and spatial domains, and this description was further enhanced via the incorporation of the ancillary assumption of a restricted end positional marker.

9

Conclusions

Introduction

The aim of this thesis has been to elucidate the principles underlying the representation and generation of serial order in short-term memory for visuospatial information. Specifically, this thesis sought to identify what combination of the explanatory principles that have been identified as contributing to the representation of serial order in verbal short-term memory are involved in the representation of serial order in visuospatial short-term memory.

The outcomes of the empirical and computational modelling work indicate that the serial order of a sequence of visual, spatial, or verbal items is represented on the basis of a competitive queuing sequence planning and control mechanism, equipped with a primacy gradient of activation, associations between items and positional markers, and suppression of recalled items. Tentative support was additionally obtained for the contribution of a restricted end positional marker. The outcomes of the current thesis therefore buttress the notion that verbal and visuospatial short-term memory depend upon at least some common principles for representing serial order. In the next section, I summarise the evidence for the inferred explanatory principles.

Evidence for inferred principles

In Chapter 1, I presented the empirical and theoretical precedents for the proposal that the functional similarities between verbal and visuospatial serial memory are attributable to their common reliance on a competitive queuing (CQ) sequence planning and control mechanism. In this section, I briefly delineate the evidence obtained in this thesis for the role of a primacy gradient, positional marking, response suppression, and a restricted end marker in representing serial order information in visual, spatial, and verbal serial memory. This evidence stems from four empirical and modelling sources: (1) the dynamics of transpositions, (2) temporal grouping effects, (3) the distribution of fill-in and infill errors, and (4) the infrequency of erroneous repetitions.

Dynamics of transpositions

In Chapter 3, I demonstrated that five different models and associated mechanisms for representing serial order that cannot be distinguished on the basis of their accuracy and latency serial position curves and transposition gradients can nevertheless be distinguished on the basis of their predicted latency-displacement functions (LDFs). I then fit a sub-set of the models to verbal serial recall data taken from Farrell and Lewandowsky (2004). The LDFs for these data exhibit a negative slope for anticipations and a shallow positive slope for postponements. The results of the quantitative model fitting exercise revealed that only the combination of a primacy gradient, positional marking, and response suppression (PG+PM+RS) predicted this empirical pattern.

In Chapters 4 and 5, I examined the pattern of transposition latencies underlying serial reconstruction of sequences of visual (Experiments 1-6) and spatial items (Experiments 7-9), respectively. These experiments consistently revealed LDFs characterised by negative anticipation slopes and either flat or shallow positive postponement slopes, again most compatible with the prediction of the PG+PM+RS model. Fits of models to representative visual serial reconstruction data taken from Experiment 2, and representative spatial serial reconstruction data taken from Experiment 9 revealed that, only this model can accommodate the empirical pattern of the observed LDFs when model parameters are estimated from the behavioural data.

In Chapter 6, I presented parameter space sensitivity analyses of the predictions of the models of Chapter 3. The results of these analyses showed that the flat/shallow-positive postponement slopes predicted by the PG+PM+RS model are a robust feature of its wider behaviour. This suggests that the superior description of the LDF provided by this model is not attributable to it being overly complex.

Temporal grouping effects

In Chapters 5 and 7, I examined the impact of temporal grouping on serial reconstruction of sequences of spatial (Experiments 7, 9, & 10), and verbal (Experiment 10) items. Relative to an ungrouped baseline, temporal grouping exerted a multiplicity of effects on spatial and verbal serial reconstruction performance, including an elevation in recall accuracy; a scalloping of the accuracy and response latency serial positions curves; and a reduction in transpositions between groups. The results of these experiments provide further support for the role of positional marking in spatial serial memory. Moreover, they support the notion that both grouped spatial and verbal sequences are represented hierarchically, with one level of representation coding the positions of groups, and a second level of representation coding the positions of items. I consider the effects of temporal grouping in further detail below.

Fill-in and infill errors

In Chapter 8, I examined the distribution of fill-in and infill errors underlying serial reconstruction of verbal, visual, and spatial sequences varying in length (Experiments 11, 1, & 12). All three experiments revealed more fill-in than infill errors, not only overall but at each possible error position throughout the sequence, for all sequence lengths examined. The ratios of fill-in to infill errors were larger under a conservative scoring procedure than under a liberal scoring procedure. Under conservative scoring, the ratios of fill-in to infill errors were generally relatively stable across error positions, but exhibited recency effects at the final error positions. Under liberal scoring, the ratios of fill-in to infill errors decreased gradually across error positions, but also exhibited recency effects at the final error positions.

Applications of four models of serial order to the data from each experiment revealed that neither a model incorporating a primacy gradient in conjunction with response suppression, nor a model incorporating positional marking in conjunction with response suppression, could accommodate the above empirical patterns. A model combining all three explanatory principles provided a satisfactory account of the pattern of results, and this account was further enhanced by the incorporation of a restricted end positional marker to model the observed effects of recency.

Erroneous repetitions

In Chapters 4 and 5, I examined the incidence of erroneous repetitions in the reconstruction of visual (Experiment 3), and spatial (Experiments 7, 8, & 9) sequences. These experiments revealed that erroneous repetitions are extremely rare, typically accounting for less than 1% of all responses, which is well below that expected by chance. The scarcity of erroneous repetitions is consistent with the notion that once an item is recalled its representation is suppressed thereby reducing the likelihood that item will be recalled again. Qualified support for the role of response suppression was provided by applications of five models of serial order to the ungrouped spatial serial reconstruction data of Experiment 9, which revealed that the position marking (PM) model, which was the only model not to incorporate response suppression, predicted a distinctly larger proportion of erroneous repetitions than the four remaining models.

Limitations

The evidence reviewed above is generally consistent with the notion that common principles represent serial order across the verbal, visual, and spatial domains of short-term memory. Nevertheless, there are some limitations of this evidence that merit comment.

First, unlike verbal serial memory, visual serial memory is apparently insensitive to effects of temporal grouping, as shown in Experiment 4. Although this experiment did reveal an enhancement in recall accuracy for grouped relative to ungrouped sequences, this improvement was only minor and accompanied by the absence of any statistically detectable effects of the grouping manipulation on the latency serial position curve and the incidence of between group interposition errors. At a cursory glance, this would appear to suggest a fundamental difference in the representation of positional information in verbal and visual serial memory. However, as emphasised in Chapter 4, the limited impact of the temporal grouping manipulation is likely attributable to the absence of an output mechanism for rehearsing visual, non-spatial information. This is pertinent, because there is evidence to suggest that rehearsal plays a pivotal role in the manifestation of grouping effects in verbal short-term memory, at least when the to-be-remembered material is presented visually. Specifically, under visual, but not auditory presentation conditions, the magnitude of effects of temporal grouping increases as a function of the length of the pause delimiting groups (Frankish, 1989) – with longer pauses presumably conferring extra time for intra-group rehearsal. Moreover, when rehearsal is prevented by having participants engage in articulatory suppression during encoding, the effects of temporal grouping are abolished under visual, but not auditory presentation conditions (Hitch et al., 1996; Hurlstone, 2006).

Second, although reliable effects of temporal grouping were observed for spatial serial reconstruction (Experiments 7, 9, & 10), including an elevation in recall accuracy, a change in recall latencies, and a reduction in between group transpositions, there was no evidence for a corresponding increase in interposition errors. Although these data are consistent with the notion that positional information in spatial serial memory can be organised on multiple dimensions, with one dimension coding the positions of groups and a second dimension coding the position of items, there is some uncertainty concerning the nature of the positional representations employed at the item level. Specifically, the absence of interpositions would appear to imply that unlike items in temporally grouped verbal sequences, items in temporally grouped spatial sequences are not coded for their within-group positions, but instead for their positions within the sequence overall. However, applications of two computational models to the grouped verbal and spatial data of Experiment 10, one assuming position in sequence representations of items (the GP-SP model), and one assuming position within-group representations of items (the GP-WP model), revealed that this assumption is unfeasible. Instead, the modelling outcomes revealed that the data from both domains are best explained by a common model assuming position within-group representations. Of course, the drawback to this conclusion is that the GP-WP model predicted a slight elevation in interpositions that was not witnessed empirically after it was fit to the spatial data. To explain this disparity between the predictions of the GP-WP model and the spatial data, I argued that the failure to detect an elevation in interpositions empirically might be attributable to the spatial organization of items interacting with the temporal organization of items and influencing the pattern of transposition errors (such an interaction could not occur in the models, because they were only given information about the temporal organization of items). Whether this assumption is valid is an empirical question that requires further investigation. However, given the poor fit of the GP-SP model to the grouped spatial data and the absence of an obvious alternative account of the data, on balance the current evidence more strongly favours the assumption that position within-group representations are a feature of grouped sequence representations in the spatial domain. In any case, the crucial point to emphasise is that the current results show that people do use positional information to encode and recall spatial sequences, and that this information can be organised hierarchically, as is known to be the case for verbal sequences.

A third limitation, concerns the failure to obtain more direct evidence for the role of response suppression in visual and spatial serial memory. The main evidence for response suppression in this thesis comes from the occurrence frequency of erroneous repetition errors. Experiments 7, 8, and 9 consistently revealed that the incidence of erroneous repetitions in the serial reconstruction of spatial stimuli is well below that expected by chance. This is consistent with the notion that once an item was recalled its memorial representation was inhibited thereby reducing the likelihood that it would be chosen again. The incidence of erroneous repetitions was also rare and below that expected by chance in visual serial reconstruction, however, the evidence in this case was more limited, since of the six experiments conducted employing visual stimuli only Experiment 3 incorporated a reconstruction condition in which repetition errors were possible. The main limitation of using erroneous repetitions as an index of response suppression is that the low incidence of these errors can also be explained by assuming that participants maintain a memory for the items that they have already chosen during the recall process. In the section below on future directions, I consider some approaches for obtaining more direct evidence for the role of response suppression in visual and spatial serial memory.

Finally, one representational principle implicated in verbal serial memory for which the current thesis has not identified evidence for in the visual and spatial domains is output interference. However, in fairness, this thesis has not made a concerted effort to test for its involvement. Indeed, the only window in which to observe the operation (or lack thereof) of output interference was in terms of a comparison of the empirically observed LDFs with the error latency prediction of a model built from the combination of positional marking, output interference, and response suppression (PM+OI+RS). This model provided a poor account of the LDFs observed in visual, spatial, and verbal serial memory and the data were best explained instead by the PG+PM+RS model. However, this does not preclude the possibility that a generalised version of the PG+PM+RS model augmented with output interference (e.g., PG+PM+RS+OI) might provide an enhanced account of the data (on grounds of parsimony such a model was not examined in the current thesis). In verbal serial memory, the evidence for the operation of output interference stems from experiments in which input and output position were deconfounded in serial recall by having participants start their recalls at different sequence positions (Cowan et al., 2002; Oberauer, 2003). Such studies have shown that recall accuracy decreases over the output positions of items, which is consistent with a contribution of output interference to the genesis of the primacy effect. However, in the experiments of the current thesis input and output position were always confounded so the observed primacy effects could be attributable to output interference, a primacy gradient of activations, or both. The absence of a direct test of output interference in the current thesis is defensible on the grounds that it has been concerned primarily with the principles that contribute to the representation of serial order in the visual and spatial domains. Output interference does not constitute a representation of serial order, but rather an ancillary assumption that has proved necessary in certain models of verbal serial recall to adequately account for effects of primacy and sequence length (e.g., Brown et al., 2000; Lewandowsky & Farrell, 2008; Lewandowsky & Murdock, 1989).

Modularity and serial order in short-term memory

Given the evidence of common principles for representing serial order in verbal, visual, and spatial short-term memory cited above, an obvious question concerns whether the mechanisms that instantiate these principles are modular, or non-modular. That is to say, is there a single mechanism underlying each principle that is shared between the verbal and visuospatial domains, or are there multiple mechanisms across domains instantiating common principles? The answer to this question has obvious implications for the interpretation of the current results within multi-component models of short-term memory, such as the working memory model (Baddeley, 1986, 2000; Baddeley & Hitch, 1974) introduced at the outset of Chapter 1. Recall that the working memory model comprises two short-term memory sub-systems: a *phonological loop* for the storage of verbal information and a *visuospatial sketchpad* for the storage of visuospatial information. A third component known as the *central executive* is responsible for coordinating the activities of the two short-term memory sub-systems, whilst a fourth component known as the *episodic buffer* is responsible for integrating information from the phonological loop and visuospatial sketchpad.

Returning to the question introduced above concerning the modularity of the mechanisms of serial order in verbal and visuospatial short-term memory, there are in fact two aspects to this question. The first is to what extent are the mechanisms of serial order separate for verbal and visuospatial information? The second is to what extent are the mechanisms of serial order separate for visual and spatial information? The answers to these questions depend for the most part on architectural constraints underlying the implementation of the different mechanisms of serial order. Below I consider these constraints with reference to each of the putative mechanisms. I also attempt to map the different mechanisms of serial order onto the different components of the working memory model.

Competitive queuing

I consider it most likely that separate CQ mechanisms are involved in the planning and control of verbal, visual, and spatial sequences. This is because the parallel planning layer in the CQ mechanism essentially constitutes an activation-based short-term memory buffer. Thus, given the compelling evidence for distinct verbal and visuospatial memory sub-systems (reviewed in Chapter 1), as well as the evidence for distinct visual and spatial sub-systems (also reviewed in Chapter 1), this seems to argue against a shared CQ mechanism. Another reason for supposing separate CQ systems is that the CQ mechanism has been invoked to explain sequential behaviour in a variety of serial performance domains, including typing (Rumelhart & Norman, 1982), speech production (Dell, 1986; Dell et al., 1997; Hartley & Houghton, 1996; Houghton, 1990; Bohland et al., 2009), sequence learning (Rhodes & Bullock, 2002), spelling (Glasspool & Houghton, 2005; Glasspool et al., 1995, 2006; Houghton et al., 1994), saccade generation (Brown, Bullock, & Grossberg, 2004), action planning (Cooper & Shallice, 2000), and of course short-term memory (Burgess & Hitch, 1992, 1999, 2006; Henson, 1998; Page & Norris, 1998, 2009). It seems highly unlikely that these diverse serial behaviours are all mediated by a domain general CQ mechanism. Instead, it seems more likely that the brain has developed specialised CQ systems to solve the problem of serial order in different domains. In terms of the working memory model, these considerations suggest that the phonological loop and visuospatial sketchpad each possess their own dedicated CQ sequence planning and control mechanisms, with the possibility that the latter sub-system contains two distinct CQ systems, one for the generation of visual sequences and one for the generation of spatial sequences. The notion that the working memory slave systems operate as CQ mechanisms is not without precedent: both of the computational models of the phonological loop, namely the Primacy model (Page & Norris, 1998, 2009) and the Burgess and Hitch (1999, 2006) model, postulate that it operates as a CQ mechanism.

Primacy gradient

Similar comments also apply to the primacy gradient mechanism. In order to code serial order information via a primacy gradient of activation levels that primacy gradient must necessarily be built up over the item representations in an activation-based short-term memory buffer, namely the parallel planning layer of the CQ system. Given the arguments above for distinct buffer storage and CQ systems in short-term memory, this suggests that separate primacy gradients represent serial order information in the phonological loop and the visuospatial sketchpad (with the possibility again that separate primacy gradients represent the serial order of visual and spatial sequences in the latter sub-system). Of course, a primacy gradient can be implemented in ways other than over the item representations in a short-term memory buffer. An alternative way of implementing the primacy gradient, which is more consistent with the notion of a domain general mechanism, is in terms of a decrease in the strength of the weights linking positional markers to item representations across serial positions, as is assumed in some models of serial recall (e.g., Brown et al., 2000; Lewandowsky & Farrell, 2008). Such a mechanism might be mediated by a common novelty-gated encoding process like that envisaged in the SOB model of short-term memory (Lewandowsky & Farrell, 2008). However, this approach to implementing the primacy gradient is incompatible with the transposition error latency and fill-in and infill error data reported in this thesis. Specifically, it can be shown that such a mechanism predicts steep positive LDF postponement slopes and an approximately equal number of fill-in and infill errors, both of which are contrary to the data.

Response suppression

Turning to response suppression, one idea is that a common mechanism mediated by an executive control system – the central executive component of the working memory model – is responsible for inhibiting the representations of recalled items in the working memory slave systems. This comes with the additional corollary that response suppression is to a large extent a strategic and wilfully controlled act. The alternative possibility, and my preferred interpretation, is that response suppression is implemented locally within the working memory slave systems by the CQ mechanisms controlling sequence planning and recall. In CQ models, when an item wins the output competition in the competitive choice layer, a large amplitude signal is generated to initiate recall of that item and to inhibit its representation in the parallel planning layer (Bullock, 2004; Bullock & Rhodes, 2003). Thus, according to the CQ model, response suppression is an automatic and obligatory process that is not under volitional control. That response suppression is an automatic, as opposed to an executive controlled process, is supported by the finding of Henson (1998b) that participants are extremely poor at recalling the second instance of a repeated item (*viz.* repetition inhibition; see Chapter 1) even when those repetitions are detected with a very high level of accuracy (85%).

Positional marking

There are less architectural constraints on the mechanism responsible for coding positional information in short-term memory. Unlike the mechanisms discussed above, the positional context signal assumed in CQ models of serial recall (e.g., Brown et al., 2000; Burgess & Hitch, 1999, 2006; Henson, 1998a) is essentially external to the CQ system itself. Because this context signal represents serial order information independently of item information, unlike the primacy gradient mechanism in which item and order information is conjunctively coded, it is possible that a common context signal might code positional information across different short-term memory domains. Indeed, Burgess and Hitch (1999) have suggested that the positional context signal in their network model of the phonological loop might also be responsible for coding the position of nonverbal items.

In terms of the working memory model, one speculation is that the positional context signal maps onto the episodic buffer component (Baddeley, 2000). In keeping with the proposed binding function of the buffer this would allow the same context signal to be flexibly associated with items from different modalities and cognitive domains. It would also provide a basis for encoding two sequences from different domains in parallel (e.g., a verbal and a spatial sequence), because it should be possible for items in the different sequences to be simultaneously bound to the same context signal. It is of course possible that multiple context signals exist that are specialised for representing positional information in different domains. Although this possibility is less attractive on grounds of parsimony it should not be ruled out, since research employing CQ models has shown that the characteristics of the positional context signal varies across serial performance domains. For example, Glasspool (2005) has noted that the positional context signal underlying CQ models of spelling differs from that underlying CQ models of verbal short-term memory, and that these in turn differ from the context signal underlying CQ models of speech production. It follows therefore that the possibility should not be dismissed that specialised context signals might also code positional information across different short-term memory domains.

Indeed, there is evidence from dual-task interference studies and neuropsychological patient dissociations that is consistent with the notion of modality-dependent positional context signals. For example, studies have shown that verbal and spatial serial recall are susceptible to interference from different kinds of secondary tasks (Guerard & Tremblay, 2008; Lange, 2005; Meisser & Klauer, 1999). For example, articulatory suppression and irrelevant background speech have been shown to interfere with verbal serial recall, but not spatial serial recall, whereas the converse is true for the secondary task of manual spatial tapping. Neuropsychological patients have also been identified that exhibit preserved verbal serial recall in conjunction with impaired spatial serial recall, whist other patients have been identified that exhibit the reverse pattern of preservation and impairment (De Renzi & Nichelli, 1975; Hanley et al., 1991; Vallar & Baddeley, 1984). Of course, the main obstacle in interpreting these data as evidence for modality-independent positional context signals is that these double dissociations may relate to some other mechanism(s) of serial order. For example, they might reflect selective interference or impairment of the primacy gradient or competitive queuing systems of the phonological loop and visuospatial sketchpad, leaving open the possibility of a common positional context signal.

Such an interpretation is given some credence by a recent interference study by Depoorter and Vandierendonck (2009), which has provided some evidence for a modality-independent positional context signal. These authors showed that performance on a verbal and a visuospatial order memory task were both disrupted by an embedded order memory task. The critical difference between this study and those reported above is that both the primary and secondary tasks in the study of Depoorter and Vandierendonck required memory for the order of a sequence of events. In contrast, in the above studies the secondary tasks, although possessing a serial element, did not involve a concurrent order memory demand. This distinction is important, because a secondary task that also requires memory for order is arguably more likely to recruit the positional context signal than a task that contains a serial element, but does not require memory for order. The results of Depoorter and Vandierendonck thus provide some tentative evidence for the notion that a common positional context signal represents serial order in the verbal and visuospatial short-term memory domains. This possibility is given further credibility by the finding from the same study that an embedded item memory task, which did not require memory for order and therefore could not have recruited the positional context signal, produced considerably less interference on the verbal and visuospatial primary memory tasks than the embedded order memory task.

An important objective for future research will be to obtain further evidence that casts light on whether or not a common context signal represents positional information across the verbal and visuospatial domains. In the section on future directions, I consider some novel approaches to tackling this issue.

Contribution to literature: Empirical and theoretical

In this section, I consider the empirical and theoretical contributions of this thesis to the short-term memory literature.

Empirical

At the outset of this thesis, I argued that the existing empirical database does not permit identification of a preferred combination of explanatory principles for representing serial order in visuospatial short-term memory. This is because the existing data can be accommodated by various different combinations of the explanatory principles delineated in Chapter 1. The main intellectual basis for this claim is the modelling work presented in Chapters 3, 4, 5, and 8, showing that models built from different combinations of representational principles predict qualitatively similar accuracy serial position curves, transposition gradients, and latency serial position curves. The empirical contribution of this thesis has been its focus on constraints that can serve to distinguish the predictions of different mechanisms for representing serial order. These constraints include the dynamics of transposition errors, effects of temporal grouping, and fill-in and infill errors. In conjunction with computational modelling work, the empirical patterns observed across these different constraints has identified a core set of principles for representing serial order in visuospatial memory that cannot be inferred on the basis of extant published data.

A core aspect of this thesis has involved the analysis of transposition latencies in verbal, visual, and spatial serial memory. Indeed, nine of the twelve experiments (Experiments 1-9) were specifically targeted towards examining the dynamics of transposition errors. The outcomes of these experiments, combined with those of Farrell and Lewandowsky (2004), have consistently shown that transposition latency is a negative function of transposition displacement, but with a reduction in the slope of the function for postponements compared to anticipations. The consistency of this empirical pattern, combined with its diagnosticity in discriminating between different mechanisms of serial order, identifies the LDF as a core benchmark that theories of serial memory in the verbal and visuospatial domain must accommodate.

Theoretical

At the end of Chapter 1, I noted the absence of any theories of serial order in visuospatial short-term memory, despite the now well-developed empirical database and extensive functional similarities with verbal serial order memory. In this thesis, I have attempted to bridge this theoretical gap. The first step in building this bridge was the proposition that the functional similarities between verbal and visuospatial serial memory are attributable to their common reliance on a CQ sequence planning and control mechanism. The benefit of this proposal is that it provides an interpretation of serial order across the two domains couched within a general theory of sequence generation. The data and modelling from subsequent chapters have provided fundamental constraints on the representation of serial order within the visuospatial and verbal CQ systems. Specifically, they provide support for a representational mechanism combining a primacy gradient of activation, associations between items and positional markers, and suppression of emitted items. In this chapter, I have also attempted to map these inferred explanatory mechanisms onto the different components of the broader working memory framework. I believe that to move from a position of complete lack of theory at the beginning of this thesis to the current theoretical framework represents considerable theoretical progress in tackling the problem of serial order in visuospatial short-term memory.

The theoretical insights offered by this thesis could not have been realized without the assistance of computational modelling. In particular, they owe a great deal to the componential approach to modelling adopted, which was inspired by the work of Simon Farrell (Farrell & Lelievre, 2009; Farrell & Lewandowsky, 2004). This approach involved contrasting models built from different combinations of the mechanisms widely employed in models of verbal short-term memory within a common modelling environment in order to tease apart their predictions. Given the complexity of some of the models and mechanisms under comparison this would not have been possible on the basis of verbal theorizing alone. Computational modelling has also played an important role in this thesis in testing novel ideas. A prime example of this is the simulation work reported in Chapter 7 evaluating the hypothesis that in spatial serial memory the absence of interposition errors when sequences are temporally grouped is due to a reliance on group position and position within sequence representations of order, rather than group position and position within-group representations of order. Although this proposal seems perfectly plausible at the verbal level, the model fits of Chapter 7 clearly indicate that such a model is a poor theory of the data. This thesis therefore highlights some of the virtues of a modelling-based approach. Specifically, it is only through the construction of formal models that the predictions of a theory can unambiguously be explored.

It is noteworthy that computational modelling has assumed a central role in theorizing about verbal short-term memory, yet surprisingly the current work constitutes the first attempt to apply computational models to visuospatial short-term memory. Furthermore, it is one of but only a handful of studies more generally to have evaluated and compared models on a competitive basis (see Farrell, 2006; Farrell & Lelievre, 2009; Farrell & Lewandowsky, 2004; Oberauer & Lewandowsky, 2008 for further examples), as opposed to evaluating the explanatory power of a single model, as is the standard practice. It is hoped that this work will inspire other researchers in this area to employ computational models for exploring and testing ideas about visuospatial short-term memory, preferably in the spirit of the current modelling work, by comparing multiple rival models and mechanisms.

A further theoretical contribution of this thesis has been to demonstrate the diagnosticity of response times in short-term memory. The time taken to generate a response is a ubiquitous measure in experimental psychology, but one that has seldom been employed in studies of short-term memory. Consequently, the majority of models of verbal short-term memory only generate predictions for response probabilities (Brown et al., 2000, 2007; Botvinick & Plaut, 2006; Burgess & Hitch, 1999, 2006; Henson, 1998; Page & Norris, 1998, 1999; although see Anderson & Matessa, 1997; Anderson et al., 1998; and Lewandowsky & Farrell, 2008 for exceptions), leaving unconsidered the temporal dynamics of the recall process. This neglect is noteworthy, because this thesis has shown that rival models of short-term memory are often indistinguishable when the dependent measure upon which they are contrasted is the probability of a response. However, the models can be effectively distinguished when the time taken to generate an error response is the fundamental dependent measure. It is hoped that this contribution will elevate the prominence of response times in the field of short-term memory, both at the theoretical and empirical levels.

Future directions: Empirical and modelling

In this section, I consider some of the future empirical and modelling directions of the work reported in this thesis.

Empirical

Direct tests of response suppression

One important objective for future research is to obtain direct evidence for the role of response suppression in visual and spatial serial memory. The clearest evidence for the operation of response suppression in verbal serial recall comes from the phenomenon of repetition inhibition (often referred to as the "Ranschburg effect"). Recall from Chapter 1 that this refers to the finding that when people are presented with sequences containing a repeated item recall of the second instance of the repeat is retarded following recall of the first instance (Crowder, 1968; Duncan & Lewandowsky, 2005; Henson, 1998b; Jahnke, 1969; Vousden & Brown, 1998). This behavioural outcome is predicted by response suppression, because the inhibition of an item once it has been recalled will render it difficult for that item to win the output competition again later in the sequence.

Recently, Farrell and Lewandowsky (2007) have identified another means of indexing the contribution of response suppression in serial recall. They exploited the fact that models of serial recall that invoke response suppression as an explanatory construct predict that it is a contributor to the recency effect. Farrell and Lewandowsky examined the accuracy of recall of the final item on sequences containing two errors that occurred in all but the last serial position. They examined how the recency effect is modulated by three different combinations of errors, either: (1) two transpositions, (2) one transposition and one extra-list intrusion, or (3) two intrusions. In all three instances two errors are committed, but in (1) all the items of the sequence have been suppressed, whereas in (2) the single extra-list intrusion means that one of the items in the sequence remains unsuppressed, whilst in (3) the two intrusions mean that two items in the sequence remain unsuppressed. Farrell and Lewandowsky found that the magnitude of the recency effect was a function of the number of items that have been suppressed (i.e., strongest under condition (1), followed by conditions (2) and (3)). This outcome is consistent with the predictions of a response suppression account of the recency effect.

Future work will examine the effects of a repeated item on visual and spatial serial reconstruction and the sensitivity of the recency effect in these tasks to the number of items that have been reported, and hence suppressed.

Tests of a common positional context signal

Another important direction for future work will be to explore further the possibility that verbal and visuospatial serial memory utilize a common positional context signal. A provisional inroad to answering this question will be to establish whether or not spatial serial memory exhibits the kinds of effects that in verbal serial recall have been attributed to a context signal. These include the multifarious effects of grouping (including interposition errors) and positional protrusion errors[1]. This thesis has already shown that spatial serial memory exhibits effect of grouping on accuracy and recall latencies, but has failed to provide any evidence for interposition errors and did not examine protrusion errors at all. Concerning interposition errors, a number of suggestions were marshalled at the end of Chapter 7 for the failure to detect an increase in these errors for grouped relative to ungrouped spatial sequences. Future work will examine whether there are conditions under which these errors do materialize, in addition to looking at protrusion errors. If spatial serial memory does exhibit these error patterns this would certainly be consistent with the possibility that verbal and spatial serial memory draw on a common context signal. However, if spatial serial memory does not exhibit these error patterns then this would suggest that verbal and spatial serial

[1] Since visual serial memory is insensitive to grouping effects this avenue of research will focus solely on spatial serial memory. As noted above, the insensitivity of visual serial memory to grouping is likely attributable to the lack of an output mechanism for rehearsing visual, non-spatial information, as opposed to a fundamental difference in the representation of positional information in the visual domain relative to the verbal and spatial domains.

memory are unlikely to recruit a common context signal. Indeed, it may even question whether a context signal contributes to spatial serial memory at all.

Another approach will address this issue by presenting people with a verbal and a visuospatial sequence in parallel, and looking for correlations between correct and error responses on the two sequences. Participants will be presented with a verbal and a visual or spatial sequence in synchrony (the visual and spatial items being presented visually and the verbal items being presented auditorily). A participant's memory for the two sequences will then be examined using a list-probe recognition memory procedure (see Henson, Hartley, Burgess, Hitch, & Flude, 2003). This will involve re-presenting one of the sequences either in the same order (a positive probe) or with the positions of two of the items transposed (a negative probe) and requiring participants to decide whether the probe sequence is in the correct order, or not. Following this memory judgement, the second sequence will be re-presented either in the same order or with the positions of two of the items transposed, after which the same memory judgement will be required (note that sometimes the verbal sequence probe will be presented first and sometimes the nonverbal sequence probe will be presented first). Critically, the same probes will be used for the verbal and spatial sequences on each trial. Thus, if a positive probe is used for the verbal sequence a positive probe will also be used for the spatial sequence. Similarly, if a negative probe involving the transposition of the items at the first and second serial position is used for the nonverbal sequence, the same negative probe would be used for the verbal sequence. If the verbal and nonverbal items are associated with the same state of a common context signal then performance on the verbal and nonverbal sequence probes should be strongly correlated with one another.

A final approach will involve estimating the parameters of the putative positional context signal in the verbal and visuospatial domains using a positional model of serial recall. This will initially involve collecting new data directly comparing verbal, visual, and spatial serial reconstruction in a within-participants design when sequence length and other task characteristics, as well as overall level of recall accuracy are equated across the different materials. A positional model of serial recall will then be fit to the transposition matrices for the verbal, visual, and spatial data of individual participants. If verbal, visual, and spatial serial memory rely on a modality-independent positional context signal then the parameters of the positional model estimated for individual participants for the different types of material should be virtually identical. In contrast, heterogeneity in these parameter estimates would point to modality-dependent positional context signals.

Modelling

Incorporating better model selection methods

In this thesis, I have attempted to test the theoretical adequacy of models by going beyond standard goodness-of-fit methods of model evaluation. Specifically, where possible I have used a maximum likelihood, rather than a least-squares approach for model parameter estimation, and augmented this with generalizability measures, such as the AIC (Akaike, 1973) and the BIC (Schwartz, 1978), which trade-off the goodness-of-fit of the models with their parametric complexity. I have also examined the flexibility of the models by examining their predictions across a large portion of their parameter settings, using parameter space sensitivity analysis.

Notwithstanding the above attempts, the model evaluations reported here have not considered model complexity based upon functional form (how the parameters of a model combine in the model equation(s)), nor have they considered the flexibility of models as a function of their total parameter space. Future work will thus examine the descriptive accuracy of the models using model selection methods such as the Generalization Criterion Method (GCM; Busemeyer & Wang, 2000; Ahn, Busemeyer, Wagenmakers, & Stout, 2008). In the GCM, a data set to-be-fitted is divided into two sections, a calibration sample and a validation sample. Model parameters are first estimated for the calibration sample and these parameters are then frozen and used to generate predictions for the validation sample. The models are then compared with reference to their descriptive accuracy for the validation sample. The benefit of the GCM is that it takes into account model complexity based upon both the number of free model parameters (when combined with the AIC or BIC), as well as functional form. Future work will also examine the flexibility of the models using the parameter space partitioning algorithm (Pitt et al., 2006, 2008) described in

Chapter 2, which explores the entire parameter space of a model and maps different predefined qualitative data patterns onto different portions of that parameter space.

Modelling response time distributions

One limitation of the modelling of response latencies reported in previous chapters is that model evaluation was based solely upon visual comparisons of the model predictions with the observed data, rather than some objective measure of discrepancy. An important future goal will be to statistically fit the models of serial recall reported here to distributions of response times for both correct and error responses using techniques such as quantile maximum likelihood estimation (Heathcote, Brown, & Mewhort, 2002). Modelling latency distributions has been very productive in many areas of psychology, as latency distributions carry information that cannot be gleaned from mean or median latencies alone (see e.g., Heathcote, Popiel, & Mewhort, 1991). Indeed, the ability of a model to predict the shape of the response time distribution is considered to be an important test of the theoretical adequacy of that model (Luce, 1986). It is therefore hoped that applying the models to latency distributions will further help to adjudicate between them.

As part of this work it is desirable to incorporate methods that can simplify and facilitate the process of obtaining model latency predictions. One limitation of the lateral-inhibition network described in Chapter 3 is that the iterative dynamic response accumulation process is time consuming to simulate, which slows down the process of estimating model parameters from the data. Future work will therefore employ the Linear Ballistic Accumulator (LBA; Brown & Heathcote, 2008) model for the derivation of choice and response latency predictions. The LBA is a model of choice behaviour and response time in which each response competitor has its own response accumulator and these accumulators race towards a common decision threshold. Crucially, the accumulation process is linear and ballistic, which means that the trajectories of the accumulators can be determined on the basis of their initial starting positions and accumulation rates. This significantly facilitates the process of estimating model parameters.

Developing a process model of spatial serial memory

I believe that the most important objective for future modelling efforts is to develop a full-blown process model of spatial short-term order memory. Given the popularity of the Corsi-Blocks task and basic variants of this task in the experimental and neuropsychology literatures as an index of spatial sequential working memory capacities there has been a need for such a model for some time, but until recently the data was not available to constrain its development. Specifically, despite much research it was unclear how precisely serial order information is represented in the spatial domain, the nature of the reference frame used to represent spatial information, and the factors that modulate spatial serial recall accuracy. This thesis has identified a core set of mechanisms of serial order that appear to be implicated in spatial serial memory, whilst recent work by Avons (Avons, 2007; Avons & Oswald, 2009) has gone some way towards identifying the nature of the spatial reference frame underlying spatial serial recall. Specifically, Avons' work suggests that locations are represented in terms of an intrinsic template-centred frame of reference, as opposed to an egocentric, or allocentric frame of reference. Put more simply, what this means is that in a computerised version of a Corsi-type task, the locations are represented in terms of local spatial relationships with the boundaries of the computer display (the template) on which they are presented.

An important next step then, is the development of a computational model that provides a process implementation of the mechanisms of serial order identified in this thesis in conjunction with a representation of spatial position consistent with the template centred frame of reference identified by the work of Avons. Such a model needs to be able to accommodate the extant data on spatial serial memory reviewed in Chapter 1, in addition to the new data presented in this thesis. However, it must also accommodate a number of recently identified constraints that shed light on the factors that modulate spatial serial recall accuracy. These constraints include:

1. *The spatial clustering effect* (De Lillo, 2004; De Lillo & Lesk, 2010): Dividing a spatial sequence into clusters based upon spatial proximity enhances spatial serial recall accuracy.

2. *The path-length effect* (Parmentier et al., 2006): Sequences containing locations separated by long distances are recalled less effectively than sequences of locations separated by short distances.
3. *The path-crossing effect* (Parmentier & Andres, 2006; Parmentier et al., 2005): Crosses in the path formed by a sequence of locations retard memory for the order of those locations.

Taken together, these constraints represent a formidable modelling challenge. In particular the three constraints identified above will require further assumptions about the mechanisms, representations, and processes that contribute to the encoding and retrieval of spatial sequences. One of the goals of this modelling effort will be to elucidate what those assumptions are.

Chapter summary

The current thesis has identified four core explanatory principles that underlie the representation and generation of serial order in visuospatial short-term memory. These include: (1) a competitive queuing sequence planning and control mechanism, (2) a primacy gradient of activation, (3) positional marking, and (4) response suppression. Some tentative evidence was also obtained for the involvement of a fifth principle: namely a restricted end positional marker. These explanatory principles have also been identified through previous research, as well as results obtained from the current thesis, as contributors to the representation and generation of serial order in verbal short-term memory. The outcomes of the current thesis therefore support the view that verbal and visuospatial short-term memory rely upon at least some common principles for processing serial order information. However, they do not preclude the possibility that domain specific processes may also help to solve the problem of serial order in the two domains.

References

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

Agam, Y., Bullock, D., & Sekuler, R. (2005). Imitating unfamiliar sequences of connected linear motions. *Journal of Neurophysiology, 94,* 2832-2843.

Agam, Y., Galperin, H., Gold, B. J., Sekuler, R. (2007). Learning to imitate novel motion sequences. *Journal of Vision, 7,* 1-17.

Ahn, W. Y., Busemeyer, J. R., Wagenmakers, E. J., & Stout, J. C. (2008). Comparison of decision learning models using the generalization criterion method. *Cognitive Science, 32,* 1376-1402.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrox & F. Caski (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.

Anderson, J. A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.

Anderon, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38,* 341-380.

Anderson, J. R., & Matessa, M. P. (1997). A production system theory of serial memory. *Psychological Review,* 104, 728–748.

Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences, 99,* 13172–13177.

Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2003). Neural activity in prefrontal cortex during copying geometrical shapes I: Single cells encode shape, sequence, and metric parameters. *Experimental Brain Research, 150,* 127-141.

Averbeck, B. B., Crowe, D. A., Chafee, M. V., & Georgopoulos, A. P. (2003). Neural activity in prefrontal cortex during copying geometrical shapes II: Decoding shape segments from neural ensembles. *Experimental Brain Research, 150,* 142-153.

Avons, S. E. (1998). Serial report and item recognition of novel visual patterns. *British Journal of Psychology, 89,* 285-308.

Avons, S. E. (2007). Spatial span under translation: A study of reference frames. *Memory and Cognition, 35,* 402-417.

Avons, S. E. & Mason, A. (1999). Effects of visual similarity on serial report and item recognition. *Quarterly Journal of Experimental Psychology Section A,* 52, 217-240.

Avons, S. E., & Oswald, K. (2008). Using temporal order to identify spatial reference frames. *Perception & Psychophysics, 70,* 1068-1080.

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance,* 24, 780-790.

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science, 7,* 25-31.

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

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

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417-423.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition, 38,* 317-338.

Bohland, J. W., Bullock, D., & Guenther, F. H. (2009). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience, 7,* 1504-1529.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113,* 201-233.

Broadbent, D. E., & Broadbent, M. H. P. (1981). Recency effects in visual memory. *Quarterly Journal of Experimental Psychology, 33A,* 1-15.

Brown, J. W., Bullock, D., & Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks, 17,* 471-510.

Brown, G. D. A., Morin, C., & Lewandowsky, S. (2006). Evidence for time-based models of free recall. *Psychonomic Bulletin & Review, 13,* 717-723.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539-576.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127-181.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57,* 153-178.

Bullock, D. (2004). Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Sciences, 8,* 426-433.

Bullock, D., & Rhodes, B. (2003). Competitive queuing for serial planning and performance. In M. Arbib (Ed.) *Handbook of brain theory and neural networks*, second edition (pp. 241-244). Cambridge, MA: MIT Press.

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

Burgess, N., & Hitch, G. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106,* 551–581.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language* 55, 627-652.

Busemeyer, J. R., & Wang, Y. M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44,* 171-189.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55,* 75-84.

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

Cooper, R. P., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology, 17,* 297–338.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International,* 34(02), 891B (University microfilms No. AA105-77717).

Couture, M., & Tremblay, S. (2006). Exploring the characteristics of the visuospatial Hebb repetition effect. *Memory and Cognition, 34,* 1720-1729.

Cowan, N., Saults, S., Elliott, E. M., & Moreno, M. (2002). Deconfounding serial recall. *Journal of Memory and Language, 46,* 153-177.

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

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

Crowder, R. G. (1972). Visual and auditory memory. In J. Kavanagh & I. Mattingley (Eds.) *Language by Ear and by Eye* (pp. 251-275). Cambridge, MA: MIT Press.

Davelaar, E. J. (2007). Sequential retrieval and inhibition of parallel (re)activated representations: A neurocomputational comparison of competitive queuing and resampling models. *Adaptive Behavior, 15,* 51-71.

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review,* 112, 3-42.

Davelaar, E. J., Haarmann, H. J., Goshen-Gottstein, Y., & Usher, M. (2006). Semantic similarity dissociates short- from long-term recency: Testing a neurocomputational model of list memory. *Memory & Cognition, 34,* 323–334.

De Lillo, C. (2004). Imposing structure on a Corsi-type task: Evidence for hierarchical organization based on spatial proximity in serial-spatial memory. *Brain & Cognition, 55,* 415-426.

De Lillo, C., & Lesk, V. (2010). Spatial clustering and hierarchical coding in immediate serial recall. *European Journal of Cognitive Psychology, 22,* 216-246.

Dell, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review, 93,* 283-321.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104,* 123-147.

Della Sala, S., Gray, C., Baddeley, A. D., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwielding visuo-spatial memory. *Neuropsychologia, 37,* 1189-1199.

Depoorter, A., & Vandierendonck, A. (2009). Evidence for modality-independent order coding in working memory. *Quarterly Journal of Experimental Psychology, 62,* 531-549.

De Renzi, E., Faglioni, P., & Previdi, P. (1977). Spatial memory and hemispheric locus of lesion. *Cortex, 13,* 424-433.

De Renzi, E., & Nichelli, P. (1975). Verbal and non-verbal short memory impairment following hemispheric damage. *Cortex, 11,* 341-354.

Duncan, M., & Lewandowsky, S. (2005). The time course of response suppression: No evidence for a gradual release from inhibition. *Memory, 13,* 236-246.

Ebbinghaus, H. (1913/1964). *Memory: A contribution to experimental psychology.* New York: Dover.

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

Estes, W. K. (1975). Some targets for mathematical psychology. *Journal of Mathematical Psychology, 12,* 263–282.

Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review, 9,* 3-25.

Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review, 12,* 403-408.

Farmer, E. W., Berman, J. V. F., & Fletcher, Y. L. (1986). Evidence for a visuo-spatial scratch-pad in working memory. *Quarterly Journal of Experimental Psychology, 38A,* 675–688.

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

Farrand, P., Parmentier, F. B. R., & Jones, D. M. (2001). Temporal-spatial memory: Retrieval of spatial information does not reduce recency. *Acta Psychologica, 106,* 285–301.

Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language, 55,* 587-600.

Farrell, S. (2008). Multiple roles for time in short-term memory: Evidence from serial recall of order and timing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 128-145.

Farrell, S., & Lelievre, A. (2009). End anchoring in short-term order memory. *Journal of Memory and Language, 60,* 209-227.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9,* 59-79.

Farrell, S., & Lewandowsky, S. (2003). Dissimilar items benefit from phonological similarity in serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29,* 838-849.

Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language, 51,* 115-135.

Farrell, S., & Lewandowsky, S. (2007). *Response suppression contributes to recency in serial recall.* Unpublished manuscript, University of Western Australia.

Fischer, M. H. (2001). Probing spatial working memory with the Corsi blocks task. *Brain and Cognition, 45*, 143-154.

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

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

Glasspool, D. W. (2005). Modelling serial order in behaviour: Evidence from performance slips. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 241-270). Hove: Psychology Press.

Glasspool, D. W., & Houghton, G. (2005). Serial order and consonant-vowel structure in a graphemic output buffer model. *Brain and Language, 94,* 304-330.

Glasspool, D. W., Houghton, G., & Shallice, T. (1995). Interactions between knowledge sources in a dual-route connectionist model of spelling. In L. S. Smith and P. J. B. Hancock (Eds.), *Neural Computation and Psychology.* London: Springer-Verlag.

Glasspool, D. W., Shallice, T., & Cipolotti, L. (2006). Towards a unified process model for graphemic buffer disorder and deep dysgraphia. *Cognitive Neuropsychology*, 23, 479-512.

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.

Grossberg, S., & Pearson, L. R. (2008). Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Towards a unified theory of how the cerebral cortex works. *Psychological Review, 115,* 677-732.

Guerard, K., Neath, I., Surprenant, A. M., & Tremblay, S. (2010). Distinctiveness in serial memory for spatial information. *Memory & Cognition, 38,* 83-91.

Guerard, K., & Tremblay, S. (2008). Revisiting evidence for functional equivalence across verbal and spatial domains in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 34,* 556- 569.

Guerard, K., & Tremblay, S., & Saint-Aubin, J. (in press). The processing of spatial information in short-term memory: Insights from eye-tracking the path length effect. *Acta Psychologica.*

Hanley, J. R., Young, A.W., & Pearson, N. A. (1991). Impairment of the visuo-spatial scratch pad. *Quarterly Journal of Experimental Psychology, 43A*, 101-125.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35,* 1-31.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review, 9,* 394-401.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin,* 109, 340-347.

Henson, R. N. A. (1996). *Short-term memory for serial order.* Unpublished doctoral thesis. Cambridge University, Cambridge, U. K.

Henson R. N. A. (1998a). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36,* 73-137.

Henson, R. N. A. (1998b). Item repetition in short-term memory: Ranschburg repeated. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24,* 1162-1181.

Henson, R. N. A. (1999). Positional information in short-term memory: Relative or absolute? *Memory & Cognition, 27,* 915-927.

Henson, R. N. A., Hartley, T., Burgess, N., Hitch, G. J., & Flude, B. (2003). Selective interference with short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *Quarterly Journal of Experimental Psychology, 56A,* 1307-1334.

Henson, R. N., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology, 49A,* 80-115.

Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating Theory and Data: Essays in Honor of Bennet B. Murdock* (pp. 39-56). Hillside, NJ: Lawrence Erlbaum Associates.

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., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *Quarterly Journal of Experimental Psychology, 49,* 116-139.

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. (2005). Introduction to connectionist models in cognitive psychology: Basic structures, processes, and algorithms. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 241-270). Hove: Psychology Press.

Houghton, G., Glasspool, D. W., & Shallice, T. (1994). Spelling and serial recall: Insights from a competitive queuing model. In G. D. A. Brown and N. C. Ellis (Eds.), *Handbook of spelling: Theory, process and intervention.* Chichester: Wiley.

Houghton, G., & Hartley, T. (1995). Parallel models of serial behaviour: Lashley revisited. *Psyche,* 2(25). Retrieved from the World Wide Web: http://psyche.cs.monash.edu.au/v2/psyche-2-25-houghton.html

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299.

Hurlstone, M. J. (2006). *Top-down processing and temporal grouping in verbal short-term order memory.* Unpublished masters thesis. University of York, York, U. K.

Jahnke, J. C. (1969). The Ranschburg effect. *Psychological Review, 76,* 592-605.

Jalbert, A., Saint-Aubin, J., & Tremblay, S. (2008). Visual similarity in short-term recall for where and when. *Quarterly Journal of Experimental Psychology, 61,* 353-360.

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). London: Lawrence Erlbaum Associates.

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, & Cognition, 21,* 1008–1018.

Kemps, E. (1999). Effects of complexity on visuo-spatial working memory. *European Journal of Cognitive Psychology, 11,* 335-356.

Kieras, D. E., Meyer, D. E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In A. L. Miyake & P. Shaw (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 183–223). Cambridge, England:

Cambridge University Press.

Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General, 133,* 355-381.

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal on Optimization, 9,* 112-147.

Lamberts, K. (2005). Mathematical modelling of cognition. In K. Lamberts and R. Goldstone (Eds.), *The Handbook of Cognition*, (pp. 422-436). London: Sage Publication.

Lange, E. B. (2005). Disruption of attention by irrelevant stimuli in serial recall. *Journal of Memory and Language, 53,* 513-531.

Lashley, K. (1951). The problem of serial order in behavior. In Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp.112–136). New York: Wiley.

Lewandowsky, S. (1993). The rewards and hazards of computer simulations. *Psychological Science, 4,* 236-243.

Lewandowsky, S., & Brown, G. D. A. (2005). Serial recall and presentation schedule: A micro-analysis of local distinctiveness. *Memory, 13,* 283-292.

Lewandowsky, S., Brown, G. D. A., Wright, T., & Nimmo, L. M. (2006). Timeless memory: Evidence against temporal distinctiveness models of short-term memory for serial order. *Journal of Memory and Language, 54,* 20-38.

Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review, 11,* 771-790.

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. *The Psychology of Learning and Motivation, 49,* 1-48.

Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language, 59,* 200-222.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review, 96,* 25–57.

Lewandowsky, S., Nimmo, L. M., & Brown, G. D. A. (2008). When temporal isolation benefits memory for serial order. *Journal of Memory and Language, 58,* 415-428.

Lewandowsky, S., Wright, T., & Brown, G. D. A. (2007). The interpretation of temporal isolation effects. In N. Osaka, R. Logie, & M. D'Esposito (Eds.). *The cognitive neuroscience of working memory: Behavioural and neural correlates* (pp. 137-152). Oxford: Oxford University Press.

Logie, R. (1995). Visuo-spatial working memory. Hove: Lawrence Erlbaum Associates.

Logie, R. H., & Marchetti, C. (1991). Visuo-spatial working memory: Visual, spatial, or central executive? In R. H. Logie & M. Dennis (Eds.), *Mental images in human cognition* (pp. 105-115). Amsterdam: North-Holland.

Logie, R., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica, 75,* 55-74.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103-189). New York: Wiley.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization.* Oxford University Press: New York

Maybery, M., Parmentier, F. B. R., & Jones, D. M. (2002). Grouping of list items reflected in the timing of recall: Implications for models of serial verbal memory. *Journal of Memory and Language, 47,* 360-385.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective–likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105, 4,* 724–760.

McLeod, P., Plunkett, K., and Rolls, E. T. (1998). *Introduction to connectionist modelling of cognitive processes*. New York: Oxford University Press.

Meiser, T., & Klauer, K. C. (1999). Working memory and changing state hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25,* 1272-1299.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27,* 272-277.

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review, 100,* 183–203.

Murdock, B. B. (1995). Developing TODAM: Three models for serial order information. *Memory & Cognition, 23,* 631–645.

Murray, D. J., (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology, 21,* 263-276.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44,* 190-204.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47,* 90-100.

Myung, I. J., Balasubramanian, V., & Pitt, M. A. (2000). Counting probability distributions: Differential geometry and model selection. *Proceedings of the National Academy of Sciences USA, 97,* 11170-11175.

Myung, I. J., Pitt, M A., & Kim, W. (2005). Model evaluation, testing and selection. In K. Lamberts and R. Goldstone (Eds.), *The Handbook of Cognition*, (pp. 422-436). London: Sage Publication.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7,* 308-313.

Ng, L. H., & Maybery, M. T. (2002). Grouping in verbal short-term memory: Is position coded temporally? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55A,* 391–424.

Ng, L. H., & Maybery, M. T. (2005). Grouping in short-term memory: Do oscillators code the positions of items? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31,* 175-181.

Nimmo, L. M., & Lewandowsky, S. (2005). From brief gaps to very long pauses: Temporal isolation does not benefit serial recall. *Psychonomic Bulletin & Review, 12,* 999-1004.

Nimmo, L. M., & Lewandowsky, S. (2006). Distinctiveness revisited: Unpredictable temporal isolation does not benefit short-term serial recall of heard or seen events. *Memory & Cognition, 34,* 1368-1375.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary learning systems approach. *Psychological Review*, 110, 611-646.

Oberauer, K. (2008). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language, 49,* 469-483.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55,* 601–626.

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review, 115,* 544-576.

Page, M. P. A. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences, 23,* 443-467.

Page, M. P. A. (2005). Connectionist models of short-term memory for serial order. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 215-240). Hove: Psychology Press.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761-781.

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society, 364B,* 3737-3753.

Parmentier, F. B. R., Andres, P., Elford, G., & Jones, D. M. (2006). Organization of visuo-spatial serial memory: Interaction of temporal order with spatial and temporal grouping. *Psychological Research, 70,* 200-217.

Parmentier, F. B. R., Elford, G., & Maybery, M. M. (2005). Transitional information in spatial serial memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31,* 412-427.

Parmentier, F. B. R., & Jones, D. M. (2000). Functional characteristics of auditory temporal-spatial memory: Evidence from serial order errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 222-238.

Parmentier, F. B. R., King, S., & Dennis, I. (2006). Local temporal distinctiveness does not benefit auditory and spatial serial recall. *Psychonomic Bulletin & Review, 13,* 458-465.

Parmentier, F. B. R., & Maybery, M. T. (2008). Equivalent effects of grouping by time, voice and location on response timing in verbal serial memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 34,* 1349-1355.

Parmentier, F. B. R., Maybery, M., & Jones, D. M. (2004). Temporal grouping in auditory spatial serial memory. *Psychonomic Bulletin & Review, 11,* 501-507.

Pearson, D. G. (2001). Imagery and the visuospatial sketchpad. In J Andrade (Ed.), *Working memory in perspective* (pp. 33-59). Hove: Psychology Press.

Phillips, W. A. (1983). Short-term visual memory. *Philosophical Transactions of the Royal Society of London, B302*, 295-309.

Phillips, W.A., & Christie, D. F. M., (1977a). Components of visual memory. *Quarterly Journal of Experimental Psychology*, *29*, 117-133.

Phillips, W.A., & Christie, D. F. M., (1977b). Interference with visualization. *Quarterly Journal of Experimental Psychology*, *29*, 637-650.

Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review, 113,* 57-83.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6,* 421-425.

Pitt, M. A., Myung, J. I., Montenegro, M., & Pooley, J. (2008). Measuring model flexibility with parameter space partitioning: An introduction and application example. *Cognitive Science, 32,* 1285-1303.

Polyn, S. M., Norman, K. A, & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116,* 129-156.

Postle, B. R., Idzikowski, C., Della Sala, S., Logie, R. H., & Baddeley, A. D. (2006). The selective disruption of spatial working memory by eye movements. *Quarterly Journal of Experimental Psychology, 59,* 100-120.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological*

Review, 88, 93–134.

Ratcliff, R. (1998). The role of mathematical psychology in experimental psychology. *Australian Journal of Psychology, 50,* 129-130.

Rhodes, B., and Bullock, D. (2002). *Neural dynamics of learning and performance of fixed sequences: Latency pattern reorganizations and the N-STREAMS model.* Boston University Technical Report CAS/CNS-02-007.

Rhodes, B. J., Bullock, D., Verwey, W. B., Averbeck, B. B., & Page, M. P. A. (2004). Learning and production of movement sequences: Behavioural, neurophysiological, and modelling perspectives. *Human Movement Science, 23,* 699–746.

Rumelhart, D, E., & McClelland, J. L. (1986). *Parallel distributed processing – volume 1: Foundations.* Cambridge, MA: MIT Press.

Rumelhart, D. E., & Norman, D. A. (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. *Quarterly Journal of Experimental Psychology, 21,* 137-147.

Ryan, J. (1969b). Temporal grouping, rehearsal and short-term memory. *Quarterly Journal of Experimental Psychology, 21,* 148-155.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464.

Sekuler, R., Siddiqui, A., Goyal, N., & Rajan, R. (2003). Reproduction of seen actions: Stimulus-selective learning. *Perception, 32,* 839-854.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145-166.

Smirni, P., Villardita, C., & Zappala, G. (1983). Influence of different paths on spatial memory performance in the block-tapping test. *Journal of Clinical Neuropsychology, 5,* 355-359.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33,* 5-42.

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6,* 11-20.

Smyth, M. M. (1996). Interference with rehearsal in spatial working memory in the absence of eye movements. *Quarterly Journal of Experimental Psychology, 40A*, 940-949.

Smyth, M. M., & Scholey, K. A. (1994). Interference in spatial immediate memory. *Memory and Cognition, 22*, 1-13.

Smyth, M. M., & Scholey, K. A. (1996). Serial order in spatial immediate memory. *Quarterly Journal of Experimental Psychology, 49,* 159-177.

Smyth, M. M., Hay, D. C., Hitch, G. J., & Horton, N. J. (2005). Serial position memory in the visual-spatial domain: Reconstructing sequences of unfamiliar faces. *Quarterly Journal of Experimental Psychology, 58,* 909-930.

Smyth, M. M., Pearson N. A., & Pendleton, L. R. (1988). Movement and working memory: Patterns and positions in space. *Quarterly Journal of Experimental Psychology, 40A*, 497-514.

Solway, A., Kahana, M. J., Addis, K. M., & Murdock, B. B. A strength-based associative chaining model of serial learning. *Submitted.*

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117-152). New York: Academic Press.

Surprenant, A. M., Kelley, M. R., Farley, L. A., & Neath, I. (2005). Fill-in and infill errors in order memory. *Memory, 13,* 267–273.

Thomas, J. G., Milner, H. R., & Haberlandt, K. F. (2003). Forward and backward recall: Different response time patterns, same retrieval order. *Psychological Science, 14,* 169–174.

Tremblay, S., Guerard, K., Parmentier, F. B. R., Nicholls, A. P., & Jones, D. M. (2006). A spatial modality effect in serial memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 32,* 1208-1215.

Tremblay, S., Macken, B., & Jones, D. M. (2001). The impact of broadband noise on serial memory: Changes in band-pass frequency increases disruption. *Memory, 9,* 323-331.

Tremblay, S., Nicholls, A., Parmentier, F. B. R., & Jones, D. M. (2005). Visual distraction and visuo-spatial memory: A sandwich effect. *Memory, 13,* 357-363.

Tresch, M. C., Sinnamon, H. M., & Seamon, J. G. (1993). Double dissociation of spatial and object visual memory: Evidence from selective interference in intact human subjects. *Neuropsychologia, 31,* 211-219.

Vallar, G., & Baddeley, A. D. (1984). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior, 23,* 151-161.

Vousden, J. I., & Brown, G. D. A. (1998). To repeat or not to repeat: The time course of response suppression in sequential behaviour. In J. A. Bullinaria, D. W. Glasspool, & G. Houghton (Eds.), *Proceedings of the fourth neural computation and psychology workshop: Connectionist representations* (pp. 301-315). London: Springer Verlag.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11,* 192-196.

Walker, P., Hitch, G. J., & Duroe, S. (1993). The effect of visual similarity on short-term memory for spatial location: Implications for the capacity of visual short-term memory. *Acta Psychologica, 83,* 203-224.

Wang, P. P., & Bellugi, U. (1994). Evidence from two genetic syndromes for a dissociation between verbal and visual-spatial short-term memory. *Journal of Clinical and Experimental Neuropsychology, 16,* 317-322.

Ward, G., Avons, S. E., & Mason, L. (2005). Serial position curves in short-term memory: Functional equivalence across modalities. *Memory, 13,* 308-317.

Wickelgren, W. A. (1965). Short-term memory for phonemically similar lists. *American Journal of Psychology, 78,* 567-574.

Appendix 1

Supplemental goodness-of-fit values

This appendix provides tables of supplemental goodness-of-fit values (e.g., values for individual participants) for the quantitative model fitting exercises of Chapters 3, 4, 5, and 8.

Chapter 3: Fits to Farrell and Lewandowsky (2004)

The minimised root mean square deviations (RMSDs) for the PG+RS and PG+PM+RS models for each individual participant after fitting to the ungrouped sequence condition of Experiment 3 of Farrell and Lewandowsky (2004), are given in Table A1-1.

Participant	PG+RS	PG+PM+RS
1	0.11	0.09
2	0.07	0.06
3	0.10	0.10
4	0.07	0.07
5	0.09	0.09
6	0.05	0.03
7	0.04	0.05
8	0.08	0.05
9	0.10	0.06
10	0.08	0.08
11	0.16	0.13
12	0.09	0.11
13	0.09	0.10
14	0.08	0.08
15	0.11	0.07
16	0.09	0.11
17	0.09	0.05
18	0.11	0.09

Table continued on next page

19	0.10	0.12
20	0.07	0.09
21	0.06	0.08
22	0.09	0.07
23	0.08	0.09
24	0.11	0.10
25	0.09	0.12
26	0.10	0.11
Mean RMSD	0.09	0.08

Table A1-1 *Minimised RMSDs for the fits of two models of serial order to the ungrouped condition of Experiment 3 of Farrell and Lewandowsky (2004).*

Chapter 4: Fits to Experiment 2

The minimised RMSDs for the PM+RS, PG+RS, and PG+PM+RS models for each individual participant after fitting to the data of Experiment 2 are shown in Table A1-2 for four-item sequences, Table A1-3 for five-item sequences, and Table A1-4 for six-item sequences.

Participant	PM+RS	PG+RS	PG+PM+RS
1	0.07	0.04	0.05
2	0.07	0.06	0.03
3	0.06	0.06	0.02
4	0.14	0.06	0.06
5	0.08	0.07	0.06
6	0.11	0.08	0.08
7	0.07	0.05	0.04
8	0.01	0.04	0.01
9	0.06	0.07	0.03
10	0.07	0.08	0.04
11	0.08	0.06	0.05
12	0.03	0.06	0.02
13	0.04	0.04	0.03
14	0.10	0.09	0.08
15	0.07	0.06	0.03
16	0.06	0.05	0.03
17	0.04	0.08	0.01
18	0.11	0.06	0.05
Mean RMSD	0.07	0.06	0.04

Table A1-2 *Minimised RMSDs for the fits of three models of serial order to the four-item sequence condition of Experiment 2.*

Participant	PM+RS	PG+RS	PG+PM+RS
1	0.08	0.05	0.05
2	0.13	0.11	0.09
3	0.13	0.09	0.07
4	0.07	0.09	0.05
5	0.18	0.09	0.10
6	0.17	0.08	0.10
7	0.16	0.07	0.09
8	0.06	0.05	0.05
9	0.12	0.08	0.07
10	0.11	0.05	0.04
11	0.13	0.04	0.04
12	0.11	0.06	0.05
13	0.08	0.04	0.04
14	0.08	0.07	0.03
15	0.19	0.09	0.10
16	0.16	0.08	0.09
17	0.11	0.09	0.07
18	0.18	0.10	0.09
Mean RMSD	0.13	0.07	0.07

Table A1-3 *Minimised RMSDs for the fits of three models of serial order to the five-item sequence condition of Experiment 2.*

Participant	PM+RS	PG+RS	PG+PM+RS
1	0.15	0.11	0.09
2	0.12	0.08	0.07
3	0.14	0.08	0.07
4	0.12	0.07	0.07
5	0.17	0.07	0.07
6	0.19	0.09	0.09
7	0.25	0.11	0.12
8	0.13	0.11	0.10
9	0.11	0.06	0.05
10	0.16	0.07	0.07
11	0.14	0.06	0.05
12	0.15	0.09	0.07
13	0.18	0.11	0.12
14	0.11	0.06	0.04
15	0.17	0.07	0.07
16	0.15	0.05	0.05
17	0.08	0.12	0.05
18	0.17	0.06	0.06
Mean RMSD	0.15	0.08	0.07

Table A1-4 *Minimised RMSDs for the fits of three models of serial order to the six-item sequence condition of Experiment 2.*

Chapter 7: Fits to Experiment 10

The maximum log-likelihood parameter estimates and associated goodness-of-fit indices for the SP, GP-SP, and GP-WP models for each individual participant after fitting to the grouped spatial sequence condition of Experiment 10 are given in Table A1-5, A1-6, and A1-7, respectively. The goodness-of-fits of the same models after fitting to the grouped verbal sequence condition of Experiment 10 are given in Table A1-8, A1-9, and A1-10, respectively.

Participant	ln$L$	AIC	ω AIC	BIC	ω BIC
1	-332	669	0.00	673	0.00
2	-292	588	0.00	593	0.01
3	-154	313	0.00	318	0.00
4	-307	618	0.00	623	0.00
5	-306	615	0.00	620	0.00
6	-237	477	0.00	482	0.00
7	-244	493	0.00	498	0.00
8	-299	602	0.00	607	0.00
9	-202	408	0.00	412	0.00
10	-253	510	0.00	515	0.00
11	-269	541	0.00	546	0.00
12	-272	548	0.00	553	0.00
13	-328	659	0.00	664	0.01
14	-252	507	0.00	512	0.00
15	-141	286	0.00	291	0.00
16	-270	544	0.00	549	0.00
17	-159	321	0.00	326	0.00
18	-295	593	0.00	598	0.00
Mean	-256	516	0.00	521	0.00

Table A1-5 *Fits of the SP model to the grouped spatial sequence condition of Experiment 10. Note—lnL = maximum log-likelihood estimate; AIC = Akaike Information Criterion; ω AIC = AIC weight; BIC = Bayesian Information Criterion; ω BIC = BIC weight.*

Participant	ln*L*	AIC	ω AIC	BIC	ω BIC
1	-308	622	0.32	630	0.32
2	-286	579	0.49	586	0.49
3	-154	313	0.00	320	0.00
4	-296	598	0.00	606	0.00
5	-301	607	0.03	615	0.03
6	-235	475	0.00	482	0.00
7	-241	489	0.00	496	0.00
8	-294	594	0.04	601	0.04
9	-196	398	0.00	406	0.00
10	-249	504	0.00	511	0.00
11	-268	542	0.00	549	0.00
12	-266	537	0.00	544	0.00
13	-325	656	0.01	664	0.01
14	-241	489	0.00	496	0.00
15	-137	281	0.00	288	0.00
16	-260	526	0.00	533	0.00
17	-157	319	0.00	326	0.00
18	-280	566	0.16	574	0.16
Mean	-250	505	0.06	513	0.06

Table A1-6 *Fits of the GP-SP model to the grouped spatial sequence condition of Experiment 10. Note—lnL = maximum log-likelihood estimate; AIC = Akaike Information Criterion; ω AIC = AIC weight; BIC = Bayesian Information Criterion; ω BIC = BIC weight.*

Participant	ln$L$	AIC	ω AIC	BIC	ω BIC
1	-307	621	0.68	628	0.68
2	-286	579	0.50	586	0.50
3	-138	282	1.00	289	1.00
4	-287	581	1.00	588	1.00
5	-297	600	0.97	608	0.97
6	-215	437	1.00	444	1.00
7	-234	475	1.00	482	1.00
8	-291	587	0.96	594	0.96
9	-190	385	1.00	392	1.00
10	-240	485	1.00	492	1.00
11	-240	487	1.00	494	1.00
12	-256	518	1.00	525	1.00
13	-321	648	0.98	655	0.97
14	-225	457	1.00	464	1.00
15	-132	270	1.00	277	0.99
16	-245	497	1.00	504	1.00
17	-120	247	1.00	254	1.00
18	-279	563	0.84	570	0.84
Mean	-239	484	0.94	491	0.94

Table A1-7 *Fits of the GP-WP model to the grouped spatial sequence condition of Experiment 10. Note—lnL = maximum log-likelihood estimate; AIC = Akaike Information Criterion; ω AIC = AIC weight; BIC = Bayesian Information Criterion; ω BIC = BIC weight.*

Participant	ln$L$	AIC	ω AIC	BIC	ω BIC
1	-210	424	0.00	429	0.00
2	-234	473	0.00	478	0.00
3	-314	632	0.00	637	0.00
4	-337	679	0.00	684	0.00
5	-307	619	0.00	623	0.00
6	-260	524	0.00	529	0.00
7	-331	665	0.00	670	0.00
8	-229	462	0.00	467	0.00
9	-289	581	0.00	586	0.00
10	-247	497	0.00	502	0.00
11	-241	485	0.00	490	0.00
12	-227	458	0.00	463	0.00
13	-245	493	0.00	498	0.00
14	-296	597	0.00	601	0.00
15	-229	461	0.00	466	0.00
16	-174	353	0.00	357	0.00
17	-328	660	0.00	665	0.00
18	-367	738	0.00	743	0.00
Mean	-270	544	0.00	549	0.00

Table A1-8 *Fits of the SP model to the grouped verbal sequence condition of Experiment 10. Note—lnL = maximum log-likelihood estimate; AIC = Akaike Information Criterion; ω AIC = AIC weight; BIC = Bayesian Information Criterion; ω BIC = BIC weight.*

Participant	ln$L$	AIC	ω AIC	BIC	ω BIC
1	-188	383	0.00	390	0.00
2	-205	416	0.00	423	0.00
3	-314	634	0.00	641	0.00
4	-328	662	0.00	669	0.00
5	-303	612	0.00	619	0.00
6	-235	476	0.00	483	0.00
7	-330	665	0.00	672	0.00
8	-224	455	0.00	462	0.00
9	-287	580	0.00	588	0.00
10	-246	499	0.00	506	0.00
11	-229	463	0.00	470	0.00
12	-212	430	0.00	437	0.00
13	-244	493	0.00	501	0.00
14	-273	553	0.00	560	0.00
15	-194	394	0.00	401	0.00
16	-165	336	0.00	344	0.00
17	-327	659	0.00	667	0.00
18	-366	738	0.00	745	0.00
Mean	-259	525	0.00	532	0.00

Table A1-9 *Fits of the GP-SP model to the grouped verbal sequence condition of Experiment 10. Note—lnL = maximum log-likelihood estimate; AIC = Akaike Information Criterion; ω AIC = AIC weight; BIC = Bayesian Information Criterion; ω BIC = BIC weight.*

Participant	ln$L$	AIC	ω AIC	BIC	ω BIC
1	-141	289	1.00	296	1.00
2	-194	395	1.00	402	1.00
3	-271	547	1.00	555	1.00
4	-303	613	1.00	620	1.00
5	-265	536	1.00	544	1.00
6	-197	399	1.00	406	1.00
7	-315	635	1.00	642	1.00
8	-214	435	1.00	442	1.00
9	-270	546	1.00	554	1.00
10	-209	423	1.00	430	1.00
11	-178	361	1.00	368	1.00
12	-185	376	1.00	383	1.00
13	-214	434	1.00	441	1.00
14	-259	525	1.00	532	1.00
15	-177	360	1.00	367	1.00
16	-100	205	1.00	212	1.00
17	-303	613	1.00	620	1.00
18	-355	717	1.00	724	1.00
Mean	-231	467	1.00	474	1.00

Table A1-10 *Fits of the GP-WP model to the grouped verbal sequence condition of Experiment 10. Note—lnL = maximum log-likelihood estimate; AIC = Akaike Information Criterion; ω AIC = AIC weight; BIC = Bayesian Information Criterion; ω BIC = BIC weight.*

Chapter 8: Fits to Experiment 11

The maximum log-likelihood parameter estimates and associated goodness-of-fit indices for the fits of the PM+RS, PG+RS, PG+PM+RS, and PG+PM+RE+RS models to the aggregate data for the five-item, six-item, and seven-item sequence length conditions of Experiment 11 are shown in Table A1-11.

		Transposition Matrix			Fill-in Ratios
Model	k	ln$L$	BIC	ωBIC	RMSD
Five-Items					
PM+RS	2	-6696	13398	0.00	3.06
PG+RS	2	-7331	14668	0.00	46.72
PG+PM+RS	3	-6618	13246	0.20	2.20
PG+PM+RE+RS	4	-6615	13243	0.80	2.08
Six-Items					
PM+RS	2	-12134	24275	0.00	1.24
PG+RS	2	-13102	26211	0.00	10.51
PG+PM+RS	3	-12031	24073	0.04	0.47
PG+PM+RE+RS	4	-12026	24066	0.96	0.37
Seven-Items					
PM+RS	2	-17964	35936	0.00	1.22
PG+RS	2	-19287	38582	0.00	2.82
PG+PM+RS	3	-17885	35782	0.00	0.79
PG+PM+RE+RS	4	-17870	35756	1.00	0.71

Table A1-11 *Goodness-of-fits of four models of serial order to the data of Experiment 11. Note— lnL = maximum log-likelihood estimate; BIC = Bayesian Information Criterion; ω BIC = BIC weight; RMSD = Root Mean Square Deviation.*

Chapter 8: Fits to Experiment 1

The maximum log-likelihood parameter estimates and associated goodness-of-fit indices for the fits of the PM+RS, PG+RS, PG+PM+RS, and PG+PM+RE+RS models to the aggregate data for the different sequence length conditions of Experiment 1 are shown in Table A1-12.

		Transposition Matrix			Fill-in Ratios
Model	k	ln$L$	BIC	ωBIC	RMSD
Four-Items					
PM+RS	2	-6364	12736	0.00	2.14
PG+RS	2	-6731	13468	0.00	11.36
PG+PM+RS	3	-6378	12762	0.00	1.10
PG+PM+RE+RS	4	-6341	12693	1.00	0.94
Five-Items					
PM+RS	2	-10575	21160	0.00	1.53
PG+RS	2	-11354	22714	0.00	3.36
PG+PM+RS	3	-10561	21128	0.00	1.10
PG+PM+RE+RS	4	-10527	21067	1.00	0.86
Six-Items					
PM+RS	2	-14987	29985	0.00	0.97
PG+RS	2	-15931	31869	0.00	1.88
PG+PM+RS	3	-14956	29919	0.00	0.73
PG+PM+RE+RS	4	-14943	29900	1.00	0.47

Table A1-12 *Goodness-of-fits of four models of serial order to the data of Experiment 1. Note— lnL = maximum log-likelihood estimate; BIC = Bayesian Information Criterion; ωBIC = BIC weight; RMSD = Root Mean Square Deviation.*

Chapter 8: Fits to Experiment 12

The maximum log-likelihood parameter estimates and associated goodness-of-fit indices for the fits of the PM+RS, PG+RS, PG+PM+RS, and PG+PM+RE+RS models to the aggregate data for the different sequence length conditions of Experiment 12 are shown in Table A1-13.

		Transposition Matrix			Fill-in Ratios
Model	k	ln$L$	BIC	ωBIC	RMSD
Seven-Items					
PM+RS	2	-15044	30096	0.00	1.77
PG+RS	2	-15923	31855	0.00	6.70
PG+PM+RS	3	-14793	29599	0.61	1.03
PG+PM+RE+RS	4	-14792	29600	0.39	1.01
Nine-Items					
PM+RS	2	-19980	39969	0.00	1.50
PG+RS	2	-20917	41843	0.00	2.60
PG+PM+RS	3	-19674	39361	0.06	0.95
PG+PM+RE+RS	4	-19669	39356	0.94	0.80

Table A1-13 *Goodness-of-fits of four models of serial order to the data of Experiment 12. Note— lnL = maximum log-likelihood estimate; BIC = Bayesian Information Criterion; ωBIC = BIC weight; RMSD = Root Mean Square Deviation.*

Appendix 2

Supplemental data

This appendix presents the latency-displacements functions (LDFs) for Experiments 11 and 12, which were not shown in Chapter 8. The LDFs for these experiments were analysed using the same protocol as described in Chapters 4 and 5.

Latency-displacement functions for Experiment 11

The LDFs for the different sequence length conditions of Experiment 11 are shown in Figure A2-1. It is apparent from inspection of this figure that the slopes of the functions for anticipations are steeply negative, whereas the slopes of the functions for postponements appear flat.

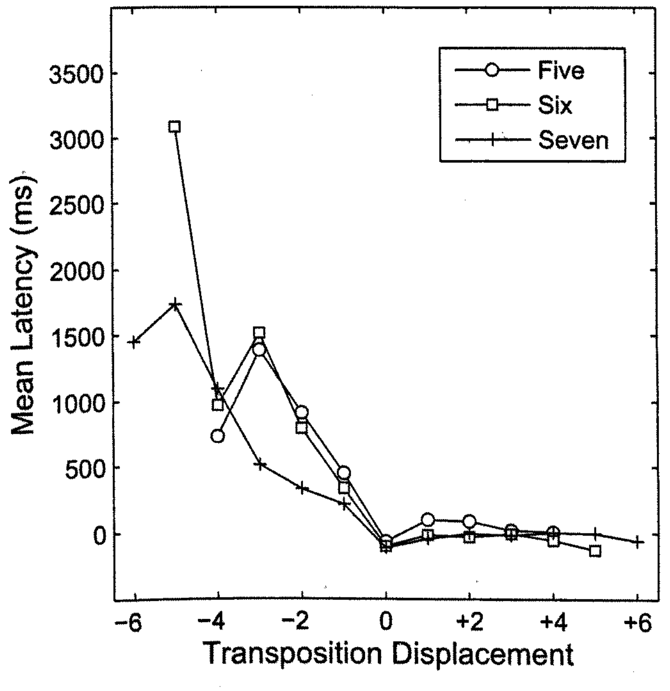

Figure A2-1 *Latency-displacement functions for Experiment 11.*

The mean regression parameter estimates for the slopes of the functions for anticipations and postponements as a function of sequence length are given in Table A2-1. From inspection of this table it can be seen that the mean slope estimates for anticipations deviated significantly from zero

Parameter	Estimate	*SE*	*t*	*p*
Five				
Anticipation	-406.41	119.68	-3.3957	.00
Postponement	13.80	14.12	0.9775	.34
Six				
Anticipation	-605.96	206.26	-2.9379	.00
Postponement	-8.71	14.96	-0.5819	.57
Seven				
Anticipation	-278.00	50.53	-5.5008	.00
Postponement	4.85	13.14	0.3692	.72

Table A2-1 *Mean regression slope parameter estimates for the latency-displacement functions of Experiment 11.*

for all sequence length conditions: $t(17) = -3.3957$, $p < .001$, for five-item sequences, $t(17) = -2.9379$, $p < .001$, for six-item sequences, and, $t(17) = -5.5008$, $p < .001$, for seven-item sequences. For postponements, the mean slope estimate deviated significantly from zero for five-item sequences, $t(17) = 0.9775$, $p < .05$, but not for six-item sequences, $t(17) = -0.5819$, $p = 0.6$, nor seven-item sequences, $t(17) = 0.3692$, $p = 0.72$.

Latency-displacement functions for Experiment 12

The LDFs for the different sequence length conditions of Experiment 12 are shown graphically in Figure A2-2. It can be seen from inspection of this figure that the slopes of the functions for anticipations are steeply negative (more so for the seven-item than the nine-item sequence condition), whereas the slopes of the functions for postponements appear relatively flat, but with a positive trend line.

The mean regression parameter estimates for the slopes of the functions for anticipations and postponements as a function of sequence length are shown in Table A2-2. As can be seen from inspection of this table, the mean slope estimates for anticipations deviated significantly from zero: $t(23) = -2.0534$, $p = .05$, for seven-item sequences, and, $t(23) = -48.0558$, $p = .01$, for nine-item sequences. For postponements, the mean slope estimate for nine-item sequences deviated

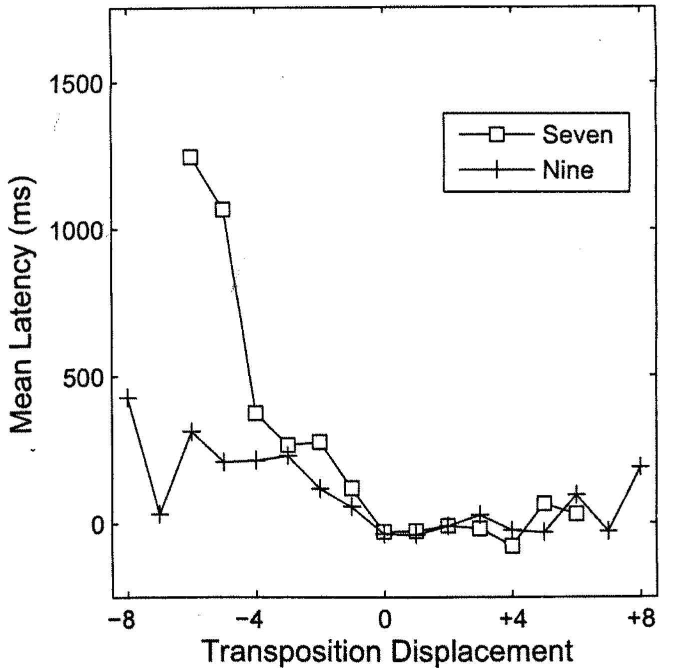

Figure A2-2 *Latency-displacement functions for Experiment 12.*

Parameter	Estimate	*SE*	*t*	*p*
Seven				
Anticipation	-194.07	94.51	-2.0534	.05
Postponement	1.88	6.60	0.2847	.08
Nine				
Anticipation	-48.06	18.66	-2.5751	.02
Postponement	11.70	7.12	1.6438	.01

Table A2-2 *Mean regression slope parameter estimates for the latency-displacement functions of Experiment 12.*

significantly from zero, $t(23) = 1.6438$, $p = .01$, but the mean slope estimate for seven-item sequences did not deviate significantly from zero, $t(23) = 0.2847$, $p = .78$.

Appendix 3

Supplemental simulations

This appendix presents supplemental simulations in the form of fits of a version of the primacy gradient and response suppression model employing a linear primacy gradient to the data of Experiments 11, 1, and 12. The purpose of these simulations is to show that like the combination of an exponential primacy gradient with response suppression, a linear primacy gradient in conjunction with response suppression generally predicts larger ratios of fill-in to infill errors than observed empirically, in addition to less realistic serial position curves and transposition gradients than the alternative models of serial order considered in Chapter 8.

The linear primacy gradient and response suppression model, defined by equations 2-5 and 2-7 of Chapter 2, was fit to the transposition matrix of response frequencies, for each sequence length condition, for each of the above experiments, using the same maximum likelihood parameter estimation procedure described to fit the models of Chapter 8 to the very same data.

Fits to Experiment 11

The maximum log-likelihood parameter estimates for the fits of the linear primacy gradient model to the transposition matrices for the five-item, six-item, and seven-item sequence conditions of Experiment 11 employing verbal stimuli were: -7407, -13218, and -19377, respectively. The root mean square deviations (RMSDs) between the predicted and observed ratios of fill-in to infill errors after fitting the model to the transposition matrices were: 23.11, 5.27, and 1.83, respectively.

The accuracy serial position curves, transposition gradients, and ratios of fill-in to infill errors predicted by the linear primacy gradient model are shown in Figure A3-1 and can be contrasted with the data and fits of the four alternative models of serial order, which are shown in Figures 8-1, 8-2, and 8-3. It can be seen from inspection of these figures that like the exponential primacy gradient model, the linear primacy gradient model predicts markedly less realistic serial position

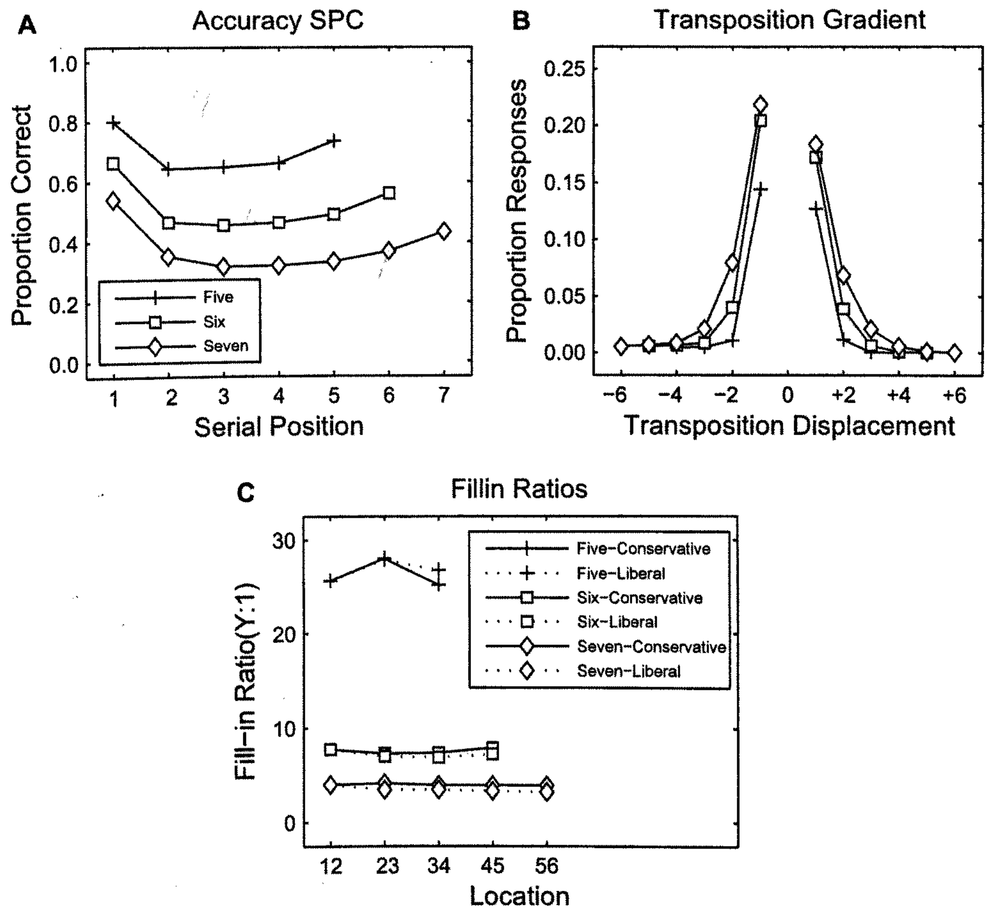

Figure A3-1 *Fits of the linear primacy gradient and response suppression model to the data of Experiment 11. Panels show accuracy serial position curves (A), transposition gradients (B), and ratios of fill-in to infill errors (C).*

curves (Figure A3-1A, versus Figure 8-1) and transposition gradients (Figure A3-1B, versus Figure 8-2), than the PM+RS, PG+PM+RS, and PG+PM+RE+RS models. It can further be seen that the linear primacy gradient model generally predicts larger ratios of fill-in to infill errors than observed empirically (Figure A3-1C, versus Figure 8-3). The mean ratios of fill-in to infill errors predicted by this model, averaged across error locations and scoring procedure, for five-item, six-item, and seven-item sequences, were: 27:1, 7:1, and 4:1, respectively, compared to empirically observed mean ratios of 3.5:1, 2:1, and 2:1. These predicted mean ratios are smaller than those predicted by the exponential primacy gradient model. Nevertheless, the predicted mean ratios for five-item and six-item sequences are still distinctly larger than the empirically observed mean ratios.

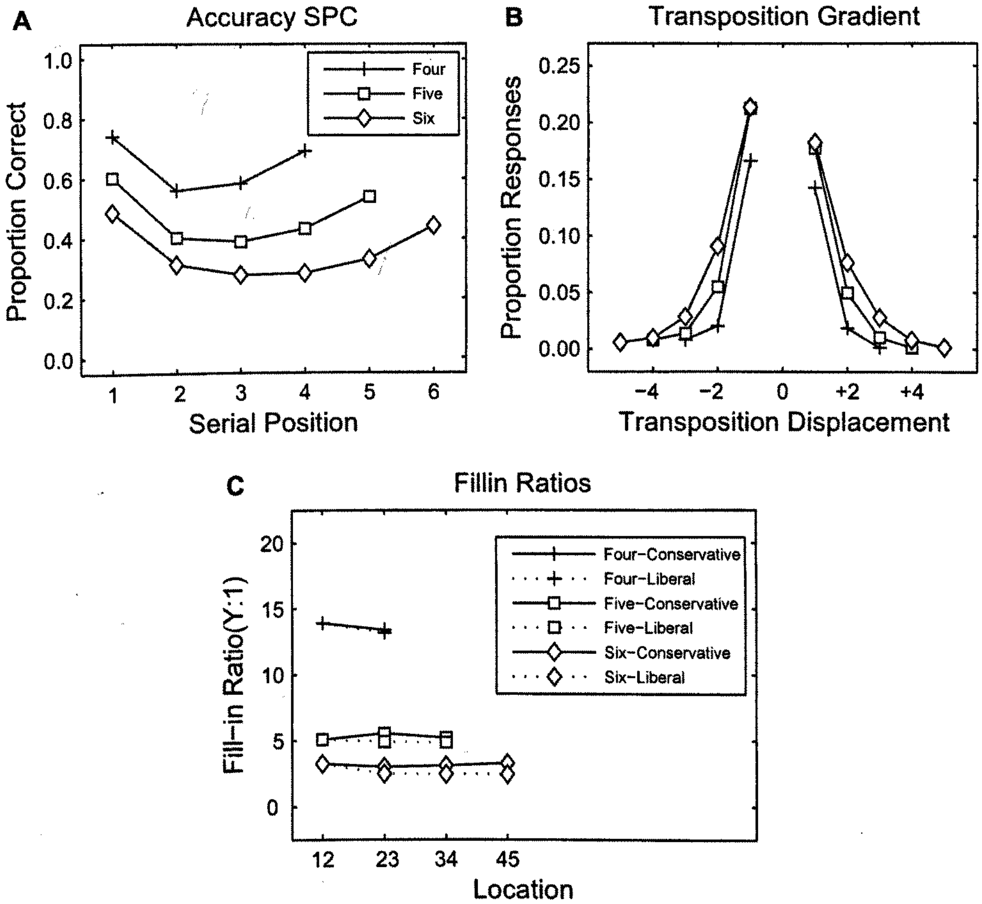

Figure A3-2 *Fits of the linear primacy gradient and response suppression model to the data of Experiment 1. Panels show accuracy serial position curves (A), transposition gradients (B), and ratios of fill-in to infill errors (C).*

Fits to Experiment 1

The maximum log-likelihood parameter estimates for the fits of the linear primacy gradient and response suppression model to the transposition matrices for the five-item, six-item, and seven-item sequence conditions of Experiment 1 employing visual stimuli were: -6683, -11297, and, -15921, respectively. The RMSDs between the predicted and observed ratios of fill-in to infill errors after fitting the model to the transposition matrices were: 10.52, 2.77, and 1.28, respectively.

The accuracy serial position curves, transposition gradients, and ratios of fill-in to infill errors predicted by the linear primacy gradient model are illustrated in Figure A3-2, and can be compared with the data and fits of the four alternative models of serial order shown in Figures 8-4, 8-5, and 8-6. As for the fits to Experiment 11, it can be seen that the linear primacy gradient model, like the exponential primacy gradient model, predicts markedly less realistic serial position curves (Figure A3-2A, versus Figure 8-4) and transposition gradients (Figure A3-2B, versus Figure 8-5) than the PM+RS, PG+PM+RS, and PG+PM+RE+RS models. It can also be seen that the linear primacy gradient model generally predicts larger ratios of fill-in to infill errors than observed empirically (Figure A3-2C, versus Figure 8-6). The mean ratios of fill-in to infill errors predicted by the linear primacy gradient model, averaged across error locations and scoring procedure, for four-item, five-item, and six-item sequences, were: 14:1, 5:1, and 3:1, compared to empirical values of 3:1, 2:1, and 1.8:1. These mean ratios are smaller than those predicted by the exponential primacy gradient model. Moreover, the predicted mean ratios for five-item and six-item sequences are only slightly larger than the empirically observed mean ratios. Nevertheless, the predicted mean ratio for four-item sequences is distinctly larger than the empirically observed mean ratio.

Fits to Experiment 12

The maximum log-likelihood parameter estimates for the fits of the linear primacy gradient and response suppression model to the transposition matrices for the seven-item and nine-item sequence conditions of Experiment 12 employing spatial stimuli were: -11136 and -21318, respectively. The RMSDs between the predicted and observed ratios of fill-in to infill errors after fitting the model to the transposition matrices were: 4.93 and 1.33, respectively.

Figure A3-3 shows the accuracy serial position curves, transposition gradients, and ratios of fill-in to infill errors predicted by the linear primacy gradient model. These predictions can be contrasted with the data and fits of the four alternative models of serial order, which are shown in Figures 8-7, 8-8, and 8-9. As per the fits to Experiments 11 and 1, it can be seen that the linear primacy gradient model, like the exponential primacy gradient model, predicts distinctly less realistic serial position curves (Figure A3-3A, versus Figure 8-7) and transposition gradients (Figure A3-3B, versus Figure 8-8) than the PM+RS, PG+PM+RS, and PG+PM+RE+RS models. Concerning the predictions for fill-in errors (Figure A3-3C, versus Figure 8-9), although the linear primacy gradient model predicted larger ratios of fill-in to infill errors for seven-item sequences than observed empirically, the predicted ratios for nine-item sequences are only slightly larger than

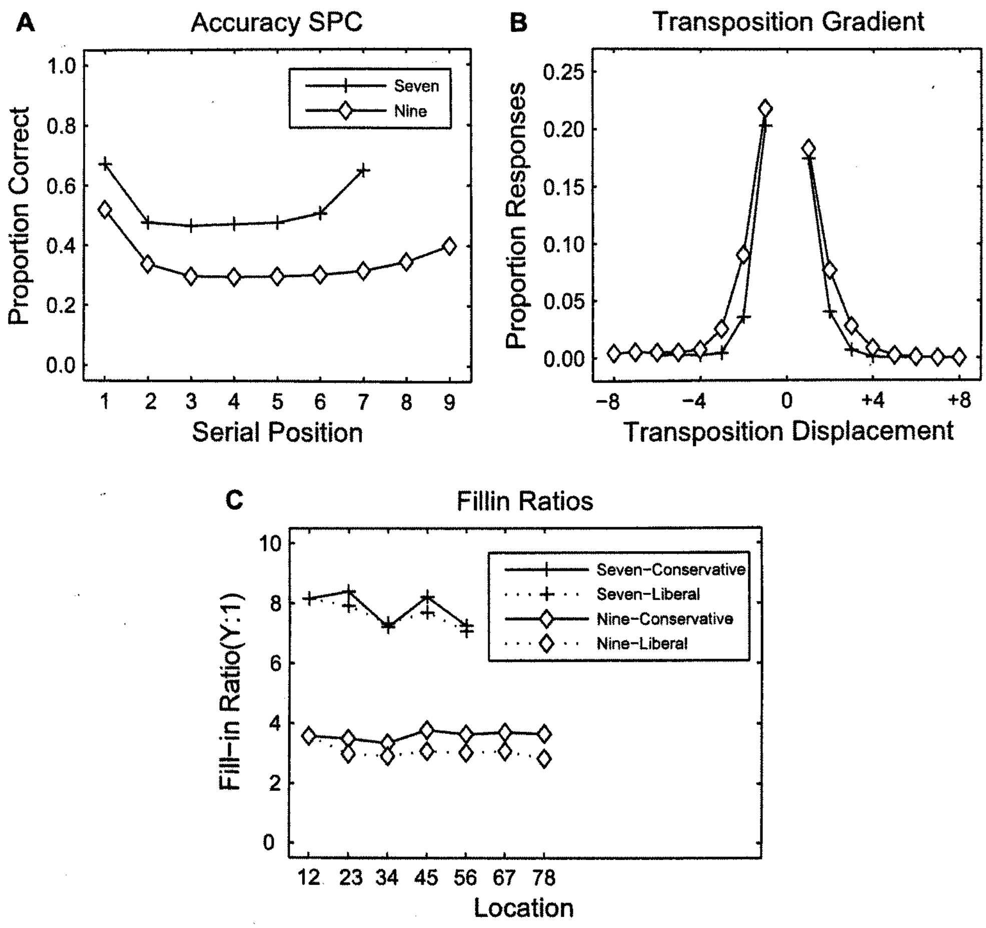

Figure A3-3 *Fits of the linear primacy gradient and response suppression model to the data of Experiment 12. Panels show predictions for: accuracy serial position curves (A), transposition gradients (B), and ratios of fill-in to infill errors (C).*

the empirically observed values. The mean ratios of fill-in to infill errors predicted by the linear primacy gradient model, averaged across error locations and scoring procedure, for seven-item and nine-item sequences, were: 8:1 and 3:1, respectively, compared to empirically observed mean ratios of 3:1 and 2:1. As for the fits to Experiments 11 and 1, these predicted mean ratios are smaller than those predicted by the exponential primacy gradient model. However, for the data for seven-item sequences, at least, the predicted ratios are still markedly larger than the empirically observed mean ratios.

Appendix 4

Implementation of restricted end marker

This appendix describes the implementation of the restricted end marker in the PG+PM+RE+RS model examined in Chapter 8.

Restricted end marker

The restricted end marker implements the idea that the final item in a to-be-remembered sequence is tagged for its position with reference to the end of that sequence. It is therefore a constrained version of the extensive end marker assumed in Henson's (1998a) Start-End Model (see Chapter 1), in which every item in a sequence is coded with reference to the end of that sequence. Farrell and Lelievre (2009) have shown that the benefits of end anchoring in serial recall, including enhanced effects of recency, can be realised through the assumption of a restricted end marker, therefore circumventing the need to posit a fully extensive end marker.

Implementation

The procedure for implementing the restricted end marker in the PG+PM+RE+RS model is as follows. First, starting activations were computed for the positional component of the model using equation 2-2 presented in Chapter 2. Specifically, the activation a of the item node j for the current output position p based upon positional marking pm was determined as follows:

$$a_{j,pm} = \lambda\phi^{|j-p|}$$

Where λ represents the weighting of activation of the position markers and ϕ represents the distinctiveness of the position markers. These activations where then modified to represent the contribution of the restricted end marker to the coding of positional information in the following manner:

$$a_{j,pmre} = \begin{bmatrix} a_{j,pm}(1-\varepsilon) & \text{If } p < ll \,\&\, j = ll \\ a_{j,pm}(1-\varepsilon) & \text{If } p = ll \,\&\, j < ll \\ a_{j,pm} & \text{Otherwise} \end{bmatrix} \quad \text{(A4-1)}$$

Where $a_{j,pmre}$ represents a node's position marking activation modified by incorporation of the restricted end marker, ll represents the list length being simulated, and ε is a new parameter governing the strength of the restricted end marker ($0 < \varepsilon < 1$). The function of equation A4-1 is three-fold: First, at all output positions except the final output position, the activation of the node representing the final item in the sequence is multiplied by ($1-\varepsilon$), yielding a proportional reduction in its activation. This therefore reduces the strength with which the final item in the sequence competes for recall at pre-terminal serial positions. Second, at the final position, the activation of all item nodes except the node representing the final item are multiplied by ($1-\varepsilon$), yielding a proportional reduction in their activation. This therefore reduces the strength with which items from pre-terminal serial positions compete for recall at the final position, thereby increasing the likelihood that the correct final item will successfully be recalled. Third, the activations of nodes that do not meet either of the two above constraints are left unchanged.

The position marking starting activations for each output position, modified by incorporation of the restricted end marker, were then normalised as follows:

$$a_{j,pmre} = \frac{a_{j,pmre}}{\sum_i a_{i,pmre}}$$

A separate set of starting activations were then specified for the primacy gradient pg component of the model, using equation 2-4 of Chapter 2:

$$a_{j,pg} = a_1 \gamma^{j-1}$$

Where a_1 represents the activation of the first item node and γ is a parameter controlling the steepness of the primacy gradient. The same primacy gradient was established for each output position. The position marking activations, modified by the incorporation of the restricted end

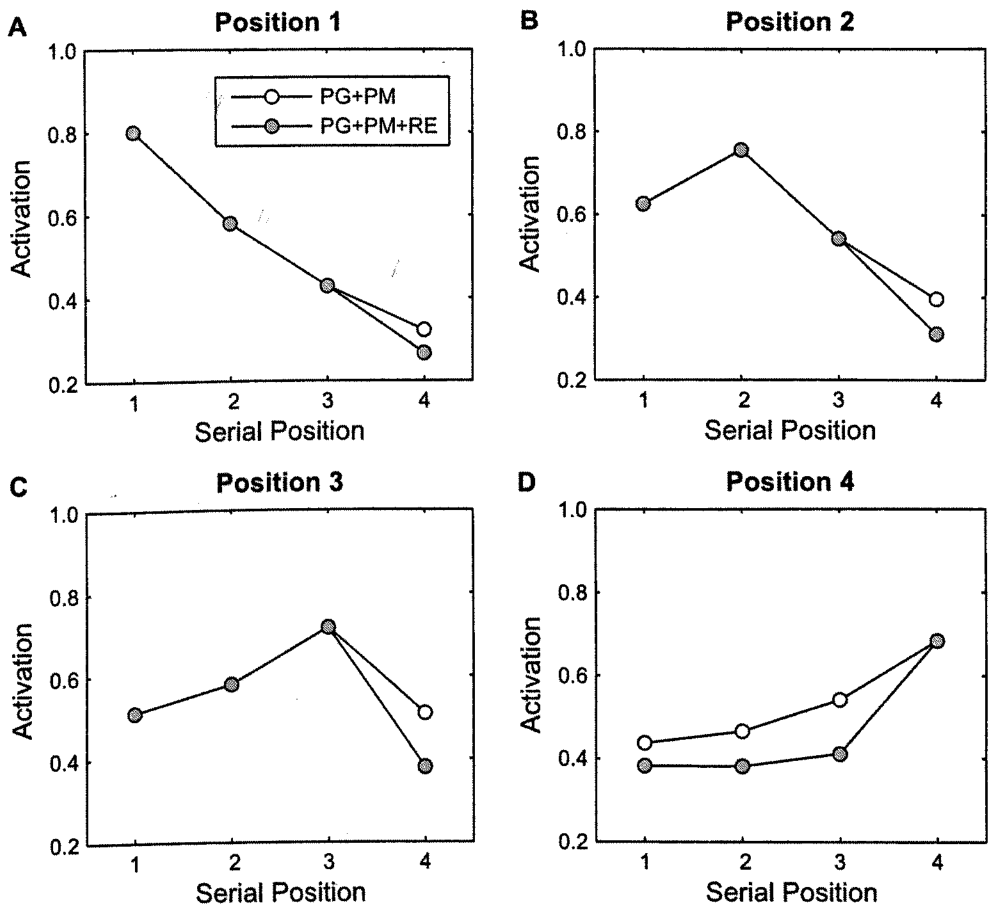

Figure A4-1 *Example starting activations for the PG+PM and PG+PM+RE models for a four-item sequence. Panels show activations for the first position (A), second position (B), third position, (C) and fourth position (D).*

marker and the primacy gradient activations, where then combined using equation 2-6 of Chapter 2:

$$a_j = (1-\omega) \cdot a_{j,pmre} + \omega \cdot a_{j,pg}$$

Where ω represents the attentional weight given to the positional and primacy dimensions of ordering, and was set to a value of .5 (as in all previous simulations of the PG+PM+RS model), thereby giving equal weight to the two dimensions.

Example starting activations

To illustrate the impact of the restricted end marker on the competition process, this section presents example starting activations for the combination of a primacy gradient with position marking (PG+PM), and the same combination of representational principles, augmented with a restricted end marker (PG+PM+RE)[1]. The example activations for the PG+PM model were generated using the following parameter settings: $\lambda = 1$; $\phi = .65$; $a_1 = .6$; and $\gamma = .85$. The same parameter settings were employed for the PG+PM+RE model, but incorporating the restricted end marker, ε, which was set to a value of .4.

The example starting activations are shown in Figure A4-1. It is apparent from inspection of this figure that the impact of adding the restricted end marker to the combination of a primacy gradient and positional marking is to reduce the extent to which the final item competes for recall at pre-terminal serial positions, whilst also reducing the extent to which pre-terminal sequence items compete for recall at the final serial position. The restricted end marker therefore reduces the likelihood that the final item will be recalled prematurely, and concomitantly increases the likelihood that the final item will be recalled in its correct end of sequence position.

---

[1] Because the aim here it to show how the restricted end marker modifies the starting activations assigned to items response suppression was not incorporated in the models.